



KLUG ENTSCHEIDEN

... auch in der Lehre!

Pilotstudie zur klinischen Entscheidungskompetenz von Medizinstudenten deckt verbreitete Fehlmeinungen auf. Auf Grundlage der „Klug entscheiden“-Empfehlungen soll nun entsprechendes Lehrmaterial entwickelt werden.



Foto: oipa

Die Klug-entscheiden-Initiative der Deutschen Gesellschaft für Innere Medizin (DGIM) und ihrer Schwerpunktgesellschaften zielt darauf ab, die medizinische Versorgung nachhaltig zu verbessern (1). Hierzu wurden unter Beteiligung zahlreicher Fachgesellschaften diagnostische und therapeutische Maßnahmen identifiziert, die häufig nicht fachgerecht erbracht werden (2). Da sowohl überflüssige als auch unterlassene Untersuchungen und Behandlungen einen Nachteil für Patienten/-innen bedeuten können, wurden in dem Konsensprozess beide gleichermaßen berücksichtigt (3). Die Adressaten der „Klug entscheiden“-Empfehlungen (KEE) sind in erster Linie die in der ärztlichen Versorgung tätigen Kolleginnen und Kollegen.

Zugleich erscheint es sinnvoll, den Grundgedanken der Initiative bereits den Medizinstudierenden zu vermitteln.

Der Erwerb klinischer Entscheidungskompetenz gehört zu den zentralen Lernzielen des Medizinstudiums. Neben den in das Studium integrierten Praxisphasen (Unterricht am Krankenbett, Blockpraktikum, Famulatur und Praktisches Jahr) existieren mit dem fallbasierten (4) und dem problemorientierten Lernen (5) spezielle Unterrichtsformate für die Einübung klinischer Entscheidungsprozesse. Bislang wird allerdings kaum geprüft, ob die entsprechenden Lernziele von den Studierenden auch erreicht wurden. Dies liegt unter anderem daran, dass die Prüfung von Entscheidungskompetenz ein anderes For-

mat erfordert als die üblicherweise eingesetzten Multiple-Choice-Fragen. Ein solches, in diesem Kontext eher geeignetes Format ist die „Key Feature“-Prüfung (6). Als Key Feature wird eine Schlüsselstelle im klinischen Entscheidungsprozess bezeichnet; Fehlentscheidungen, die an diesen Schlüsselstellen getroffen werden, können die Diagnosestellung verzögern oder dazu führen, dass Patienten falsche oder unnötige Therapien erhalten. Bei Key-Feature-Prüfungen werden Studierende mit klinischen Szenarien konfrontiert und an den jeweiligen Schlüsselstellen des diagnostischen und therapeutischen Prozedere werden offene Fragen zum optimalen Vorgehen gestellt.

Nach Beantwortung jeder Frage erhalten die Studierenden Feedback

zu den jeweils richtigen und falschen Antworten. Bei computer-gestützten Key-Feature-Prüfungen lassen sich die Antworten auf die offenen Fragen leicht auswerten, weil die Studierenden ihre Antwort anhand eines Drop-down-Menüs aus einer Vielzahl (ca. 4 000) hinterlegter Optionen auswählen können und die Auswertung automatisch erfolgt. Eine aktuelle Übersichtsarbeit belegt, dass das Format der Key-Feature-Prüfung dazu geeignet ist, klinische Entscheidungskompetenz zu messen (7).

Wie in anderen Prüfungen auch wird bei der Auswertung von Key-Feature-Prüfungen üblicherweise der Anteil der Studierenden mit richtiger Antwort betrachtet (8). Hinsichtlich der Ziele der Klug-entscheiden-Initiative sind aber auch die Falschantworten von Interesse, denn aus ihnen können sich Hinweise auf Defizite in der Lehre ergeben. So könnte eine ungünstige Gewichtung von Lehrinhalten dazu führen, dass angehende Ärztinnen und Ärzte nicht indizierte Maßnahmen zu häufig und andere, indizierte Maßnahmen zu selten durchführen.

Bislang wurde die Häufigkeitsverteilung der Falschantworten von Key-Feature-Prüfungen noch nicht untersucht. Ziel dieser Studie war es, das Antwortverhalten von Studierenden der Humanmedizin in Key-Feature-Prüfungen auszuwerten. Von besonderem Interesse waren dabei Änderungen der Antwortverteilung im Laufe eines Semesters und der mittelfristige Lernerfolg nach 6 Monaten.

Methoden

An der Universitätsmedizin Göttingen nahmen zwischen dem Wintersemester 2013/14 und dem Wintersemester 2015/16 alle Studierenden jeweils zu Beginn und am Ende des 7. Fachsemesters an nicht benoteten Key-Feature-Prüfungen („Eingangstestat“ und „Ausgangstestat“) teil; darin war ein Querschnitt der während des Semesters behandelten Themen abgebildet. Zur Messung der mittelfristigen Retention dieser Inhalte fand im Folgesemester eine unangekündigte (und ebenfalls nicht benotete) Nachprüfung im

TABELLE 1

Studierenden-Antworten auf Fragen zu diagnostischen Maßnahmen. Angegeben ist der prozentuale Anteil der Studierenden (n), die zu den jeweiligen Zeitpunkten in den Key-Feature-Prüfungen die jeweiligen Antwortoptionen gewählt haben.

	Eingangstestat	Abschlusstestat	Nachprüfung
Entscheidungskriterium für/gegen eine D-Dimer-Bestimmung: Wells-Score			
Richtige Antwort	21,7 (78)	73,1 (263)	68,6 (247)
Hämodynamische Stabilität	0,8 (3)	3,1 (11)	1,1 (4)
Bildgebung/Labordiagnostik	8,6 (31)	3,3 (12)	3,6 (13)
Andere Falschantwort	68,9 (248)	20,6 (74)	26,7 (96)
Maßnahme bei hoher Wahrscheinlichkeit für das Vorliegen einer Lungenembolie: CT des Thorax			
Richtige Antwort	18,1 (65)	49,2 (177)	41,4 (149)
Andere Bildgebung	9,4 (34)	8,9 (32)	10,8 (39)
Systemische Thrombolyse	20,8 (75)	16,4 (59)	21,1 (76)
D-Dimer-Bestimmung	13,6 (49)	9,7 (35)	10,3 (37)
Andere Falschantwort	38,1 (137)	15,8 (57)	16,4 (59)
Maßnahme zur Sicherung der Diagnose „Lupus-Nephritis“: Nierenbiopsie			
Richtige Antwort	23,1 (83)	79,2 (285)	83,9 (302)
Bildgebung	14,7 (53)	1,1 (4)	1,4 (5)
Labordiagnostik	17,2 (62)	3,6 (13)	3,3 (12)
Andere Falschantwort	45,0 (162)	16,1 (58)	11,4 (41)
Parameter zur Diagnose einer Obstruktion: FEV₁/VC (Ist-Wert)			
Richtige Antwort	20,8 (75)	60,0 (216)	55,3 (199)
FEV ₁	29,7 (107)	13,6 (49)	12,8 (46)
FEV ₁ /VC (% vom Soll)	15,0 (54)	15,8 (57)	17,2 (62)
Peak-flow	8,9 (32)	0,3 (1)	1,9 (7)
Andere Falschantwort	25,6 (92)	10,3 (37)	12,8 (46)
Diagnostik bei Patienten mit COPD und Vigilanzminderung unter O₂-Gabe: Blutgasanalyse			
Richtige Antwort	31,9 (115)	63,6 (229)	60,8 (219)
Bildgebung	28,9 (104)	8,3 (30)	10,0 (36)
Andere Labordiagnostik	7,8 (28)	4,7 (17)	3,1 (11)
Andere Falschantwort	31,4 (113)	23,3 (84)	26,1 (94)
Diagnostische Methode bei Verdacht auf orthostatisch bedingte Synkope: Schellong-Test			
Richtige Antwort	29,4 (106)	66,7 (240)	65,0 (234)
Kipptisch-Untersuchung	26,4 (95)	26,1 (94)	28,1 (101)
Andere Falschantwort	44,2 (159)	7,2 (26)	6,9 (25)

Key-Feature-Format statt („Nachprüfung“).

Das Eingangs- und das Ausgangstestat sowie die Nachprüfung bestanden aus jeweils 28 Fragen, von denen sich 13 auf die am ehesten vorliegende Erkrankung bezogen, 6 auf die indizierten diagnostischen Maßnahmen, 4 auf Methoden der Risikostratifizierung und des Patientenmanagements sowie 5 auf therapeutische Maßnahmen. Im Fo-

kus der hier vorgestellten Analyse stehen die letzten 3 Fragengruppen, da sie einige der im Kontext der Klug-entscheiden-Initiative besonders relevanten Aspekte thematisieren.

Zur Auswertung der Prüfungsergebnisse wurden für jede Frage Kategorien von Falschantworten gebildet. Für jede Frage wurde, nach Prüfungszeitpunkten getrennt, die Häufigkeitsverteilung der Antworten de-

TABELLE 2

Studierenden-Antworten auf Fragen zur Risikostratifizierung und zum Patientenmanagement. Angegeben ist der prozentuale Anteil der Studierenden (n), die zu den jeweiligen Zeitpunkten in den Key-Feature-Prüfungen die jeweiligen Antwortoptionen gewählt haben.

	Eingangstestat	Abschlusstestat	Nachprüfung
Risikostratifizierung bei bestehender Lungenembolie: Rechtsherzbelastung			
Richtige Antwort	8,1 (29)	52,8 (190)	48,6 (175)
Andere apparative Methoden	31,9 (115)	7,8 (28)	13,9 (50)
Andere Labordiagnostik	19,2 (69)	7,5 (27)	5,6 (20)
Andere Falschantwort	40,8 (147)	31,9 (115)	31,9 (115)
Kriterium zur Entscheidung über eine stationäre Aufnahme bei Pneumonie: CRB-65-Score			
Richtige Antwort	3,1 (11)	75,8 (273)	68,1 (245)
Labordiagnostik	59,7 (215)	4,2 (15)	3,1 (11)
Andere Falschantwort	37,2 (134)	20,0 (72)	28,9 (104)
Vordringliche Maßnahme bei schwerer Hyponatriämie (117 mM): stationäre Einweisung			
Richtige Antwort	10,0 (36)	45,3 (163)	50,8 (183)
Natrium-Substitution	31,9 (115)	20,0 (72)	19,7 (71)
Absetzen von Diuretika	9,2 (33)	17,5 (63)	10,6 (38)
Diuretika-Gabe	10,0 (36)	2,5 (9)	4,2 (15)
Andere Falschantwort	38,9 (140)	14,7 (53)	14,7 (53)
Entscheidung über die Indikation zur Antikoagulation bei Vorhofflimmern: CHA₂DS₂-VASc-Score			
Richtige Antwort	17,5 (63)	80,3 (289)	79,4 (286)
Vorliegen von Vorhofflimmern	14,7 (53)	2,2 (8)	1,1 (4)
Labordiagnostik	8,1 (29)	0,6 (2)	0,8 (3)
Andere Falschantwort	59,7 (215)	16,9 (61)	18,6 (67)

skriptiv analysiert. Im Interesse einer longitudinalen Betrachtung wurden dabei nur Daten von Studierenden berücksichtigt, die an allen 3 Prüfungen teilgenommen hatten.

Ergebnisse

Die kompletten Prüfungsdaten von 360 Studierenden gingen in die Analyse ein. Von den 15 ausgewerteten Fragen wurden im Eingangstestat durchschnittlich $3,7 \pm 2,2$ richtig beantwortet, im Abschlusstestat lag dieser Anteil mit $9,8 \pm 3,4$ deutlich höher, und in der Nachprüfung wurden noch $9,1 \pm 3,3$ Fragen richtig beantwortet. Das Antwortverhalten der Studierenden in den 3 Prüfungen ist in den *Tabellen 1–3* dargestellt.

Tabelle 1 zeigt die Ergebnisse für Fragen zu diagnostischen Maßnahmen. Für einige Inhalte (z. B. Stellenwert des Wells-Score im Rahmen der Diagnostik einer Lungenembolie) zeigt sich im zeitlichen

Verlauf eine deutliche Zunahme der Häufigkeit richtiger Antworten bei gleichzeitiger Abnahme der Häufigkeit spezifischer und unspezifischer Falschantworten. Für andere Lernziele (z. B. lungenfunktionelle Obstruktionsdiagnostik) nahm im Semesterverlauf zwar die Häufigkeit spezifischer Falschantworten zugunsten der richtigen Antwort ab; wichtige Falschantworten wurden aber in allen 3 Prüfungen von einem nicht unerheblichen Teil der Studierenden eingetragen.

Im Hinblick auf die KEE sind die Ergebnisse zur Frage nach der Diagnostik der Wahl bei einer hohen Wahrscheinlichkeit für das Vorliegen einer Lungenarterienembolie interessant: Zwar konnten zu Semesterende 50 % der Studierenden und 6 Monate später noch 40 % der Studierenden das CT als adäquate Maßnahme benennen; der Anteil der Studierenden, die auch ohne Sicherung der Diagnose eine systemi-

sche Thrombolyse für indiziert hielten, lag jedoch durchgehend unverändert um 20 %.

Tabelle 2 fasst die Daten für Fragen zur Risikostratifizierung und zum Patientenmanagement zusammen. Aus Sicht der KEE sind die Ergebnisse zur Frage nach dem Kriterium für/gegen eine Antikoagulation bei Vorhofflimmern interessant: Hier stieg der Anteil der Studierenden, die korrekterweise den CHA₂DS₂-VASc-Score als Entscheidungskriterium identifizierten, während des Semesters von 20 % auf 80 %; dieser zuletzt hohe Anteil blieb auch nach einem halben Jahr unverändert.

In *Tabelle 3* sind die Studierenden-Antworten auf Fragen zu therapeutischen Maßnahmen dargestellt. Auch für diese Aspekte zeigte sich mitunter ein substanzieller und nachhaltiger Lernerfolg, wenngleich im Abschlusstestat von zahlreichen Studierenden weiterhin Falschantworten gegeben wurden, die im Falle einer Umsetzung in der klinischen Praxis Patienten gefährden könnten (z. B. Gabe von Schilddrüsenhormonen oder Jod bei manifester Hyperthyreose). Im Hinblick auf die KEE war zu beobachten, dass der Anteil der Studierenden, die bei bestehender Indikation zur Langzeit-Sauerstoff-Therapie stattdessen fälschlicherweise eine (nichtinvasive) Beatmung empfahlen, während des Semesters und auch in der Nachbeobachtungsphase sogar zunahm.



Diskussion

Die vorliegende Analyse der Prüfungsdaten von 360 Studierenden der Humanmedizin bestätigt, dass das genutzte Fragenformat dazu geeignet ist, die diagnostische und therapeutische Entscheidungskompetenz von Studierenden zu beurteilen. Insbesondere war es erstmals möglich, aus einer quantitativen Auswertung der Falschantworten von Key-Feature-Fragen Hinweise auf Defizite in der Lehre abzuleiten. Für alle 3 betrachteten Berei-

TABELLE 3

Studierenden-Antworten auf Fragen zu therapeutischen Maßnahmen. Angegeben ist der prozentuale Anteil der Studierenden (n), die zu den jeweiligen Zeitpunkten in den Key-Feature-Prüfungen die jeweiligen Antwortoptionen gewählt haben.

	Eingangstestat	Abschlusstestat	Nachprüfung
Maßnahme bei Lungenarterienembolie (LAE) mit hämodynamischer Instabilität: Thrombolyse			
Richtige Antwort	37,8 (136)	67,8 (244)	57,2 (206)
Andere Medikamente	27,2 (98)	12,2 (44)	13,6 (49)
Thrombektomie	2,5 (9)	0,8 (3)	0,8 (3)
Andere Falschantwort	32,5 (117)	19,2 (69)	28,3 (102)
Therapie bei CO₂-Narkose: nichtinvasive Beatmung			
Richtige Antwort	16,9 (61)	56,9 (205)	41,9 (151)
CPAP	4,2 (15)	11,9 (43)	18,3 (66)
Sauerstoff-Gabe	28,1 (101)	15,3 (55)	21,1 (76)
Azidoseausgleich	11,9 (43)	3,3 (12)	5,6 (20)
Intubation und Beatmung	16,4 (59)	2,8 (10)	2,8 (10)
Andere Falschantwort	22,5 (81)	9,7 (35)	10,3 (37)
Maßnahme bei trockenem Reizhusten unter Herzinsuffizienz-Medikation: ACE-Hemmer absetzen			
Richtige Antwort	69,4 (250)	90,8 (327)	89,7 (323)
Medikamentöse Therapie	19,2 (69)	3,9 (14)	3,6 (13)
Andere Falschantwort	11,4 (41)	5,3 (19)	6,7 (24)
Maßnahme bei manifester Hyperthyreose, thyreostatischer Therapie und Amiodaron-Einnahme: Amiodaron absetzen			
Richtige Antwort	11,9 (43)	64,7 (233)	64,7 (233)
Thyreoidektomie	15,0 (54)	4,4 (16)	6,4 (23)
Andere medikamentöse Therapie	21,4 (77)	7,8 (28)	9,7 (35)
Radio-Jod-Therapie	4,4 (16)	1,7 (6)	1,7 (6)
Gabe von Jod oder T ₃ /T ₄	7,2 (26)	3,6 (13)	3,9 (14)
Andere Falschantwort	40,0 (144)	17,8 (64)	13,6 (49)
Maßnahme bei einem pO₂ < 55 mmHg: Langzeit-Sauerstoff-Therapie			
Richtige Antwort	52,8 (190)	54,7 (197)	41,7 (150)
Beatmung (CPAP/NIV)	10,6 (38)	25,8 (93)	30,0 (108)
Andere Falschantwort	36,7 (132)	19,4 (70)	28,3 (102)

che (Diagnostik, Risikostratifizierung/Management und Therapie) ließen sich Inhalte mit hohem und nachhaltigem Lernerfolg, aber auch Inhalte mit einer erheblichen Persistenz von Falschantworten identifizieren. Eine Umsetzung der falschen Antworten in der klinischen Praxis hätte mitunter negative Konsequenzen für Patienten.

Die in dieser Studie eingesetzten Key-Feature-Fragen wurden bereits im Jahr 2013, also vor Beginn der Arbeiten an den Klug-entscheiden-Empfehlungen der DGIM entworfen. Dennoch weisen die Fragen zum

diagnostischen Vorgehen bei Verdacht auf Lungenembolie, zum Management von Vorhofflimmern und zur Indikation einer Langzeit-Sauerstoff-Therapie enge Bezüge zu den KEE auf. In einem nächsten Schritt sollen – analog zur Entwicklung einer landesweiten Key-Feature-Prüfung in Amerika (9) – in einem von der DGIM geförderten Projekt auf der Grundlage der KEE neue Key-Feature-Fragen formuliert und an mehreren medizinischen Fakultäten pilotiert werden. Auf diese Weise wird es erstmals möglich sein, die Verankerung des Konzepts und der

konkreten Inhalte der Klug-entscheiden-Initiative im Medizinstudium zu beurteilen und hieraus gegebenenfalls Empfehlungen für die Weiterentwicklung der Lehre abzuleiten.

Fazit

Eine Analyse der Falschantworten von Studierenden in fallbasierten Prüfungen trägt dazu bei, Erfolge und Defizite in der universitären medizinischen Lehre zu identifizieren. Das Format der Key-Feature-Frage erscheint geeignet, um zu untersuchen, inwieweit die Klug-entscheiden-Empfehlungen der DGIM im Rahmen des Medizinstudiums vermittelt werden.

Kernaussagen

- Die Vermittlung klinischer Entscheidungskompetenz ist ein zentrales Ziel des Medizinstudiums.
- Mit fallbasierten Key-Feature-Fragen kann diese Kompetenz geprüft werden.
- In bisherigen Studien wurde zu meist nur die Anzahl der richtigen Antworten berichtet; hierbei blieben die Häufigkeit und die Qualität der Falschantworten unberücksichtigt.
- Eine longitudinale Analyse der Prüfungsdaten von 360 Studierenden des 3. klinischen Semesters an der Universitätsmedizin Göttingen ergab, dass wichtige Fehlmeinungen bei Studierenden langfristig persistieren.
- In weiterführenden Studien soll untersucht werden, inwieweit die „Klug entscheiden“-Empfehlungen der DGIM in der universitären medizinischen Lehre thematisiert werden.

Milena Goldmann
Göttingen

Prof. Dr. med. Gerd Hasenfuß
Klinik für Kardiologie und Pneumologie,
Universitätsmedizin Göttingen

Terese Dehl
Bereich Medizindidaktik und Ausbildungsforschung,
Studiendekanat, Universitätsmedizin Göttingen

Prof. Dr. med. Tobias Raupach
Klinik für Kardiologie und Pneumologie und
Bereich Medizindidaktik und Ausbildungsforschung,
Studiendekanat, Universitätsmedizin Göttingen

@ Literatur im Internet:
www.aerzteblatt.de/lit4716
oder über QR-Code.



LITERATURVERZEICHNIS HEFT 47/2016, ZU:

KLUG ENTSCHEIDEN

... auch in der Lehre!

Pilotstudie zur klinischen Entscheidungskompetenz von Medizinstudenten deckt verbreitete Fehlmeinungen auf. Auf Grundlage der „Klug entscheiden“-Empfehlungen soll nun entsprechendes Lehrmaterial entwickelt werden.

LITERATUR

1. Hasenfuß G, Märker-Herrmann E, Hakkel M, Fölsch U: Gegen Unter- und Überversorgung. Dtsch Ärztebl 2016; 113: A600–2.
2. Jung N: Klug entscheiden in der Infektiologie. Dtsch Ärztebl 2016; 113: A608–10.
3. Fölsch U, Faulbaum F, Hasenfuß G: Wie Internisten das Problem von Über- und Überversorgung werten. Dtsch Ärztebl 2016; 113: A604–6.
4. Thistlethwaite JE, Davies D, Ekeocha S, et al.: The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. Med Teach 2012; 34: e421–44.
5. Barrows HS, Mitchell DL: An innovative course in undergraduate neuroscience. Experiment in problem-based learning with „problem boxes“. Br J Med Educ 1975; 9: 223–30.
6. Page G, Bordage G, Allen T: Developing key-feature problems and examinations to assess clinical decision-making skills. Acad Med 1995; 70: 194–201.
7. Hrynchak P, Takahashi SG, Nayer M: Key-feature questions for assessment of clinical reasoning: a literature review. Med Educ 2014; 48: 870–83.
8. Raupach T, Andresen JC, Meyer K, et al.: Test-enhanced learning of clinical reasoning: a crossover randomised trial. Med Educ 2016; 50: 711–20.
9. Bronander KA, Lang VJ, Nixon LJ, et al.: How we developed and piloted an electronic key features examination for the internal medicine clerkship based on a US national curriculum. Med Teach 2015; 37: 807–12.