

# Validität der Empfehlungen zur Sekundärprävention von Dickdarmkrebs

## Validity of Recommendations for Secondary Prevention of Colon Cancer

### Autoren

F. Porzsolt<sup>1</sup>, D. Ittner<sup>1</sup>, D. Henne-Bruns<sup>2</sup>

### Institute

<sup>1</sup> Allgemein- und Viszeralchirurgie, Universitätsklinikum Ulm, AG Versorgungsforschung, Ulm, Deutschland

<sup>2</sup> Allgemein- und Viszeralchirurgie, Universitätsklinikum Ulm, Zentrum für Chirurgie, Ulm, Deutschland

### Schlüsselwörter

- Screening
- Überleben
- Krebserkrankung

### Key words

- screening
- survival
- cancer

### Zusammenfassung



**Hintergrund:** Präventionsprogramme sollten nur empfohlen werden, wenn sie leisten, was sie zu leisten versprechen. Deshalb haben wir die Variation und Validität der Empfehlungen zum Darmkrebs-Screening von 9 Organisationen geprüft.

**Methoden:** Angaben zum Screening-Alter, chemischen Stuhltest (gFOBT), immunologischen Stuhltest (FIT), DNA-Stuhltest, Sigmoidoskopie, Koloskopie, Doppelkontrastuntersuchung/Barium-Doppelkontrasteinlauf und zur virtuellen Koloskopie/CT-Kolonografie wurden in 3 Schritten analysiert. 1) Erfassung aller Referenzen, die von den 9 Organisationen ausgewählt wurden, 2) Zuordnung der Referenzen zur Mortalität, Inzidenz und Sensitivität/Spezifität und 3) Bewertung der Validität bei Mortalitätsstudien.

**Ergebnisse:** Der Nachweis von okkultem Blut wird als einzige Screening-Methode von allen 9 Organisationen empfohlen. Die Koloskopie wird von 7 Organisationen empfohlen. Von 33 Referenzen zur Begründung des Screenings hatten 15 andere Endpunkte als die Mortalität. Eine Publikation war eine Metaanalyse. Elf von 17 Veröffentlichungen untersuchten den chemischen Stuhltest, 3 die Sigmoidoskopie und je eine Studie den immunologischen Stuhltest, die Koloskopie oder generell die Diagnostik des Darms. Durchschnittlich wurden in allen Studien 2 von 9 Validitätskriterien vollständig, 5 teilweise und 2 nicht erfüllt. In 2 Veröffentlichungen war keines der Validitätskriterien vollständig erfüllt.

**Schlussfolgerungen:** Die Analyse der Empfehlungen zum Darmkrebs-Screening zeigte, dass 9 Organisationen unterschiedliche Ziele verfolgen und unterschiedliche Empfehlungen aussprechen. Die gewissenhafte und aufwendige Prüfung der wissenschaftlichen Studien zur Mortalität, die diesen Empfehlungen zugrunde liegen, weisen erhebliche Mängel auf und können damit die internationalen Empfehlungen zum Darmkrebs-Screening

### Abstract



**Background:** Prevention programmes should only be recommended if they achieve what they promise to achieve. Therefore, we checked the variation and validity of recommendations for screening for colorectal cancer of nine organisations.

**Methods:** We analysed the information concerning recommended screening age, guaiac faecal occult blood test (gFOBT), faecal immunological test (FIT), faecal DNA test, sigmoidoscopy, colonoscopy, double-contrast examination/double-contrast barium enema, and virtual colonoscopy/CT colonography in the following three steps: 1) we gathered the references quoted by the nine organisations; 2) references were categorised according to mortality, incidence and sensitivity/specificity; 3) the validity of references that reported reduced mortality attributed to screening were evaluated.

**Results:** Evidence of occult faecal blood was the only screening method recommended by all nine organisations. Colonoscopy was recommended by seven organisations. Fifteen of the 33 references used endpoints other than mortality to justify screening. One publication was a meta-analysis. Eleven of 17 publications evaluated the gFOBT, three evaluated sigmoidoscopy, one FIT, one colonoscopy, and one general diagnosis of the intestine. On average, two of nine validity criteria were completely fulfilled, five only partially, and two were not fulfilled. In two publications, none of the validity criteria were completely met.

**Conclusion:** Analysis of screening for colorectal cancer revealed that nine organisations had different goals and different recommendations. Scrupulous and thorough evaluation of the scientific studies in relation to mortality, upon which these recommendations are based, revealed numerous shortcomings and therefore could not sufficiently substantiate the international recom-

### Bibliografie

**DOI** <http://dx.doi.org/10.1055/s-0032-1328074>  
 Online-publiziert 25. 1. 2013  
 Zentralbl Chir 2015; 140:  
 390–396 © Georg Thieme  
 Verlag KG Stuttgart · New York ·  
 ISSN 0044-409X

### Korrespondenzadresse

**Prof. Franz Porzsolt**  
 AG Versorgungsforschung  
 Allgemein- und Viszeral-  
 chirurgie,  
 Universitätsklinikum Ulm  
 Albert-Einstein-Allee 23  
 89081 Ulm  
 Deutschland  
 Tel.: 0173/3 2997 74  
 Fax: 0731/5 00-5 35 03  
 franz.porzolt@uniklinik-ulm.de

ning nicht hinreichend begründen. Sinnvoll wäre, über die zu erhebenden Daten einen Konsensus herbeizuführen, mit welchem eine belastbare Basis für Versorgungsentscheidungen generiert werden kann.

## Einleitung

Krebs stand im Jahr 2004 an 3. Stelle der Todesursachenstatistik [1]. Das kolorektale Karzinom war mit etwa 1,2 Millionen Erkrankungsfällen und 609 000 Todesfällen in 2008 die dritthäufigste Krebsart und vierthäufigste Krebstodesursache weltweit [2]. Demografischen Trends zufolge könnte die Erkrankungsrate in den nächsten 2 Jahrzehnten zivilisationsbedingt um bis zu 80% von heute 1,2 auf 2,2 Millionen Darmkrebsfälle weltweit zunehmen [2]. In den USA und Deutschland ist Darmkrebs mit etwa 144 000 bzw. 68 000 Neuerkrankungen pro Jahr der zweithäufigste bösartige Tumor [3, 4].

In Politik und Gesellschaft findet die Sekundärprävention breite Unterstützung [5], obwohl renommierte wissenschaftliche Organisationen (z.B. die Cochrane Collaboration) den Nutzen dieser Maßnahmen durchaus infrage stellen. In Deutschland wurde im Jahr 1977 der chemische Stuhltest, gFOBT und im Jahr 2002 zusätzlich die Koloskopie zur Sekundärprävention von Darmkrebs per Leitlinie eingeführt [6]. In Großbritannien und Frankreich hingegen beschränkt sich die Darmkrebsfrüherkennung auf den chemischen Stuhltest. Deshalb sollte die Frage beantwortet werden, ob wir in Deutschland den Nutzen der Koloskopie überschätzen oder dieser in anderen Ländern, z.B. in Großbritannien und Frankreich unterschätzt wird. Dazu haben wir die Validität der Empfehlungen ausgewählter internationaler Organisationen zur Sekundärprävention des kolorektalen Karzinoms analysiert.

## Material und Methoden

Es wurden Screeningempfehlungen von 9 deutschen, amerikanischen und internationalen Organisationen geprüft (► Tab. 1). Im 1. Schritt wurden die Referenzen der Empfehlungen erfasst. Im 2. Schritt wurden diese Referenzen den Themen Mortalität, Inzidenz und Sensitivität/Spezifität zugeordnet. Im 3. Schritt wurde die Validität der Referenzen geprüft, die sich mit der Reduktion der Mortalität durch Früherkennungsmaßnahmen befassen, weil Mortalität ein klinisch relevanter Endpunkt ist [7], der sowohl Personen einschließt, die nie an Darmkrebs erkranken wie auch geheilte Patienten. Die Mortalität wird durch vorgezogene Diagnosen oder Überdiagnosen<sup>1</sup> im Gegensatz zur Überlebenszeit kaum beeinflusst [8].

Die Validität der ausgewählten Publikationen wurden mithilfe einer publizierten Strategie [9, 10] analysiert. Diese Methode umfasst 9 Kriterien und 2 Werturteile (► Tab. 2). Da wir in einer früheren Analyse zeigen konnten, dass sich zwar keine systematischen Unterschiede, aber erhebliche Streuungen bei der Bedeutung dieser 9 Validitätskriterien nachweisen lassen [11], haben wir jedes vollständig erfüllte Kriterium einheitlich mit einem Punkt und jedes teilweise erfüllte Kriterium mit 0,5 Punkten bewertet. Somit konnten pro Publikation maximal 9 Validitäts-

<sup>1</sup> Überdiagnose: Entdeckung einer Veränderung, die mit Ausnahme fehlender Symptome von einer klinisch manifesten Erkrankung nicht zu unterscheiden ist, aber weder die Lebensdauer noch die Lebensqualität beeinflusst (s.a. ‚pseudodisease‘)

recommendations for screening for colorectal cancer. It would be useful to establish a consensus about which data have to be collected to provide a reliable basis for health-care decisions.

punkte erreicht werden. Arbeiten mit einer Punktzahl von 7–9 Punkten wurden als valide, Arbeiten mit 4–6 Punkten als eingeschränkt valide und Arbeiten mit weniger als 4 Punkten als nicht valide eingestuft.

## Ergebnisse

### Empfehlungen

Alle 9 Organisationen empfehlen, bei der asymptomatischen Bevölkerung mit der Darmkrebsvorsorge ab dem 50. Lebensjahr zu beginnen. Die deutschen Organisationen empfehlen zur Früherkennung 2 Verfahren. Den chemischen Stuhltest jährlich ab dem 50. Lebensjahr und die Koloskopie im Abstand von 10 Jahren ab dem 55. Lebensjahr. Die US-amerikanischen Organisationen sprechen Empfehlungen für diverse Verfahren aus. Alle 3 amerikanischen Organisationen zählen den chemischen und immunologischen Stuhltest sowie die Sigmoidoskopie und Koloskopie zu den geeigneten Verfahren. Die American Cancer Society und das National Cancer Institute empfehlen zusätzlich den DNA-Stuhltest, den Barium-Doppelkontrasteinlauf, DCEB, und die virtuelle bzw. CT-Koloskopie. Die International Agency for Research on Cancer und der European Code against Cancer haben am Anfang des Jahres 2011 lediglich den chemischen Stuhltest im Abstand von 1–2 Jahren empfohlen. Laut IARC war die Datenlage zum FIT zu diesem Zeitpunkt nicht ausreichend.

### Referenzen und deren Validität

In 18 von 33 Referenzen, die zur Begründung einer Screening-Maßnahme aufgeführt wurden, wurde die Mortalität als Endpunkt ausgewählt; in 15 Referenzen wurden andere Endpunkte gewählt. Elf der verbleibenden 18 Veröffentlichungen untersuchten den chemischen Stuhltest [12–22], 3 die Sigmoidoskopie [23–25] und jeweils eine den immunologischen Stuhltest [26], die Koloskopie [27] oder generell die Diagnostik des Darmes [28]. Bei einer weiteren Publikation handelte es sich um eine Metaanalyse [29]. ► Tab. 3 beschreibt die Verfasser, das jeweilige Verfahren, den Studientyp und die Validität der einzelnen Veröffentlichungen. Die vergebenen Punkte bei Bewertung der Validität sind in ► Tab. 4 dargestellt. Von 9 erreichbaren Punkten wurden maximal 6 und minimal 2 Punkte erreicht. Im Median wurden 4,5

Tab. 1 Ausgewählte Empfehlungen zur Darmkrebsprävention.

| Deutsche Organisationen                            |
|--|
| Felix Burda Stiftung                               |
| Deutsche Krebshilfe                                |
| Deutsche Krebsgesellschaft (DKG)                   |
| Deutsches Krebsforschungszentrum (DKFZ)            |
| US-amerikanische Organisationen                    |
| American Cancer Society (ACS)                      |
| National Cancer Institute (NCI)                    |
| U. S. Preventive Services Task Force (USPSTF)      |
| Internationale Organisationen                      |
| European Code against Cancer                       |
| International Agency for Research on Cancer (IARC) |

**Tab. 2** Kriterien zur Beurteilung der Validität publizierter Studien.

| Nr. | Fragen zur Prüfung der Kriterien  | Antwortmöglichkeiten  |
|-----|---|---|
| 1   | Sind die Ziele der Studie und die Dimensionen benannt, in welchen die Ziele gemessen werden sollen? | Ja; Falls nicht, letzten Satz/Absatz der Einleitung zitieren. |
| 2   | Wurde das ideale Studiendesign gewählt?   | Ja; Falls nicht, weshalb?                                     |
| 3   | Sind die Risiken der Studienpopulationen ähnlich?   | Ja; Falls nicht, weshalb?                                     |
| 4   | War die Allokation zu den Studiengruppen geheim?  | Ja; Falls nicht, weshalb?                                     |
| 5   | Waren Ärzte und Patienten kontinuierlich verblindet?  | Ja; Falls nicht, weshalb?                                     |
| 6   | War die Nachbeobachtung lang genug, um aussagekräftige Endpunkte messen zu können?                  | Ja; Falls nicht, weshalb?                                     |
| 7   | Wurden alle Patienten in die Analyse einbezogen?  | Ja; Falls nicht, weshalb?                                     |
| 8   | Ist die angewandte Statistik angemessen?  | Ja; Falls nicht, weshalb?                                     |
| 9   | Wurden Interessenkonflikte ausgeschlossen?  | Ja; Falls nicht, weshalb?                                     |
| 10  | Halten Sie die Aussage der Studie für valide?   | Ja; Eingeschränkt; Nein                                       |
| 11  | Ist der Effekt der Studie klinisch relevant?  | Ja; Eingeschränkt; Nein                                       |

Dieser Fragenkatalog zur Validität wissenschaftlicher Publikationen hat sich als Leitfaden für die gemeinsame Diskussion mehrerer (idealerweise 3–4) Beurteiler bewährt, weil bei gemeinsamer Beurteilung mehr Unzulänglichkeiten als bei Einzelbeurteilung aufgedeckt werden und die Diskussion der Unzulänglichkeiten ausgewogener und weniger ermüdend ist als die solitäre Bewertung. Die Mehrfachbeurteilung/gemeinsame Diskussion erfordert aber deutlich mehr Ressourcen, die möglicherweise nur mit Unterstützung durch Medizinstudenten oder Doktoranden gewährleistet werden kann.

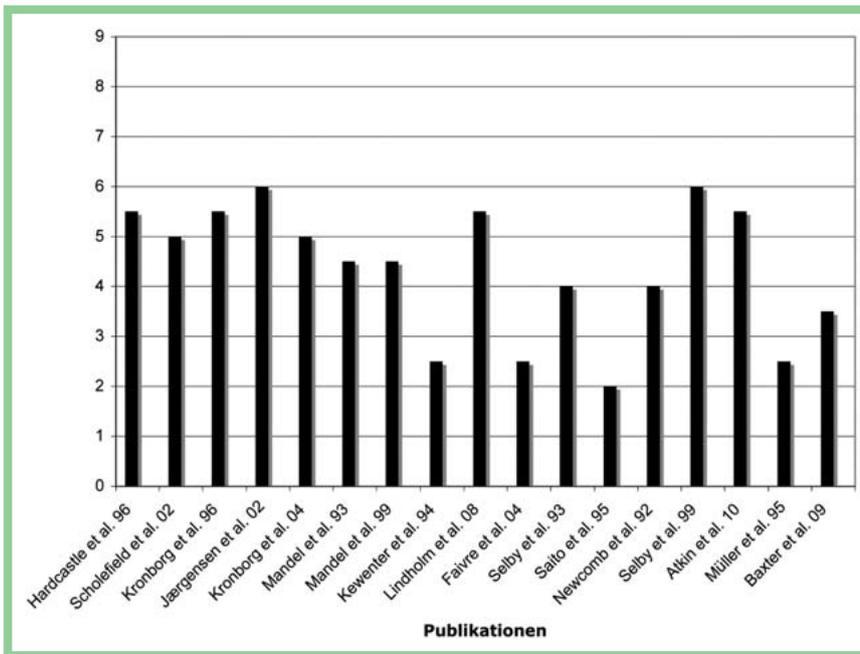
**Tab. 3** Studien, die zur Beurteilung der Validität ausgewählt wurden.

| Autoren, Studienname                           | Studientyp                               | Verfahren                       | Validität der Publikation |
|--|--|---------------------------------|---------------------------|
| Hardcastle et al. 1996, Nottingham-Studie [12] | randomisierte kontrollierte Studie (RCT) | chemischer Stuhltest (gFOBT)    | eingeschränkt valide      |
| Scholefield et al. 02, Nottingham-Studie [13]  | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Kronborg et al. 96, Fünen-Studie [14]          | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Jørgensen et al. 02, Fünen-Studie [15]         | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Kronborg et al. 04, Fünen-Studie [16]          | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Mandel et al. 93, Minnesota-Studie [17]        | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Mandel et al. 99, Minnesota-Studie [18]        | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Kewenter et al. 94, Göteborg-Studie [19]       | RCT                                      | gFOBT                           | nicht valide              |
| Lindholm et al. 08, Göteborg-Studie [20]       | RCT                                      | gFOBT                           | eingeschränkt valide      |
| Faivre et al. 04, Burgund-Studie [21]          | kontrollierte Studie                     | gFOBT                           | nicht valide              |
| Selby et al. 93 [22]                           | Fallkontrollstudie (FKS)                 | gFOBT                           | eingeschränkt valide      |
| Saito et al. 95 [26]                           | FKS                                      | immunologischer Stuhltest (FIT) | nicht valide              |
| Newcomb et al. 92 [23]                         | FKS                                      | Sigmoidoskopie                  | eingeschränkt valide      |
| Selby et al. 92 [24]                           | FKS                                      | Sigmoidoskopie                  | eingeschränkt valide      |
| Atkin et al. 10 [25]                           | RCT                                      | Sigmoidoskopie                  | eingeschränkt valide      |
| Müller et al. 95 [28]                          | FKS                                      | diverse                         | nicht valide              |
| Baxter et al. 09 [27]                          | FKS                                      | Koloskopie                      | eingeschränkt valide      |

**Tab. 4** Ergebnisse der Validitätsprüfung.

| Kriterium                    | 1    | 2   | 3    | 4    | 5   | 6   | 7   | 8   | 9   | Punkte |
|------------------------------|------|-----|------|------|-----|-----|-----|-----|-----|--------|
| Hardcastle et al. 1996 [12]  | +    | ±   | ±    | +    | ±   | +   | –   | ±   | ±   | 5,5    |
| Scholefield et al. 2002 [13] | ±    | ±   | ±    | +    | ±   | +   | –   | ±   | ±   | 5,0    |
| Kronborg et al. 1996 [14]    | +    | ±   | +    | +    | ±   | –   | ±   | ±   | ±   | 5,5    |
| Jørgensen et al. 2002 [15]   | +    | ±   | +    | +    | ±   | ±   | ±   | ±   | ±   | 6,0    |
| Kronborg et al. 2004 [16]    | –    | ±   | +    | +    | ±   | ±   | ±   | ±   | ±   | 5,0    |
| Mandel et al. 1993 [17]      | +    | –   | ±    | +    | –   | ±   | ±   | ±   | ±   | 4,5    |
| Mandel et al. 1999 [18]      | ±    | –   | ±    | +    | –   | ±   | +   | ±   | ±   | 4,5    |
| Kewenter et al. 1994 [19]    | –    | –   | ±    | +    | –   | –   | –   | ±   | ±   | 2,5    |
| Lindholm et al. 2008 [20]    | +    | –   | ±    | +    | ±   | ±   | +   | ±   | ±   | 5,5    |
| Faivre et al. 2004 [21]      | ±    | ±   | –    | –    | ±   | –   | –   | ±   | ±   | 2,5    |
| Selby et al. 1993 [22]       | ±    | ±   | ±    | –    | ±   | –   | ±   | +   | ±   | 4,0    |
| Saito et al. 1995 [26]       | ±    | ±   | ±    | –    | –   | –   | –   | ±   | –   | 2,0    |
| Newcomb et al. 1992 [23]     | ±    | ±   | ±    | –    | ±   | –   | +   | ±   | ±   | 4,0    |
| Selby et al. 1992 [24]       | ±    | ±   | +    | –    | ±   | ±   | +   | +   | +   | 6,0    |
| Atkin et al. 2010 [25]       | +    | ±   | +    | +    | ±   | ±   | –   | ±   | +   | 5,5    |
| Müller et al. 1995 [28]      | ±    | –   | +    | –    | –   | ±   | –   | –   | ±   | 2,5    |
| Baxter et al. 2009 [27]      | ±    | ±   | ±    | –    | –   | ±   | –   | ±   | +   | 3,5    |
| <b>Punkte</b>                | 10,5 | 6,0 | 11,0 | 10,0 | 5,5 | 6,5 | 6,5 | 9,0 | 9,5 |        |

+ Kriterium erfüllt, ± Kriterium teilweise erfüllt, – Kriterium nicht erfüllt



**Abb. 1** Vergebene Punkte für Validität. Maximal konnten 9 Punkte erreicht werden. Wenn ein Kriterium nur mit Einschränkung zutraf, wurde ½ Punkt vergeben.

Punkte erreicht, wobei im Durchschnitt 2 von 9 Validitätskriterien vollständig, 5 teilweise und 2 nicht erfüllt waren. In 2 Studien war keines der 9 Validitätskriterien vollständig erfüllt.

### Identifizierte Mängel

Allgemein wurden Mängel in den Bereichen Studiendesign, Verblindung, Nachverfolgungsdauer und Studienanalyse bzw. -auswertung festgestellt (siehe [Tab. 4](#)). In nur 2 der analysierten randomisierten kontrollierten Studien (eine Doppelpublikation) wurde berichtet, dass vorab eine Einverständniserklärung, ein sog. „Informed Consent“, von allen Studienteilnehmer eingeholt wurde [17, 18, 25]. Unterschiede in den Screening-Intervallen, der Anzahl der Wiederholungen der Screening-Maßnahmen, der Art der Probenbearbeitung und der Vorgehensweise nach einem positiven Testergebnis – als Hinweis auf einen Detection Bias – führten im Falle der Minnesota- und Göteborg-Studie insgesamt zu einer negativen Bewertung [17–20]. In allen analysierten Fallkontrollstudien bestand das Problem, dass nicht immer klar war, ob die Untersuchungen ohne ein bestehendes Problem – als Screening – oder bei bestehendem Problem – als Diagnostik durchgeführt wurde [22–24, 26–28]. Diese Unschärfe der Untersuchungsindikation ist als Hinweis auf einen „Sampling Bias“ zu werten. Außerdem waren die Fallzahlen in den Kontrollstudien zu klein [22–24, 26]. Sie lagen in den 4 genannten Studien zwischen 262 und 1212 Personen.

Bei allen analysierten RCTs handelte es sich mit Ausnahme der Minnesota-Studie [17, 18] um ein „asymmetrisch verblindetes“ Design [12–16, 19–21, 25]. Das bedeutet, dass nur die Screening-Gruppe über die Studie informiert wurde. Diese unterschiedliche Information erfordert die Diskussion eines möglichen „Performance Bias“.

Da die Nachverfolgungsdauer in lediglich 2 Publikationen zur Nottingham-Studie aufgrund der dargestellten Mortalitätskurven als ausreichend bewertet wurde [12, 13], ist in der Mehrzahl der Studien ein möglicher „Detection Bias“ zu bedenken. Die Auswertung der Studienergebnisse berücksichtigt in nur wenigen Fällen alle Studienteilnehmer. In der Nottingham- und Fünen-Studie wurden z. B. zunächst bzw. ausschließlich nur die Teilneh-

mer zu weiteren Screenings eingeladen, die an den vorherigen Screenings teilgenommen hatten [12–16]. Deshalb kann in diesen Studien ein „Attrition Bias“ nicht ausgeschlossen werden. Die Fragestellung der Studie wurde in 14 von 17 Fällen klar formuliert ([Tab. 4](#)).

Die einzige Veröffentlichung zum FIT [26], eine monozentrische Studie mit kleiner Fallzahl, wurde als nicht valide bewertet. Die relative Risikoreduktion in der Studie zur Sigmoidoskopie von Atkin et al. [25] ist schwer zu interpretieren, weil die Daten dieser Studie suggerieren, dass als Folge der Sigmoidoskopie auch die Rate der nicht kolorektalen Karzinome reduziert wurde. Die exakte Analyse dieser Studie zeigte, dass die erheblichen Unterschiede in den Ausgangsrisiken der experimentellen Gruppe und der Kontrollgruppe diesen unlogischen Befund erklären. Außerdem konnte in dieser Studie eine Intervention aus diagnostischen Gründen nicht ausgeschlossen werden. Das trifft auch auf die Fallkontrollstudie von Baxter et al. [27] zur Koloskopie zu. Zusätzlich bezieht sich die beobachtete relative Risikoreduktion von 31% ausschließlich auf linksseitige Darmtumoren. Die Studie von Müller et al. [28] zu verschiedenen Untersuchungsverfahren des Darmes wurde als nicht valide und nicht repräsentativ bewertet, da der Einfluss von 18 verschiedenen Diagnoseverfahren fast ausschließlich bei Männern mit einem Durchschnittsalter von 70 Jahren analysiert wurde.

Zusammenfassend wurden von den 17 analysierten Publikationen keine als uneingeschränkt valide beurteilt. In [Abb. 1](#) und [Tab. 4](#) sind die konkreten Resultate dargestellt.

### Diskussion

Voraussetzung für eine positive Bewertung der Validität einer Publikation ist eine explizite Formulierung der Fragestellung. Eine fehlende oder unklare Fragestellung führt zwangsläufig zu Schwächen der Studienmethodik. Wenn die Fragestellung einer Studie, z. B. die Göteborg-Studie [19], weder die verglichenen Gruppen noch den Zielparameter der Studie konkret benennt, kann der Leser nicht entscheiden, ob eine angemessene Methode

in der Studie gewählt wurde. Ein Teil der in dieser Arbeit eingeschlossenen Studien wurde vor etwa 20 Jahren publiziert und entsprechend früher konzipiert. Zwischenzeitlich hat sich aber auch die Klinische Epidemiologie weiterentwickelt und konkrete Vorstellungen zu den am häufigsten auftretenden Formen systematischer Fehler (Bias) entwickelt [30]. Gerade in Studien zur sekundären Prävention ist darauf zu achten, dass in der experimentellen und in der Kontrollgruppe ein Screening Bias vermieden wird, d. h., dass die Ausgangsrisiken der in diese Gruppen eingeschlossenen Patienten vergleichbar sind. Wenn nicht ausgeschlossen werden kann, dass gesundheitsbewusste Bürger, die möglicherweise generell ein geringeres Erkrankungsrisiko als andere Bürger haben, auch häufiger als Andere an Screening-Programmen teilnehmen („healthy screenee bias“), kann das Ergebnis solcher Studien durch den Selection Bias verzerrt sein. In der Studie von Atkin [25] kann dieser Selection Bias bereits anhand der publizierten Daten nachvollzogen werden. Ein Performance Bias ist in Erwägung zu ziehen, wenn die beiden miteinander verglichenen Gruppen – mit Ausnahme der geprüften Intervention – nicht gleich behandelt werden. Eine ungleiche Behandlung ist immer in Betracht zu ziehen, wenn die Testpersonen der beiden Gruppen unterschiedlich informiert werden, z. B. zum Screening eingeladen bzw. nicht eingeladen werden. Die wissenschaftliche Evidenz zu dieser unerwartet stark ausgeprägten „Macht der Information“ haben wir an anderer Stelle zusammengestellt [31]. Ein „Attrition Bias“ tritt auf, wenn im Rahmen einer Studie aus den verglichenen Gruppen Teilnehmer mit unterschiedlichen Ausgangsrisiken ausscheiden. So ist anzunehmen, dass die Durchführung mehrerer Screening-Runden vom Gesundheitsbewusstsein der Bürger und/oder von den Ergebnissen vorausgegangener Screening-Untersuchungen abhängig ist. Der Detection Bias bezeichnet systematische Fehler, die entstehen, wenn diagnostische oder Screening-Methoden in den verglichenen Gruppen unterschiedlich häufig oder mit unterschiedlichen Methoden (Sensitivität/Spezifität) durchgeführt werden. Diese 4 Formen von Bias betreffen die interne Validität von Studien. Die externe Validität kann durch den sogenannten Sampling Bias beeinträchtigt werden. Dieser Bias liegt vor, wenn sich die Population, an der die Studie durchgeführt wird, von der Grundgesamtheit, auf die das Ergebnis der Studie angewandt werden soll, unterscheidet. Dieser Sampling Bias tritt in sehr vielen randomisierten Studien auf, in welchen das durch die Einschlusskriterien grob definierte Studienkollektiv durch eine Vielzahl von Ausschlusskriterien präzisiert ist. Diese Präzisierung ist für die randomisierte Studie erwünscht und vorteilhaft, weil die Chance steigt, dass die verglichenen Gruppen tatsächlich ähnliche Ausgangsrisiken aufweisen. Dieser Vorteil kann aber zum Verhängnis werden, wenn das Ergebnis, das an einer eng definierten Population gewonnen wurde, nicht nur für diese eng begrenzte Population, sondern für alle Mitglieder der Grundgesamtheit als gültig erklärt wird. Bei der Sekundärprävention ist dieser Sampling Bias (Vergleichbarkeit der Studiengruppen mit der Grundgesamtheit) und der Selection Bias (Vergleichbarkeit der in die Studie eingeschlossenen Gruppen untereinander) auszuschließen. Wenn diese potenziellen Störfaktoren nicht ausgeschlossen werden, steigt das Risiko, Gesundheitsleistungen zu erbringen, die zwar unserer Erwartungshaltung entsprechen, einer kritischen Betrachtung aber nicht standhalten. Wenn in einer Studie diese Kriterien der Validität zutreffen, bleibt zu klären, ob der beschriebene Effekt aus der Sicht der Patienten und der behandelnden Ärzte den Aufwand und die potenziellen Nachteile und Risiken der Intervention aufwiegt. Wir

bezeichnen dieses subjektive Urteil als Bewertung der klinischen Relevanz. Natürlich macht es nur Sinn, die klinische Relevanz zu bewerten, wenn tatsächlich ein statistisch signifikanter Unterschied nachgewiesen ist. So führt der gFOBT zwar zu relativen Risikoreduktionen zwischen 11–33%. Die korrespondierenden absoluten Risikoreduktionen betragen nur 0,01–0,02/100 Personennjahre. Das entspricht einer Number needed to screen (NNS) von etwa 500–1000, was bedeutet, dass ein Screening bei 500 bis 1000 Patienten durchzuführen ist, um durch das Screening bei einem zusätzlichen Patienten den Tod infolge eines Darmkrebses zu verhindern. Da in keiner Studie eine Reduktion der Gesamtsterblichkeit beobachtet wurde, kann nicht beurteilt werden, ob lediglich die anderen Todesursachen um ein Vielfaches häufiger sind oder ob andere Ursachen, z. B. Unterschiede im Lebensstil von Teilnehmern und Nichtteilnehmern an Screening-Programmen das Ergebnis beeinflussen. Zu einem ähnlichen Ergebnis kommt ein Review der Cochrane Collaboration [29]. Somit ist der potenzielle Mehrwert, der durch das kolorektale Screening erzielt werden kann, sicher nur marginal.

Bemerkenswert ist, dass in der EU-Leitlinie zur Qualitätssicherung von Darmkrebs-Screening und -diagnose [32] und die Empfehlung des chemischen Stuhltests auf Arbeiten basiert, die in unserer Studie im Detail analysiert wurden [12–21]. Kritische Aspekte wie Schwächen im Studiendesign, z. B. ein über das Einladungsverfahren provozierter Bias, werden in den Leitlinien nicht thematisiert. Die Evidenz zum immunologischen Stuhltest wird als angemessen beschrieben, beruht aber auf einem RCT, in dem lediglich eine Reduzierung der Mortalität an rektalem Krebs, aber nicht an Darmkrebs insgesamt beobachtet wurde [33]. Auch die Evidenz zur Wirksamkeit der Sigmoidoskopie wird als angemessen beurteilt. Die Leitlinien berufen sich auf das hier analysierte RCT von Atkin et al. [25], in dem ein Selection Bias vorliegt, und auf 2 weitere RCTs, in denen die beobachtete Risikoreduktion nicht signifikant war [34, 35]. Zusätzlich wurden die Ergebnisse der Fallkontrollstudien von Newcomb et al., Selby et al. und Müller et al. in die Bewertung mit einbezogen. Die Veröffentlichungen zu diesen Studien wurden im Rahmen dieser Arbeit als eingeschränkt valide bzw. nicht valide und nicht klinisch relevant beurteilt [23, 24, 28]. Die wissenschaftliche Datenlage zur Koloskopie schließlich wird in den EU-Leitlinien als unzureichend beschrieben [32].

Laut Nationalem Krebsplan [36] sind niedergelassene Ärzte in Deutschland zurzeit aber nicht dazu verpflichtet, die korrekte Ausführung und die Ergebnisse des gFOBT zu dokumentieren. Eine Evaluierung der Maßnahme ist somit nicht möglich. Auch die Evaluierung des Koloskopie-Screenings gestaltet sich in Deutschland schwierig, da derzeit keine vollständige Erfassung der Koloskopien erfolgt und keine zuverlässigen Daten zur regionalen Darmkrebsinzidenz und -mortalität vorliegen. Die Festlegung auf 2 Methoden zur Früherkennung von Darmkrebs in Deutschland könnte im solidarisch organisierten Krankenkassensystem seine Begründung finden, da für neuere Verfahren zu wenig aussagekräftige Daten vorliegen. Auch die als Goldstandard bezeichnete Koloskopie ist nicht unproblematisch, da flache Läsionen und rechtsseitige Neoplasien häufig übersehen werden und das Auftreten von schnell wachsenden Intervallkarzinomen sowie eine Zunahme von Komplikationen mit einer Ausweitung der Screening-Koloskopien nicht auszuschließen sind [37]. Weitere wichtige Erkenntnisse hierzu werden durch die Koloskopie-Studien in Spanien und Nordeuropa [38, 39] beigetragen, deren Ergebnisse im Jahr 2022 erwartet werden.

Nach unserem Verständnis von Versorgungsforschung hat dieses neue Arbeitsgebiet die Frage zu beantworten, ob Verfahren zur Prävention, Diagnostik und Therapie, die von der theoretischen oder klinischen Forschung unter Idealbedingungen untersucht wurden und für die Anwendung im Versorgungsalltag empfohlen werden, auch das halten, was sie versprechen. Neben der Analyse publizierter Daten, wie sie hier vorgenommen wurde, stellt die Verifizierung von „efficacy“-Daten unter „effectiveness“-Bedingungen eine der großen Herausforderungen der Versorgungsforschung dar. Eine Sammlung von Koloskopiedaten wird nur auswertbar sein, wenn vorab ein Konsens erzielt wird, welche Daten für eine verlässliche Interpretation gefordert werden. Beispiele dieser Forderung sind Daten zur Indikation der Koloskopie (Unterscheidung von Screening ohne Symptome von der Diagnostik wegen bestehender Symptome), zum Ausgangsrisiko (eigene und familiäre Risikofaktoren) oder zur Qualität der Untersuchungsergebnisse (Plausibilitätskontrollen zur Häufigkeit von Befunden, histologische Bestätigung).

Bei der Bewertung von Therapien werden hohe Anforderungen an die Validität der zugrunde liegenden Daten gestellt; im Bereich der Diagnostik werden seit einigen Jahren die Kriterien diskutiert, anhand welcher klinisch bedeutende Diagnoseverfahren von überflüssigen Testen unterschieden werden können. Was für die Therapie und die Diagnostik gilt, sollte auch für die Prävention zutreffen. Das hier diskutierte Beispiel der sekundären Prävention des Kolonkarzinoms zeigt, dass die Validität der vorliegenden Daten – gemessen an der geforderten Validität für therapeutische Interventionen – nicht ausreicht, um wissenschaftlich fundierte Schlussfolgerungen abzuleiten. Wegen dieser Mängel ist es wesentlich wahrscheinlicher, dass wir in Deutschland den Nutzen der Koloskopie überschätzen als dass andere Länder, z. B. Großbritannien oder Frankreich den tatsächlichen Nutzen unterschätzen.

Als eines der wirtschaftlich stabilen europäischen Länder sollten wir uns verpflichtet fühlen, diese kostenintensive Herausforderung anzunehmen, um zu gewährleisten, dass die langfristig etablierten Gesundheitsleistungen deren finanziellen Aufwand rechtfertigen. Dieses Ziel können wir nur erreichen, wenn wir unsere eigenen Leistungen kritisch hinterfragen und identifizierte Schwachstellen durch bessere Lösungen ersetzen. Solche Lösungen lassen sich sicher nicht kurzfristig umsetzen; die Diskussion, wie diese Lösungen zu gestalten sind, könnten aber kurzfristig eingeleitet werden.

**Interessenkonflikt:** Nein

## Literatur

- 1 Mathers C, Fat DM, Boerma JT. The global burden of disease: update 2004. Genf: World Health Organization; 2008
- 2 Karsa von L, Lignini TA, Patnick J et al. The dimension of the CRC problem. *Best Pract Res Clin Gastroenterol* 2010; 24: 381–396
- 3 National Cancer Institute. Im Internet: <http://www.cancer.gov/cancer-topics/types/commoncancers>; Stand: 15.05.2012
- 4 Husmann G, Kaatsch P, Katalinic A et al. Darm. In: Robert-Koch-Institut und Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., Hrsg. Krebs in Deutschland 2005/2006. Häufigkeiten und Trends. 7. Aufl. Berlin 2010; 36–39
- 5 Moyer VA. What We Don't Know Can Hurt Our Patients: Physician Innumeracy and the overuse of Screening Tests. *Ann Intern Med* 2012; 156: 392–393
- 6 Gemeinsamer Bundesausschuss. Richtlinie des Gemeinsamen Bundesausschusses über die Früherkennung von Krebserkrankungen. *Bundesanzeiger* 2009; 148a: 1–48
- 7 Maniapan S, Garrido MV. Surrogatparameter als Parameter der Nutzenbewertung. In: Deutsches Institut für Medizinische Dokumentation und Information, Hrsg. Schriftenreihe Health Technology Assessment (HTA) in der Bundesrepublik Deutschland. Bd 91. 1. Aufl. Berlin: DIM-DI; 2009
- 8 Wegwarth O, Schwartz LM, Woloshin S et al. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012; 156: 340–349
- 9 Porzolt F, Bonotto de O, Costa IC and Thomaz TG for the SHUFFLE group. Advantages and Limitations of Twin Assessment of Clinical Trials (TACT). *J Publ Health* 2009; 17: 425–435
- 10 Rosati P, Porzolt F. A practical educational tool for teaching child-care hospital professionals attending evidence-based practice courses for continuing medical education to appraise internal validity in systematic reviews. *J Eval Clin Pract* 2012 Jul 29 [epub ahead of print]
- 11 Braubach P, Porzolt F. Messung und Bewertung der Validität. In: Porzolt F, Hrsg. Grundlagen der Klinischen Ökonomik. Berlin: PVS Verlag; 2011: 133–134
- 12 Hardcastle J, Chamberlain J, Robinson M et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996; 348: 1472–1477
- 13 Scholefield JH, Moss S, Sufi F et al. Effect of faecal occult blood screening on mortality from colorectal cancer: results from a randomised controlled trial. *Gut* 2002; 50: 840–844
- 14 Kronborg O, Olsen J, Søndergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996; 348: 1467–1471
- 15 Jørgensen OD, Kronborg O, Fenger C. A randomised study for colorectal cancer using faecal occult blood testing: results after 13 years and seven biennial screening rounds. *Gut* 2002; 50: 29–32
- 16 Kronborg O, Jørgensen OD, Fenger C et al. Randomised study of biennial screening with a faecal occult blood test: results after nine screening rounds. *Scand J Gastroenterol* 2004; 9: 846–851
- 17 Mandel JS, Bond JH, Church TR et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *Minnesota Colon Cancer Control Study*. *N Engl J Med* 1993; 328: 1365–1371
- 18 Mandel JS, Church TR, Ederer F et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst* 1999; 91: 434–437
- 19 Kewenter J, Brevinge H, Engarås B et al. Results of screening, rescreening, and follow-up in a prospective randomised study for detection of colorectal cancer by fecal occult blood testing. Results for 68,308 subjects. *Scand J Gastroenterol* 1994; 29: 468–473
- 20 Lindholm E, Brevinge H, Haglund E. Survival benefit in a randomised clinical trial of faecal occult blood screening for colorectal cancer. *Br J Surg* 2008; 95: 1029–1036
- 21 Faivre J, Dancourt V, Lejeune C et al. Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. *Gastroenterology* 2004; 126: 1674–1680
- 22 Selby JV, Friedman GD, Quesenberry CP et al. Effect of fecal occult blood testing on mortality from colorectal cancer. *Ann Intern Med* 1993; 118: 1–6
- 23 Newcomb PA, Norfleet RG, Storer BE et al. Screening sigmoidoscopy and colorectal cancer mortality. *J Natl Cancer Inst* 1992; 84: 1572–1575
- 24 Selby JV, Friedman GD, Quesenberry CP et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992; 326: 653–657
- 25 Atkin WS, Edwards R, Kralj-Hans I et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010; 375: 1624–1633
- 26 Saito H, Koeda J, Wada T et al. Reduction in risk of mortality from colorectal cancer by fecal occult blood screening with immunochemical hemagglutination test. A case-control study. *Int J Cancer* 1995; 61: 465–469
- 27 Baxter NN, Goldwasser MA, Paszat LF et al. Association of colonoscopy and death from colorectal cancer. *Ann Intern Med* 2009; 150: 1–8
- 28 Müller AD, Sonnenberg A. Protection by endoscopy against death from colorectal cancer. A case-control study among veterans. *Arch Intern Med* 1995; 155: 1742–1748
- 29 Hewitson P, Glasziou P, Irwig L et al. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev* 2007; 1: CD001216
- 30 Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42–46
- 31 Clouth J, Porzolt F. Praktische Effekte zur Steuerung des Gesundheitssystems. In: Porzolt F, Hrsg. Grundlagen der Klinischen Ökonomik. Berlin: PVS Verlag; 2011, 85–89

- 32 *European Commission*. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition. Im Internet: <http://screening.iarc.fr/doc/ND3210390ENC.pdf>; Stand 05.04.2012
- 33 *Zheng S, Chen K, Liu X et al*. Cluster randomisation trial of sequence mass screening of colorectal cancer. *Dis Colon Rectum* 2003; 46: 51–58
- 34 *Thiis-Evensen E, Hoff GS, Sauar J et al*. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. Telemark Polyp Study I. *Scand J Gastroenterol* 1999; 34: 414–420
- 35 *Hoff G, Grotmol T, Skovlund E et al.; for the Norwegian Colorectal Cancer Prevention Group*. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised trial. *BMJ* 2009; 338: B1846
- 36 *Bundesministerium für Gesundheit*. Im Internet: <http://www.bmg.bund.de/praevention/nationaler-krebsplan/was-haben-wir-bisher-erreicht/ziel-2b-weiterentwicklung-der-darmkrebsfrueherkennung.html>; Stand: 04.04.2012
- 37 *Ransohoff DF*. Have we oversold colonoscopy? *Gastroenterology* 2005; 129: 1815
- 38 *Colorectal Cancer Screening in Average-risk Population: Immunochemical Fecal Occult Blood Testing Versus Colonoscopy*. Im Internet: <http://www.clinicaltrials.gov/ct2/show/NCT00906997>; Stand: 04.04.2012
- 39 *The Northern-European Initiative on Colorectal Cancer (NordICC)*. Im Internet: <http://clinicaltrials.gov/ct2/show/NCT00883792>; Stand: 04.04.2012