



Is everything we eat associated with cancer? A systematic cookbook review

Schoenfeld Jonathan D, Ioannidis John PA. *The American Journal of Clinical Nutrition*, 97:127-134 (2013)

Hintergrund Ernährung ist und bleibt ein kontroverses Thema, zu dem jeder etwas zu sagen zu haben scheint und an bestimmte Theorien glaubt. Warum ist das so? Sicher hat die Ernährungswissenschaft gegenüber reinen Naturwissenschaften den Nachteil, experimentell keine exakten Gesetze ableiten zu können, da die Bedingungen eines Experimentes am Menschen nie die gleichen sein können und daher nicht exakt reproduzierbar sind. Weiterhin müssen viele Hypothesen aus Beobachtungen generiert werden, wobei es sehr schwierig ist, andere Einflüsse auf die zu untersuchenden Zusammenhänge zu erkennen und zu berücksichtigen. Allerdings haben es sich die Ernährungswissenschaftler auch selbst zuzuschreiben, dass so viele Halbwahrheiten über ihre „Wissenschaft“ kursieren. Ein gutes Beispiel liefern Schoenfeld und Ioannidis mit dieser Arbeit, in der sie zeigen, dass die Einflüsse einzelner Nahrungsmittel auf das Krebsrisiko zu oft überinterpretiert werden, und zwar in einem Ausmaß welches meiner Meinung nach die gesamte Literatur zu diesem Thema zweifelhaft, wenn nicht lächerlich erscheinen lässt.

Material und Methoden Per Zufallsgenerator wurden einzelne Seiten des Kochbuches *The Boston Cooking-School Cook Book* ausgewählt und alle Zutaten des ersten Rezeptes auf der ausgewählten Seite in die Liste der zu untersuchenden Zutaten übernommen, solange bis 50 individuelle

Zutaten erreicht waren. Für jede Zutat wurde dann in der Online-Datenbank PubMed nach epidemiologischen Studien gesucht, die einen Zusammenhang dieser Zutat mit der Entstehung von Krebs untersuchten. Es wurde nach Möglichkeit versucht, für jede Zutat die zehn neusten Studien zu finden. Zusätzlich wurde für jede Zutat nach der aktuellsten Metaanalyse gesucht, die einen Zusammenhang mit dem Krebsrisiko herstellte.

Aus all diesen Studien und Metaanalysen wurden dann folgende Daten extrahiert: Die Zutat, die untersuchte Krebsart, die Einschätzung der Studienautoren in Bezug auf das Krebsrisiko (erhöht, erniedrigt, kein Effekt oder grenzwertiger Effekt), das Relative Risiko als Effektgröße (RR; *Hazard Ratio* für Kohortenstudien oder *Odds Ratio* für Fall-Kontrollstudien), der zugehörige p-Wert sowie der Kontrast zwischen zwei Konsummengen auf den sich das RR bezog (z.B. mehr als fünf Tassen pro Tag gegenüber höchstens eine Tasse pro Tag). Die p-Werte wurden entsprechend einer Normalverteilung in zugehörige z-Scores umgewandelt¹ und daraus eine Verteilung erstellt. Die z-Score Verteilung der Einzelstudien wurde mit derjenigen der Metaanalysen verglichen, um herauszufinden ob mögliche Effekte einer bestimmten Zutat auf das Krebsrisiko in Metaanalysen anders ausfallen als in den Originalstudien. Mit Hilfe des χ^2 Testes wurde untersucht, ob die p-Werte, die in den Abstracts erwähnt wurden, sich von

¹Für eine gegebene Wahrscheinlichkeit p ergibt sich der z-Score als die Anzahl der Standardabweichungen σ , die man vom Mittelwert entfernt liegt, so dass die Fläche unter der Gaußkurve zwischen $-z\sigma$ und $z\sigma$ gleich $1-p$ ist. So entspricht z.B. $p=0,05$ einem z-Score von 2, da die Fläche der Gaußkurve zwischen -2σ und 2σ ungefähr gleich 0,95 ist. Negative bzw. positive z-Scores entsprechen dabei p-Werten, die einem erniedrigtem bzw. erhöhtem Risiko zugeordnet sind.

denen, die nur im vollständigen Text aufgeführt waren, unterschieden.

Ergebnis Für 40 der 50 Zutaten (80%) wurde mindestens eine Studie gefunden, die einen Zusammenhang zwischen der betreffenden Zutat und einem Krebsrisiko untersuchte. Dies waren folgende Zutaten bzw. Nahrungsmittel²: Salz, schwarzer Pfeffer, Mehl, Eier, Brot, Schwein, Butter, Tomaten, Kalb, Zitrone, Ente, Zwiebel, Sellerie, Karotte, Petersilie, Muskat, Sherry, Oliven, Pilze, Milch, Käse, Kaffee, Speck, Hummer, Zucker, Kartoffel, Rind, Lamm, Senf, Nüsse, Wein, Erbsen, Mais, Zimt, Cayenne Pfeffer, Orangen, Tee, Rum und Rosinen. 39% der Studien fanden ein erhöhtes Krebsrisiko für die betreffende Zutat, 33% ein erniedrigtes Risiko, 5% einen grenzwertigen statistisch signifikanten Effekt und 23% ergaben keinen klaren Einfluss auf das Krebsrisiko. Interessanterweise gab es für 36 der 40 oben genannten Zutaten mindestens eine Studie, die ein erhöhtes oder erniedrigtes Krebsrisiko festgestellt hatte. Allerdings war die Glaubwürdigkeit dieser Aussagen wenig überzeugend: in 75% bzw. 76% der Studien, welche ein erhöhtes bzw. erniedrigtes Risiko festgestellt hatten, waren die Effekte statistisch gesehen entweder nicht nominell signifikant ($p > 0,05$) oder nur schwach ($0,001 < p < 0,05$); und das obwohl 65% der Studien Extremwerte des Konsums einer bestimmten Zutat miteinander verglichen, um ein relatives Risiko zu ermitteln (z.B. größtes gegenüber kleinstes Quintil oder mehr als fünf Tassen gegenüber höchstens einer Tasse am Tag). Generall wurde eine sehr große Uneinheitlichkeit unter den Studien festgestellt was diese Unterteilung der Konsummengen einer bestimmten Zutat betraf, so dass der Vergleich einzelner Studien untereinander nicht trivial ist. Die meisten Metaanalysen benutzten die Risikoabschätzungen zwischen den Gruppen mit den geringsten gegenüber den höchsten Einnahmemengen um ein RR aus mehreren Studien zu ermitteln. Die Ergebnisse

der Metaanalysen fielen dabei ernüchternd aus, d.h. beim Zusammenwürfeln der Daten aus den Originalstudien wurden die Effekte einzelner Zutaten auf das Krebsrisiko kleiner: Während das RR in den Studien, die ein erhöhtes bzw. erniedrigtes Krebsrisiko fanden, im Median 2,20 (Spanne von 1,60–3.44) bzw. 0.52 (0.39–0.66) betrug, war der zugehörige Effekt in den Metaanalysen im Schnitt Null (medianes RR 0,96, Spanne 0,85–1,10).

Interessanterweise waren die den Effektgrößen einzelner Zutaten zugehörigen z-Scores bimodal (also mit zwei Spitzen) verteilt, während Werte um 0 kaum vorkamen; dies galt sowohl für die Originalstudien als auch die Metaanalysen. Das bedeutet, es wurden hauptsächlich solche Ergebnisse berichtet, die auf ein signifikant erhöhtes (positive z-Scores) bzw. erniedrigtes (negative z-Scores) Risiko hindeuteten. Wurde kein signifikanter Effekt gefunden, fand dies meist nur im Haupttext, nicht aber im Abstract der Studien Erwähnung. Dieser Unterschied in der Berichterstattung der Effektgrößen zwischen Abstract und Volltext war hoch signifikant ($p < 0,0001$).

Sponsoren Keine.

Kommentar An sich kann man schon darüber schmunzeln, wenn man hier erfährt was denn so alles von dem, was auf unseren Tellern landet, vor Krebs schützt oder ihn fördert – zumindest wenn man den publizierten wissenschaftlichen Studien glaubt. Dabei ging es Schoenfeld und Ioannidis nicht darum, uns Angst davor zu machen, morgens unseren Speck zu essen, sondern im Gegenteil Skepsis bei der Bewertung solcher epidemiologischen Studien zu erzeugen. Das muss nicht mal heißen, dass diese Studien von schlechter Qualität sind. Das Problem ist eher die Interpretation der Ergebnisse und deren Überbewertung durch die Studienautoren. Was von den Medien gerne aufgegriffen und unter das Volk verteilt wird. Dass p-Werte falsch bewertet werden,

²Man sollte sich diese Liste gedanklich auf der Zunge zergehen lassen

ist ein weit verbreitetes Problem und liegt auch an der willkürlichen Schwelle von 0,05 für „Signifikanz“. So wurden sehr oft grenzwertige ($0,05 < p < 0,1$) oder nur schwache ($0,001 < p < 0,05$) Ergebnisse von den Studienautoren als Hinweis auf eine Dosis-Effekt-Beziehung interpretiert. Dieses Problem wird weiter verschärft durch die selektive Wiedergabe von scheinbar signifikanten Ergebnissen im Abstract der Artikel, während weniger deutliche Ergebnisse meist nur im Haupttext erwähnt werden. Der Leser, der sich für die Effekte eines bestimmten Lebensmittels interessiert, wird bei der Literaturrecherche so leichter den Eindruck erhalten, dass es tatsächlich statistisch signifikante Effekte gibt.

Ein weiteres Problem, das generell die gesamte biomedizinische Literatur betrifft, ist die selektive Publikation von positiven Resultaten. Nullergebnisse, die auf keine Effekte hinweisen, werden allzu oft garnicht nicht erst zur Publikation eingereicht. Wenn aus diesem Grund hauptsächlich signifikante Ergebnisse publiziert werden, gibt es zwei Auswirkungen: Erstens könnten die Ergebnisse rein zufällig signifikant ausgefallen sein und würden dann in nachfolgenden Studien nicht repliziert werden. In der Tat kann die Mehrzahl der veröffentlichten medizinischen Studienergebnisse nicht unabhängig bestätigt werden (Ioannidis et al., 2014). Würden auch Nullergebnisse publiziert, so würde man bei der Interpretation eines schwachen Effektes wahrscheinlich vorsichtiger sein. Zweitens könnten Metaanalysen Nullresultate, die nie publiziert wurden, nicht berücksichtigen und verfälschte Schlüsse ziehen.

Allein der gesunde Menschenverstand

müsste ausreichen, um bei einer potentiellen Krebsrisikoerhöhung oder -senkung durch ein einzelnes Nahrungsmittel skeptisch zu sein. Zunächst mal müsste man überlegen, ob es tatsächlich biochemische Mechanismen für dessen Wirkung gibt oder geben könnte, und ob man wirklich das Zusammenwirken mit anderen Nahrungsmitteln ausschließen kann. Da letzteres in Beobachtungsstudien praktisch nie erfüllt ist und reale Effekte oft sehr klein und deshalb schwierig zu erfassen sind, kann man davon ausgehen, dass Studienresultate oft durch diese und andere Fehlerquellen verfälscht sind. Um das sogenannte “Effect-to-bias ratio” künstlich zu erhöhen, benutzen Forscher oft Tricks im Studiendesign, die die Relevanz der Studienergebnisse für das reale Leben zweifelhaft machen (Ioannidis et al., 2014). Ein gutes Beispiel, das von Schoenfeld und Ioannidis herausgestellt wurde, ist der Vergleich zweier Extreme, um einen Effekt zu berechnen. Wie soll man eine Risikobewertung interpretieren, die auf dem Vergleich zwischen Personen, die keinen Kaffee trinken, und solchen, die mehr als fünf Tassen am Tag trinken, beruht?

Es ist klar, dass in der biomedizinischen Literatur viel Müll publiziert wird – diesen zu reduzieren ist ein dringliches Ziel, dem die renommierte Zeitschrift *The Lancet* kürzlich sogar eine ganze Artikelserie gewidmet hat (Macleod et al., 2014). Mittlerweile existiert ein ganz neues Forschungsgebiet zum Thema “Research of research”, dessen Ziel es ist, die Missstände aufzudecken und Lösungen für mehr Wissensfortschritt und weniger „Müll“ zu finden. Das Paper von Schoenfeld und Ioannidis trägt sicherlich seinen Teil dazu bei.

Literatur

Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*, 383 (9912):166-175 (2014)

Macleod MR, Michie S, Roberts I, et al. Biomedical Research: increasing value, reducing waste. *Lancet*, 383 (9912):101-104 (2014)