

Hinweis: Dieses Dokument basiert auf dem Aufsatz „Evaluating Research“ von Professor Alex Edmans, London School of Economics. Den Originalaufsatz finden Sie [hier](#). Der Text wurde stellenweise an den deutschen Kontext angepasst. Die Verantwortung für die deutsche Version liegt bei Professor Alexandra Niessen-Ruenzi, Lehrstuhl für Corporate Governance, Universität Mannheim.

Forschung richtig bewerten

Eine Handreichung für Journalist*innen und
Autor*innen

Autor*innen und Journalist*innen zitieren häufig forschungsbasierte Studien, um ein Argument zu untermauern. Dies ist eine willkommene Entwicklung in einer Zeit, in der wissenschaftlichen Fakten zuweilen wenig Beachtung geschenkt wird. Allerdings ist hierbei zu beachten, dass die Qualität forschungsbasierter Studien einer großen Bandbreite unterliegt:

- Einige Studien beinhalten grundlegende methodische Fehler. Dies wäre zum Beispiel der Fall, wenn eine Studie bei der Berechnung von Aktienrenditen Dividendenzahlungen ignoriert.
- Einige Studien interpretieren die Daten falsch. So untersuchen beispielsweise viele Studien den Zusammenhang zwischen zwei Faktoren (A und B), z. B. Ungleichheit in einem Land und Gesundheit der Bevölkerung, und finden eine *Korrelation* zwischen den beiden: Länder, in denen die Ungleichheit hoch ist, haben auch eine weniger gesunde Bevölkerung. Es wird allerdings statt einer *Korrelation* eine *Kausalität* von A nach B behauptet - Ungleichheit verursacht eine schlechte Gesundheit, und deshalb sollte ein Land die Ungleichheit bekämpfen, um seine Gesundheit zu verbessern. Bekanntlich *impliziert die Korrelation jedoch keine Kausalität*, und zwar aus zwei Gründen:
 - Es kann eine *umgekehrte Kausalität* geben - die Kausalität kann in die andere Richtung verlaufen. Es kann zum Beispiel sein, dass Krankheit Ungleichheit verursacht, da sie die Menschen davon abhält, arbeiten zu gehen. Wenn dies jedoch der Fall ist, besteht die Lösung darin, Gesundheitsthemen direkt anzugehen und nicht das Problem der Ungleichheit.
 - Ebenso kann es *nicht in die Analyse einbezogene Variablen* geben, die sowohl A als auch B beeinflussen. Demokratisch gewählte Regierungen können zum Beispiel sowohl Ungleichheit reduzieren als auch die Gesundheit verbessern. In diesem Fall, bestände die Lösung darin, die Demokratie zu stärken bzw. zu verbessern – und damit die Grundursache der Ungleichheit zu behandeln - und nicht die Ungleichheit selbst, die nur ein Symptom des Ganzen ist.
- Einige Studien stellen auf Basis ihrer Ergebnisse Behauptungen auf, die von den Daten tatsächlich nicht gestützt werden. Malcolm Gladwells berühmte "10.000-Stunden-Regel"

besagt zum Beispiel, dass man in jedem kognitiv komplexen Bereich, vom Schachspielen bis zur Neurochirurgie, ein außerordentlich konsistentes Muster findet, das zeigt, dass man in einer Sache erst gut sein kann, wenn man 10.000 Stunden geübt hat. Die Studie, auf die sich seine Behauptung stützt, untersuchte jedoch nur Geigenspieler.¹ Die Ergebnisse gelten daher möglicherweise nicht für Schachspieler oder das Gebiet der Neurochirurgie.

Diese Unterscheidungen sind wichtig. Viele Leser*innen werden bereitwillig einen Artikel teilen, der ihre eigene Sicht der Welt bestätigt, selbst wenn die Forschung, die zitiert wird, fehlerhaft oder nicht verlässlich ist. Das wiederum kann zur Verbreitung von Fehlinformationen führen.

So wurde zum Beispiel die Anti-Impf-Bewegung tatsächlich auf fadenscheinigen Forschungen begründet, in denen behauptet wurde, dass Impfungen Autismus verursachen, ebenso wie der Vandalismus von 5G-Masten durch unbestätigte Geschichten ausgelöst wurde, dass diese das Coronavirus verschlimmern. Darüber hinaus müssen Zeitungen regelmäßig Widerrufe drucken, weil sie Forschungsergebnisse falsch zitiert oder sich auf Forschungsarbeiten minderer Qualität berufen haben.

Eine große Herausforderung besteht darin, dass Autor*innen und Journalist*innen häufig nicht die Zeit oder das Fachwissen haben, um sich in jede Arbeit zu vertiefen und deren Methodik zu überprüfen. Der Zweck des vorliegenden Textes ist es daher, einen kurzen Leitfaden zur Verfügung zu stellen, anhand dessen Sie erkennen können, ob Sie den jeweiligen Forschungsergebnissen vertrauen können. Der Leitfaden beinhaltet sechs einfache und schnelle Überprüfungen:

1. Gibt es die Forschung wirklich?

Dies mag eine offensichtliche Frage sein, aber es hat mehrere Fälle gegeben, in denen Journalist*innen Artikel auf der Grundlage einer Pressemitteilung der Autor*innen geschrieben haben, obwohl die eigentliche Studie noch nicht existierte. Darüber hinaus sollte eine Studie öffentlich zugänglich sein und nicht zum Verkauf stehen oder nur auf Anfrage bei den Autoren erhältlich sein. Eine Studie nicht öffentlich zugänglich zu machen, kann bedeuten, dass die Autoren nicht gewillt sind, sich der Überprüfung durch Experten zu stellen.

Manchmal ist eine Studie verfügbar, aber in einer verkürzten Version - sie beschreibt beispielsweise nur die Ergebnisse, aber nicht die zugrunde liegende Methodik. Um ein Beispiel zu nennen, sollte eine Studie, die behauptet, dass "ethische Unternehmen bessere Ergebnisse liefern", erklären, welche Maßstäbe angelegt werden, um ein Unternehmen als "ethisch" zu bezeichnen. Inwieweit wir der Behauptung trauen können, hängt entscheidend davon ab, wie zuverlässig die Messung der Ethik ist, daher ist es entscheidend, die Methodik zur Verfügung zu stellen, damit diese überprüft werden kann.

Wichtig ist jedoch im ersten Schritt nur, dass ein/e Journalist*in prüft, ob die Studie an sich existiert und ob es einen entsprechenden Abschnitt zur Methodik gibt. Er/Sie muss die Studie nicht im Detail lesen.

¹ Ericsson, K. Andres, Ralf Th. Krampe, and Clemens Tesch-Romer (1993): "The Role of Deliberate Practice in the Acquisition of Expert Performance." *Psychological Review* 100, 363-406

2. Basiert die Forschung auf tatsächlichen Daten?

Auch dies mag eine offensichtliche Frage sein, aber einige Studien basieren nicht auf tatsächlichen Daten, sondern auf einfachen Meinungsumfragen. Zum Beispiel kann eine Studie berichten, dass "70% der Vorstandsmitglieder denken, dass CEOs die Unternehmensleistung nicht verbessern." Diese Aussage unterscheidet sich jedoch erheblich davon, eine solche Frage mit tatsächlichen Daten zu untersuchen – also zu erforschen, ob CEOs tatsächlich die Leistung eines Unternehmens verbessern. Umfragen sind zweifelsohne nützlich, um beispielsweise ein Meinungsbild zu erstellen. Allerdings sollte bei der Verwendung von Umfrageergebnissen deutlich gemacht werden, dass sie nur die Meinung der befragten Menschen wiedergeben und es sich dabei nicht um tatsächliche Fakten handelt.

Außerdem müssen die Daten auch mit den von den Autoren aufgestellten Behauptungen in Verbindung stehen. Ein Beispiel: Ein Beratungsunternehmen veröffentlichte eine Pressemitteilung zu seiner Studie und titelte dazu "CEO-Vergütungspakete verhindern aktiv Innovationen in Großbritanniens Top-Unternehmen". Die Studie enthielt jedoch keinerlei Daten zu Innovationen. Sie sammelte Daten zu CEO-Vergütungspaketen und zeigte, dass diese Boni enthielten, dabei *nahm sie an*, dass diese Boni Innovationen verhindern.

3. Wurde die Studie in einer führenden akademischen Fachzeitschrift mit peer-review Verfahren veröffentlicht?

Viele Studien werden von Organisationen wie professionellen Dienstleistungsunternehmen und Think Tanks durchgeführt. Solche Studien sind von großem Wert. Die Unternehmen haben häufig einen besseren Zugang zu Daten als Akademiker und sind oft eine bessere Quelle für Statistiken. Allerdings haben Akademiker eine besondere Expertise, wenn es darum geht, Statistiken auszuwerten und Beziehungen *zwischen* den Daten herzustellen - Kausalität von Korrelation zu trennen und sich mit alternativen Erklärungen zu befassen.

Entscheidend ist, dass akademische Studien einem strengen Peer-Review Verfahren unterzogen werden müssen, um ihre wissenschaftliche Genauigkeit zu überprüfen. Solche Überprüfungen gibt es weder bei Studien von Praktiker*innen noch bei Büchern. Ein Peer-Review Verfahren befasst sich nicht nur mit tatsächlichen Fehlern in der Ausführung, sondern auch mit absichtlicher Voreingenommenheit, z. B. wenn ein Pharmakonzern einen Bericht über seine eigenen Medikamente finanziert oder wenn ein Wissenschaftler behauptet, einen Skandal aufgedeckt zu haben, nur um berühmt zu werden. Die höchsten Standards setzen dabei die wissenschaftlichen Spitzenzeitschriften, sie beschäftigen weltweit führende Spezialisten, um ein Manuskript zu prüfen, und sie lehnen bis zu 95% der eingereichten Manuskripte ab. Die 5%, die nicht abgelehnt werden, werden jedoch nicht sofort angenommen, sondern haben zunächst den Status "revise-and-resubmit". Die Gutachter*innen zeigen Themen auf, die die Autor*innen nochmals bearbeiten müssen (zum Beispiel die drei oben erwähnten Fehler), und es ist möglich., dass die Arbeit auch in der nächsten Runde noch abgelehnt wird.

Das akademische Peer-Review Verfahren ist ein fundamentaler Teil der Qualitätssicherung wissenschaftlicher Artikel; es ist nicht nur ein formeller Stempel. So können die Ergebnisse einer Studie durch den Peer-Review-Prozess tatsächlich komplett umgestoßen werden. Bei

einer Untersuchung zur Corporate Governance im britischen Unterhaus im Jahr 2016 zitierte ein Zeuge Beweise, die "zeigen, dass die Produktivität eines Unternehmens negativ mit dem Lohngefälle zwischen Top-Managern und Mitarbeitern der unteren Ebenen korreliert", und bezog sich dabei auf einen Entwurf einer Studie aus dem Januar 2010. Die fertige Version der Forschungsergebnisse wurde 2013 veröffentlicht.² Nachdem die Studie ein Peer Review Verfahren durchlaufen hatte und die Methodik nachgebessert wurde, fand sie das gegenteilige Ergebnis:

- "Wir konnten keinen negativen Zusammenhang zwischen der relativen Vergütung und der Mitarbeiterproduktivität finden."
- "Wir konnten feststellen, dass sowohl der Unternehmenswert als auch die operative Leistung mit der relativen Vergütung steigen."

Die Stringenz und Kompromisslosigkeit des Peer-Review-Prozesses ist dabei entscheidend. Dass sich eine Zeitschrift "peer-reviewed" nennt, reicht bei weitem nicht als Qualitätsmerkmal aus, da es eine enorme Bandbreite in der Qualität der Begutachtungsstandards gibt. Das Analyseunternehmen Cabell's hat eine schwarze Liste mit 8.700 Zeitschriften erstellt, die als "unlauter" einstuft werden, weil sie vorgeben, "peer-reviewed" zu sein, obwohl sie es nicht sind.

Wichtig ist auch zu wissen, dass die Qualität von Zeitschriften leicht überprüft werden kann, indem man sich eine der frei verfügbaren Listen ansieht, die die besten Zeitschriften benennt. Für die Wirtschaftswissenschaften gibt die *Financial Times* eine Liste der Top 50 Journale heraus; für andere Bereiche der Wissenschaft erstellen z.B. das National Institute of Environmental Health Sciences und Scimago ebenfalls solche Listen. Zudem ist der "Impact Factor" einer Zeitschrift ein Maß für die Qualität, das leicht nachgeschlagen werden kann.

Der Peer-Review-Prozess ist nicht perfekt – auch hier werden Fehler gemacht. Manchmal werden schlecht ausgeführte Arbeiten angenommen und gute Arbeiten abgelehnt. Dennoch ist es besser, etwas Geprüftes zu zitieren als etwas Ungeprüftes. Einige Autor*innen nehmen das Argument der "Publikationsverzerrung" als Freibrief dafür, jede beliebige Studie zu zitieren, die die ihnen gefällt und dabei zu ignorieren, ob die Arbeit tatsächlich veröffentlicht wurde. Ihr Vorwurf ist, dass Journale nur Arbeiten veröffentlichen, die traditionelle Lehrmeinungen unterstützen. So funktioniert der Publikationsprozess jedoch nicht. Herausgeber von Zeitschriften möchten neue Arbeiten veröffentlichen, die die Denkweise der Menschen verändern. Der Impact Factor einer Zeitschrift bemisst sich unter anderem danach, wie häufig ihre Artikel zitiert werden. Die erste Arbeit in einem neuen Gebiet wird sehr oft zitiert werden, die zehnte in einem etablierten Gebiet eher nicht mehr.

Ein Vorbehalt zu dieser Frage ist, dass es durchaus eine kleine Anzahl von Praxisstudien gibt, deren Qualität ausreichend wäre, um in einer akademischen Zeitschrift veröffentlicht zu werden, die aber ein solches Ergebnis nicht anstreben, da eine akademische Veröffentlichung nicht das Ziel ist. Ein Beispiel hierfür sind Studien von Wirtschaftsberatungsunternehmen (oder der Wirtschaftsabteilung eines Dienstleistungsunternehmens), die von der Regierung oder einer Regulierungsbehörde in

² Faley, Olubunmi, Ebru Reis and Anand Venkateswaran (2013): "The Determinants and Effects of CEO- Employee Pay Ratios." *Journal of Banking and Finance* 37, 3258–3272.

Auftrag gegeben wurden. Es ist daher wichtig, zwischen Studien zu unterscheiden, die zur *Wissensgewinnung* durchgeführt werden, und Studien, die zur *Werbung* für eine Marke oder die Dienstleistungen einer Organisation durchgeführt werden. Eine wichtige Frage, die man sich dabei stellen muss, ist: *Hätte die Organisation die Studie veröffentlicht, wenn sie das gegenteilige Ergebnis festgestellt hätte?* Zum Beispiel werden wahrscheinlich viele Beratungsunternehmen Studien veröffentlichen, die nachweisen, dass ethische Unternehmen bessere Ergebnisse erzielen (da diese Behauptung gut für ihre Marke ist), jedoch von der Veröffentlichung eher absehen, wenn die Studie zur gegenteiligen Erkenntnis kommt. Diese Motivation ist besonders stark, wenn die Studie das eigene Maß für ethisches Verhalten des Unternehmens verwendet hat, da dies zeigt, dass dieses Maß "funktioniert". Selbst wenn die Antwort "Nein" lautet, bedeutet dies nicht automatisch, dass die Studie falsch ist, sie sollte jedoch mit Vorsicht genossen werden.

4. Welche Referenzen hat der/die Autor*in?

Natürlich beginnt jede wissenschaftliche Arbeit unveröffentlicht. Wie können wir daher die Qualität einer neuen Arbeit einschätzen? Die vierte Dimension, die es zu prüfen gilt, ist die Legitimation der Autor*innen. Ein relevanter Faktor ist z.B. die Qualität der Institution, bei der sie arbeiten und die wir mit frei verfügbaren Listen von Top-Universitäten vergleichen können. Dies ist kein Elitedenken, sondern einfach der Wunsch, die bestmögliche Evidenz zu nutzen. Wir würden eher auf eine medizinische Meinung der Berliner Charité hören als auf die Meinung von einer Institution, von der wir noch nie gehört haben.

Sicherlich ist es nicht so, dass Studien von Top-Institutionen immer richtig und die von anderen Institutionen immer falsch sind. Ein weiterer Faktor ist es daher, die Erfolgsbilanz von Top-Publikationen eines/r Autors/Autorin zu prüfen. Da fast alle Akademiker*innen ihre Lebensläufe zur Verfügung stellen, ist dies ein leichtes Unterfangen. Wir prüfen die Referenzen eines Sachverständigen in einem Prozess stets sehr sorgfältig. Auch dies bedeutet nicht, dass gut veröffentlichte Autor*innen immer Recht haben. Referenzen sind lediglich ein Faktor, der bei der Bewertung zu berücksichtigen ist, genauso wie die Marke eines Unternehmens ein Faktor bei einer Kaufentscheidung ist oder die Universität eines Studenten ggf. ausschlaggebend bei einer Einstellungsentscheidung sein kann. Hilfreich ist es, sich die folgende Frage zu stellen: *Wenn dieselbe Studie von denselben Autoren und mit denselben Referenzen verfasst worden wäre und die gegenteiligen Ergebnisse geliefert hätte, wären wir dann immer noch bereit, ihr zu glauben?*

Ebenso wichtig ist es, zu prüfen, ob die Autor*innen auch über Referenzen auf dem *relevanten Gebiet* verfügen. Dies hilft, das Problem des "Halo-Effekts" zu vermeiden, bei dem eine Person mit Expertise in einem Bereich auch als Guru in anderen Bereichen angesehen wird. Der Manager des FC Liverpool, Jürgen Klopp, lehnte es ab, seine Meinung zum Coronavirus zu äußern, weil er befürchtete, dass die Leute ihm nur deshalb glauben könnten, weil er berühmt ist. Während es offensichtlich erscheinen mag, dass ein Fußballmanager nicht die beste Referenz bei einer wissenschaftlichen Frage ist, gibt es Schauspieler*innen, deren Meinung Menschen dazu gebracht hat, ihre Kinder nicht zu impfen.

Weniger offensichtlich ist dieser Effekt, wenn führende CEOs Ansichten zu

Wirtschaftsthemen äußern. Zum Beispiel wird der ehemalige GE-CEO Jack Welch häufig mit der Behauptung zitiert, Shareholder Value sei "die dümmste Idee der Welt". Welch gilt sicherlich als einflussreicher CEO in einem Unternehmen, er hat jedoch nicht untersucht, wie sich eine Shareholder-Value-Orientierung auf die Leistung von Unternehmen im Allgemeinen auswirkt - eine Frage, die eher durch eine akademische Studie als durch Erfahrung in der Unternehmensführung beantwortet werden kann. Einige Ärzte verkünden Wunderheilungen für das Coronavirus, auch wenn ihre Expertise nicht in der Pathologie oder Epidemiologie liegt.

Ein Warnzeichen sind Gurus, die ihre Referenzen mit unüberprüfbar behauptungen aufbauen, was häufig bedeutet, dass es ihnen an echter Expertise mangelt. Gängige Beispiele sind dabei:

- "Bestseller-Autor". Es gibt keine klare Definition für diesen Status. Es ist nicht definiert, ob man in den Top 10, 100, 1000 sein muss; ob dies für alle Bücher oder für Bücher in einem sehr kleinen Teilbereich gilt; oder wie lange man in der Bestsellerliste sein muss (Amazons Bestsellerliste wird jede Stunde aktualisiert).
- "Einer der besten Ökonomen der Welt." Auch hier gibt es keine klare Rangfolge der Ökonomen. Die Behauptung ist nur valide, wenn eine entsprechende Metrik angeführt wird (z. B. Anzahl Publikationen oder Zitate)
- "International keynote speaker". Dies bedeutet lediglich, dass die Person einen Hauptvortrag in einem Land außerhalb Deutschlands gehalten hat, auch bei einer kleineren Konferenz.
- "Award-winning professor." Auch dies bedeutet nur, dass der Professor einen Preis gewonnen hat, sagt jedoch nichts über die Wettbewerbsfähigkeit des Preises aus.

5. Gibt es alternative Erklärungen?

Die fünfte zu prüfende Dimension ist, ob es alternative Erklärungen für die Ergebnisse der Autoren gibt. Man muss kein akademischer Insider sein, um diese Prüfung durchzuführen, da die meisten alternativen Erklärungen eher auf gesundem Menschenverstand als auf methodischen Besonderheiten beruhen. Die Leser können sich fragen, ob die Ergebnisse durch umgekehrte Kausalität oder ausgelassene Variablen getrieben sein könnten.

Dabei muss es nicht lange dauern, um herauszufinden, ob es alternative Erklärungen gibt. Die Einleitung einer akademischen Arbeit (typischerweise 4-6 doppelseitige Seiten) sollte sowohl in sich abgeschlossen als auch in nicht-technischer Sprache verfasst sein. Sie sollte alle wichtigen Punkte der Arbeit enthalten - einschließlich des Umgangs mit alternativen Erklärungen -, ohne dass der Leser sich in die eigentliche Arbeit vertiefen muss.

Auch diese Frage kann sowohl für veröffentlichte als auch für unveröffentlichte Arbeiten gestellt werden. Die meisten unveröffentlichten Arbeitspapiere aus der Sozialwissenschaft sind im Social Science Research Network, www.ssrn.com, frei zugänglich. Selbst wenn ein Papier in einer Zeitschrift veröffentlicht wird, die hinter einer Paywall liegt, bleibt die Vorabversion normalerweise auf SSRN verfügbar.

6. Ist die Arbeit ausgewogen?

Die letzte zu prüfende Dimension ist, ob ein Artikel ausgewogen ist. In den Sozialwissenschaften (z. B. in der Betriebs- oder Volkswirtschaftslehre) gibt es zu fast jedem

Thema zwei Seiten, und es ist schwierig, ein Ergebnis endgültig und zweifelsfrei zu beweisen. Anders als in den Naturwissenschaften kann man keine kontrollierten Experimente durchführen, bei denen man nur eine Variable (z.B. Ungleichheit) verändert und alles andere konstant hält (z.B. die Qualität der Regierung eines Landes).

Autor*innen sollten sich daher vor Studien hüten, die behaupten, "eindeutige Beweise" oder "Beweise" gefunden zu haben oder etwas "zweifelsfrei" aufzuzeigen. Ähnlich wie übertriebene Autorennachweise verbergen übertriebene Beweisbehauptungen häufig tatsächliche Mängel und Fehler in der Analyse. Forscher, die "eindeutige Beweise" versprechen, haben möglicherweise alternative Erklärungen nicht ernsthaft in Betracht gezogen. Darüber hinaus besteht ein erheblicher Anreiz, von "eindeutigen Beweisen" zu sprechen, da eine solche Studie mit größerer Wahrscheinlichkeit weiter verbreitet wird als eine mit nuancierten Ergebnissen. Im Gegensatz dazu werden gründliche Forscher deutlich machen, was ihre Studie tatsächlich zeigen kann und was nicht.

Das oben Gesagte trifft auf Forschungen zu, bei denen eigene Primärdaten gesammelt und Analysen durchgeführt werden. Viele Praxisstudien erheben keine Primärdaten, sondern aggregieren die Ergebnisse bestehender akademischer Arbeiten. Auch hier ist es selten der Fall, dass die akademischen Erkenntnisse eindeutig zugunsten einer Seite ausfallen. Studien, die dies behaupten, haben sich wahrscheinlich dafür entschieden, nur die akademischen Beweise einzubeziehen, die unterstützen, was die Studie zeigen möchte, und widersprüchliche Beweise absichtlich auszuschließen. Wenn eine Studie behauptet: "Es gibt eindeutige Beweise dafür, dass X Y erhöht", ist eine einfache Überprüfung, "X verringert Y" in eine Suchmaschine einzugeben und zu sehen, ob es Links zu hochwertigen Beweisen gibt, die das Gegenteil zeigen.

Zusammenfassend lässt sich sagen, dass eine ausgewogene Studie eher mit der Absicht geschrieben wird, zu informieren und zum Wissen beizutragen, als die Autoren berühmt zu machen oder die Marke einer Organisation zu stärken.