

Aspekte von Evidence-Based Medicine und Medical Decision Making

M. F. Fischmeister

Evidence based Medicine is the conscientious, explicit und judicious use of current best evidence in making decisions about the care of individual patients. (Sackett 1997)

Decision analysis is a systematic approach to decision making under conditions of uncertainty. (Raiffa 1968)

Einleitung

Die Tätigkeit eines Arztes erfordert neben anderem das Treffen von Entscheidungen. Beschäftigt man sich als Chirurg mit diesem Thema stößt man sehr bald an Grenzen, die im eigenen mathematischen Wissen liegen und darin, dass es kaum interessierte Chirurgen als Gesprächspartner gibt. Dieser Essay soll einige Gedanken zur Evidence Based Medicine zur Diskussion stellen, die Verbindung zu Medical Decision Making offen legen und zum Nachdenken anregen.

Das Gedankenexperiment

Eine der besten Methoden, Einsichten über einen Gegenstand, ein physikalisches Gesetz oder eine Hypothese zu bekommen, ist, sich an den Rand des Gültigkeitsbereichs desselben zu begeben. Im Rahmen einer solchen Überlegung dachte ich, es könnte gelingen in einem einfachen Gedankenexperiment eine randomisierte kontrollierte Studie zum Vergleich zweier Behandlungsgruppen unter Einhaltung aller Spielregeln und Vorschriften durchzuführen – einem Behandlungsvergleich zwischen einer konservativen Behandlung, um es vielleicht anschaulicher zu machen, einer konservativ funktionellen Frakturbehandlung und einer chirurgischen Therapie mit einer modernen Osteosynthese.

Nun könnte es durch Zufall so sein, dass, mit einem idealen Messinstrument gemessen, die Outcome-Daten beider Behandlungsgruppen eine Normalverteilung aufweisen und beide Datensätze denselben Mittelwert aber eine unterschiedliche Varianz aufweisen (Abb. 1). Wir kennen Ähnliches aus der chirurgischen Praxis. Eine operative Behandlung in der Unfallchirurgie vermag die anatomischen Verhältnisse genauer wiederherzustellen und damit zu besseren klinischen Ergebnissen führen als eine konservative Behandlung, aber eben auch zu Infektionen und schlechteren Ergebnissen, die mit einer konservativen Behandlung nicht eingetreten wären.

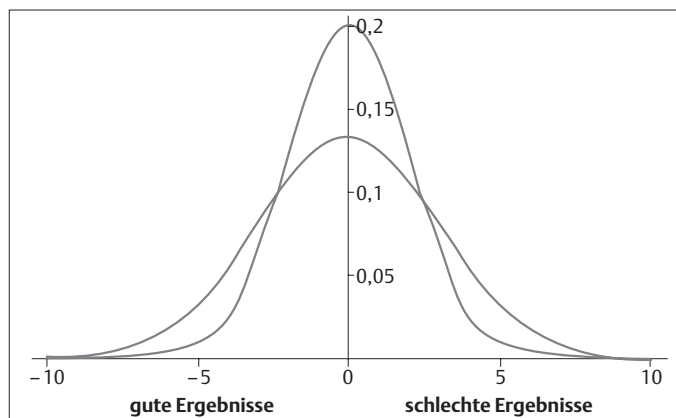


Abb. 1 Vergleich zweier Behandlungen.

Analysiert man das Ergebnis eines RCT (2), so ist die erste Frage immer: Gibt es einen klinisch interessanten Unterschied zwischen den Behandlungsergebnissen (quantitative Signifikanz)? Der zweite Punkt einer Analyse besteht darin die Verteilungen der beiden Datensätze zu vergleichen und zu fragen: Welche Rolle kommt dem Zufall in der Beurteilung der Unterschiede zu (stochastische Signifikanz)? Zuerst sollten wir uns den Daten selbst zuwenden.

Probleme mit den Daten

Daten sollten valide, reliabel, sensibel und patientennah sein (3). Geht man davon aus, dass diese Bedingungen in unserem Gedankenexperiment erfüllt sind, so bleiben immer noch einige Fragen offen.

Eine klinische Situation ist durchaus komplex und mit den normal verteilten Messparametern oft nicht ausreichend zu beschreiben. Eine postoperative Infektion ist zum Beispiel ein Ereignis, das qualitativ von den anderen Behandlungsausgängen und deren Messungen entscheidend abweicht. Damit ist eine Voraussetzung, dass dieses statistische Modell eines Vergleichs von Verteilungen verwendet werden kann, eigentlich nicht erfüllt. Immer wenn in einem Vergleich eines chirurgischen Heilverfahrens Amputationen, Ostitis aut alia als Outcome beurteilt werden müssen, stellt sich die Frage, ob eine derartige Behandlungskatastrophe in

einem statistischen Modell qualitativ genauso beurteilt werden kann wie ein guter oder weniger guter Outcome.

Ein weiteres Problem besteht darin, dass Gruppen von Patienten verglichen werden, es sich aber eigentlich um den Vergleich von Behandlungsprozessen handelt. In vielen Studien erfahren wir nichts darüber, ob diese sich auch in einem Zustand der statistischen Kontrolle befanden. Lernkurven etc. spielen hier eine wesentliche Rolle. Auch in unserem Gedankenexperiment nehmen wir dazu nicht Stellung. Erst die Darstellung der Daten in einer Zeitreihe oder einer Process Behaviour Chart (4) würde uns darüber Auskunft geben.

In der Realität arbeiten wir aber meist mit kategorialen Skalen und nonparametrischen Daten, und die können in unserem Experiment vielleicht so wie in Abb. 2 aussehen. Wenn wir unsere Beurteilungsskala wie auf Abb. 2 streng auswählen, sodass die Unterschiede der Verteilung im Segment der guten Ergebnisse genauer dargestellt werden, dann geht es um die Antwort auf die Frage: Wie gut sind die Ergebnisse? Im umgekehrten Fall freuen wir uns über die durch den Plafond-Effekt erzielten vielen guten und exzellenten Ergebnisse (die meisten Skalen sind so, dass sie unser psychologisches Bedürfnis gute Ergebnisse zu erzielen, befriedigen). Eigentlich beantworten sie aber die Frage: Wie schlecht sind unsere Ergebnisse?

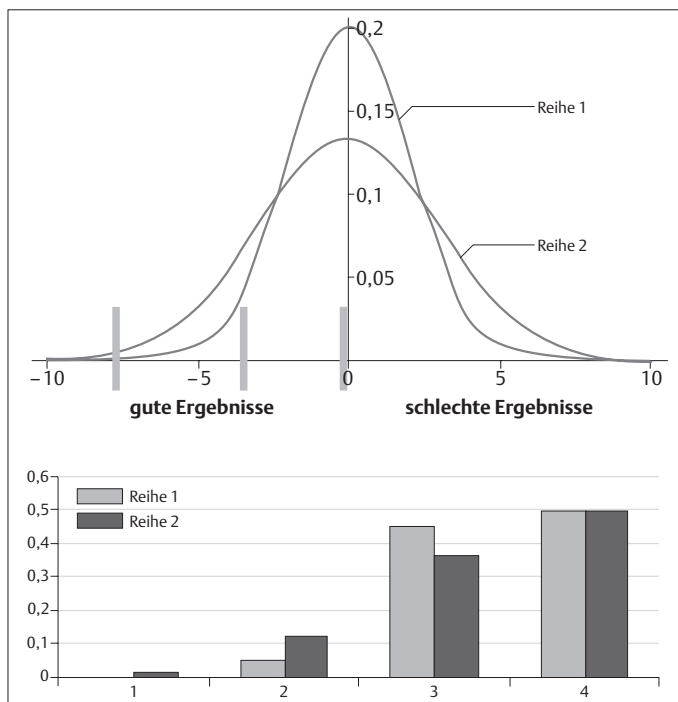


Abb. 2 Vergleich zweier Behandlungen in unterschiedlichen Skalen.

Die Verwendung von diesen Skalen stellt natürlich immer einen Informationsverlust dar. Wenn es zu Plafond- oder Floor-Effekten kommt, dann muss davon ausgegangen werden, dass darin ein wesentlicher Informationsverlust versteckt sein kann. Die Daten sind in diesem Bereich zumindest als zu wenig sensibel zu bezeichnen.

Statistische Auswertung

Wenn man die normal verteilten Daten unseres Gedankenexperimentes auswertet, fällt auf, dass mit einem einfachen Mittelwertvergleich kein statistischer Unterschied zu beobachten ist. Betrachtet man die Auswertung der kategorialen Daten mittels eines nonparametrischen Tests, so zeigt sich, dass man je nach Wahl der Skala das Ergebnis manipulieren kann. Ist man daran interessiert zu beweisen, dass eine Therapie A besser ist als B, dann könnte man eine Beurteilungsskala wählen, die im Bereich der schlechten Ergebnisse nicht sensibel ist – und umgekehrt.

Clinical Decision Making

Medizinische Entscheidungen werden von verschiedenen Faktoren (5) beeinflusst. Diese sind:

- Art der Endzustände der verschiedenen Behandlungszweige zu einem gewählten Zeitpunkt,
- Wahrscheinlichkeit dieser Endzustände in den verschiedenen Behandlungszweigen,
- Utility (Nutzen) dieser Endzustände für den Patienten, gemessen mit einem der geeigneten Instrumente: Qaly oder Standard Reference Gamble,
- Einstellung des Patienten und behandelnden Arztes zum Risiko der Behandlung (der Wahrscheinlichkeit, durch die Behandlung Schaden zu erleiden),
- Art der Patienten-Arzt-Beziehung (Paternalismus versus Patientenautonomie),
- Verfügbarkeit und Kosten der Behandlung für den Kostenträger,
- Art des Vergütungssystems für den Arzt.

Betrachtet man ein formales Entscheidungsbaumsystem zur Analyse unseres Gedankenexperimentes, so ist klar absehbar, dass es sich kongruent zur Analyse der Verteilungen verhält. Also auch hier lässt sich mit der Wahl der Skala das Ergebnis manipulieren. Der Wert der Beurteilung eines Entscheidungsbaumes liegt aber auch darin mittels einer Sensitivitätsanalyse die Stabilität der Entscheidungen abschätzen zu können (6).

Entscheidungsbaumanalysen gehen davon aus, dass die Patienten sich für den Pfad zum wahrscheinlich besten Behandlungsergebnis (7) entschließen werden. Dies ist in der Praxis des chirurgischen Alltags aber nicht immer so. Betrachten wir Abb. 3, so

Themen der Zeit



Gegründet 1872
Sitz Berlin

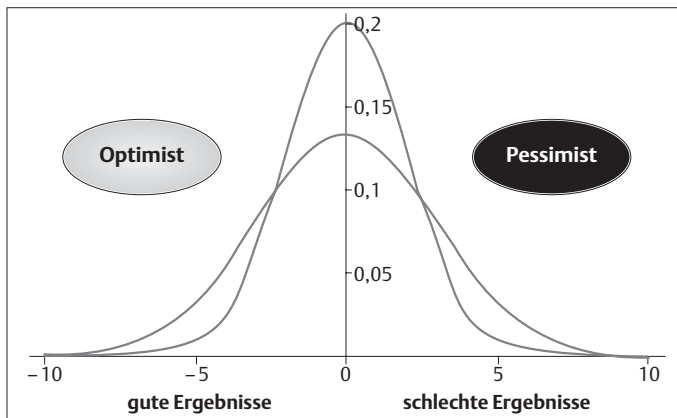


Abb. 3 Vergleich zweier Behandlungen und Risikoabschätzung.

könnte zum Beispiel ein Leistungssportler die Haltung haben: „Ich will jede Chance wahrnehmen, wieder ein exzellentes Ergebnis zu haben; die schlechten Ergebnisse interessieren mich nicht.“ Er zeigt eine Haltung der Gewinnmaximierung und wird eine Operation unbedingt haben wollen, und er wird nur Interesse an der linken Seite der Abb. 3 haben. Ein Chirurg als Patient wird vielleicht meinen: „Ich weiß, was alles passieren kann. Ich will bitte die Behandlung mit dem geringsten Risiko, auch wenn die Ergebnisse nicht so gut sind!“ Er zeigt, wie die meisten Patienten, eine Haltung der Verlustminimierung und wird vor allem an der rechten Seite der Abbildung interessiert sein.

Eine Lösung dieses Problems liegt darin, zur Messung von Endzuständen nach Behandlungsverfahren Nutzwerte (Utilities, Quality adjusted life years aut al.) zu verwenden. In diesen ist eine Beurteilung des Risikos enthalten. Die Ermittlung von Nutzwerten ist in weiten Bereichen der Unfallchirurgie und Orthopädie aber unbekannt.

Diskussion

Betrachten wir die Modelle der Statistik und Entscheidungsfindung in ihren Grenzbereichen, so erfahren wir sehr viel über ihre Anwendbarkeit in den Standardsituationen. Insbesondere zeigt dieses Gedankenexperiment: Eine medizinische Entscheidung kann, auch wenn sie gegen das Ergebnis eines RCT getroffen wird, für einen individuellen Patienten richtig sein. Eine medizinische Entscheidung kann, auch wenn sie gegen das Ergebnis einer formalen Entscheidungsbaumanalyse getroffen wird und zur Messung des Outcome keine individuellen Nutzwerte verwendet wurden, richtig sein. In der Erstellung von Leitlinien und Standard Operating Procedures steht das Wohl von Patientengruppen im Vordergrund. Die individuelle Entscheidung am individuellen Patienten kann aus guten Gründen davon abweichen. In der Beurtei-

lung von RCTs und Entscheidungsbaumanalysen ist auf die Art und Qualität der Daten besondere Aufmerksamkeit zu legen, insbesondere auf das Vorliegen von Scale Bias und darauf, dass die untersuchten Prozesse sich in einem Zustand statistischer Kontrolle befinden.

Es ist einigermaßen erstaunlich, wie leicht unsere Systeme der Wissensgewinnung und medizinischen Entscheidungsfindung infrage zu stellen sind, und trotzdem sind sie das beste, was wir haben, um unseren Patienten besser helfen zu können.

Zusammenfassung

Individuelle medizinische Entscheidungen können für den einzelnen Patienten richtig sein, auch wenn RCTs und Entscheidungsbaumanalysen andere Entscheidungen bevorzugen würden. Besonders auf den Einfluss von Scale Bias und die Möglichkeit, in bestimmten Situationen Ergebnisse von RCTs und Entscheidungsbaumanalysen zu manipulieren, wurde hingewiesen. Im Vergleich von Behandlungsprozessen ist darauf zu achten, dass diese im Zustand der statistischen Kontrolle sind. Die breitere Verwendung von Nutzwerten zur Outcome-Messung auch in der Unfallchirurgie und Orthopädie wäre wünschenswert.

Literatur

- 1 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine. New York, Edinburgh: Churchill Livingstone, 1997
- 2 Feinstein AR. Clinical Epidemiology. Philadelphia, London: W. B. Saunders, 1985
- 3 Feinstein AR. Clinimetrics. New Haven, London: Yale University Press, 1987
- 4 Wheeler DJ, Poling SR. Building Continual Improvement. Knoxville: SPC Press, 2001
- 5 Andersen TF, Mooney G. The Challenges of Medical Practice Variations. Basingstoke: Macmillan Press, 1990
- 6 Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia, London: W. B. Saunders, 1980
- 7 Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston, London: Butterworths, 1988

Anschrift des Verfassers:

Dr. Martin F. Fischmeister
Facharzt für Unfallchirurgie
Ferihumerstraße 11
4040 Linz
Österreich
Tel.: +43/(0)70716617
Fax: +43/(0)707166173
E-Mail: martin@fischmeister.info
Internet: www.fischmeister.info