

«Lies, damn lies, and statistics»: Bemerkungen zu den Outcome-Zahlen des Bundesamtes für Gesundheit

Thomas F. Lüscher

Departement für Innere Medizin, Klinik für Kardiologie, Herzkreislaufzentrum, UniversitätsSpital Zürich und Institut für Physiologie, Kardiovaskuläre Forschung, Universität Zürich

«Lies, damn lies, and statistics»...

Benjamin Disraeli, britischer Premierminister von 1874–1880 (Abb. 1), soll diese Worte gesprochen haben, die später von Mark Twain popularisiert wurden: «*There are three kinds of lies: Lies, damn lies, and statistics*». Vielleicht war Disraeli nicht der erste, der berufsbedingt bemerkte, dass Statistiken nicht selten Ergebnisse hervorbringen, die in der einen oder anderen Art gedeutet werden können – vor allem abhängig von den herrschenden Interessen oder dem Standpunkt einer Partei – aber er hat es am Besten auf den Punkt gebracht. Als Politiker in Debatten im Unterhaus gestählt, erkannte er den grundsätzlichen Unterschied zwischen Zahlen und ihrer Bedeutung, zwischen Daten und ihrer Interpretation – und davon handeln diese Zeilen.

Vom Schätzen zum Rechnen

Einst schätzte man die Einwohnerzahl ebenso das Handelsgut, die Stärke von Armeen und die Grösse von Ländern. Dann kam das Zählen, zunächst durch die Erfindung der Keilschrift bei den Sumerern, danach durch die Einführung vergleichbarer Zahlensysteme in anderen Kulturen. Alsbald wusste man über alles immer genauer Bescheid.

Auch in der Medizin hielt spät, aber umso heftiger, die Statistik Einzug: Noch im Jahre 1837 stritten sich in Paris Risueno d'Amador und Pierre Charles Alexandre Louis in der *Académie Royale de Médecine*, welche Louis XVIII 1820 ins Leben gerufen hatte, über die Frage, ob Medizin als Kunst oder als exakte Wissenschaft zu betrachten sei. Die einen hielten jeden Patienten für einmalig und die Behandlung daher für einen Akt der Kreativität des jeweiligen Arztes, sie verdamnten quantitative Angaben über Patienten und Erkrankungen. «*L'homme moyen n'existe pas!*» war ihr Credo. Pierre Charles Alexandre Louis aber hielt dagegen, dass wir nur lernen, wenn wir unserer Erfahrung Zahlen zuordnen, ihren Wert und ihr Gewicht bestimmen, ihre Veränderung bei Auftreten von Seuchen und Epidemien beschreiben, was zuletzt zur naturwissenschaftlichen Medizin führte, wie wir sie heute kennen [1].

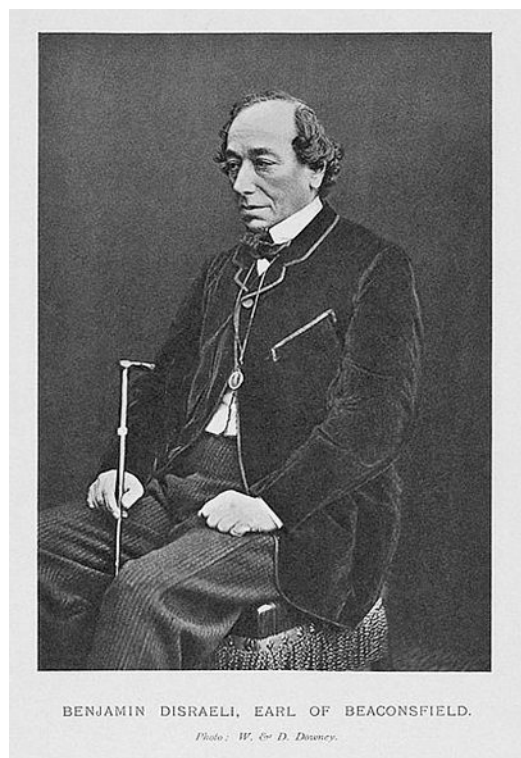


Abbildung 1

Benjamin Disraeli. Foto: Harvard Law School Library / Wikimedia Commons.

Die Epidemiologie als Wissenschaft entwickelte sich dennoch langsam, vielleicht, weil ihr die mathematischen Grundlagen fehlten, möglicherweise aber auch, weil sie unserer Erfahrung nicht entspricht: Unser Hirn lernt episodisch, überzeugt sich lieber durch Geschichten als durch Zahlen. Die narrative Ver-

Korrespondenz:

Prof. Thomas F. Lüscher
Vorsteher, Departement für Innere Medizin
Direktor, Klinik für Kardiologie
HerzKreislaufZentrum
UniversitätsSpital
Rämistrasse 100
CH-8091 Zürich
cardiotfl@gmx.ch

zerrung [2], die die Statistik zu korrigieren versucht, wird von vielen für selbstverständlich gehalten, ihre Überwindung ist daher ein Kraftakt, der nicht leicht gelingt.

Der Aufstieg der Epidemiologie

Das mussten auch die Pioniere der Disziplin erfahren, die sich mit der Beziehung von Rauchen und Krebsleiden befassten. Sir Richard Dole und Sir Austin Bradford Hill begannen Ende der 1940er Jahre Umwelteinflüsse als Ursache der bedrohlich steigenden Lungenkrebszahlen im Vereinigten Königreich zu untersuchen. Obschon anfänglich die Kohleminen Englands und Abgase aus Industrie und Verkehr als die wahrscheinlichsten Ursachen angenommen wurden, zeigten ihre Daten bald, dass das Rauchen am engsten mit dem Bronchuskarzinom in Verbindung stand. Ihr erster Bericht im Jahre 1950 wurde sehr skeptisch aufgenommen und von vielen Seiten kritisiert; später publizierte Arbeiten bestätigten aber ihre Ergebnisse und brachten das Rauchen auch mit der koronaren Herzkrankheit in Zusammenhang. Damit begann der Aufstieg der Epidemiologie, die mit berühmten Kohorten wie der Framingham-Studie aus Boston wesentlich zur Identifikation der kardiovaskulären Risikofaktoren beitrug, die wir heute täglich behandeln. Dabei lieferte die Epidemiologie zunächst nur eine überzeugende Korrelation, welche durch experimentelle Untersuchungen und Interventionsstudien bestätigt werden musste, bevor sie in das medizinische Wissen aufgenommen und in der Praxis umgesetzt werden konnte. Und in der Tat blieb die epidemiologische Beziehung zwischen Cholesterin und Infarkt eine Hypothese, bis der erfolgreiche Einsatz der Statine das Konzept stützte. Nicht jeder statistischen Korrelation war des gleichen Schicksal beschieden: Die epidemiologische Beziehung zwischen Homocystein und koronarer Herzkrankheit erwies sich als Täuschung; in kontrollierten Megatrials senkte Folsäure zwar die Homocysteinpiegel, aber nicht das Auftreten von Infarkt und Tod.

Outcome-Forschung

Nach der krankheitsorientierten Epidemiologie kam die Outcome-Forschung, die sich mit den Ergebnissen medizinischen Handelns befasst. Ein Anfang wurde mit der ersten randomisierten Studie von Sir Austin Bradford Hill zur Wirkung des Streptomycins bei Lungentuberkulose gemacht [3]. Dabei ging es aber noch ums Prinzip: Wirkt eine therapeutische Massnahme oder nicht? Danach wandte man sich der praktischen Umsetzung des in kontrollierten Studien Erprobten zu: Wird eine wirksame therapeutische Massnahme auch umgesetzt und wenn ja, wie gut sind die Ergebnisse im Alltag, in einem bestimmten Spital, eines bestimmten Arztes? Hier geht es um Qualitätskontrolle und nicht

um die Erforschung des Neuen. Die Methodik ist aber nicht grundsätzlich anders; auch hier wird gezählt, berechnet und interpretiert.

Amtlich zertifizierte Zahlen

Die Behandlungsergebnisse von Zentren wurden anfänglich von den Betroffenen selber erhoben, ja sie gehören auch heute noch zu den zentralen Gebieten akademischen Forschens und werden an zahllosen Kongressen und Publikationen vorgestellt und diskutiert. Neu ist die Transparenz solcher Daten, die Leser sind heute nicht nur Kollegen, vielmehr Behörden, Vorgesetzte, ja die Öffentlichkeit an sich. Den ersten Schritt machten die Amerikaner. So sind in New York die Daten sämtlicher herzchirurgischer Zentren und ihrer Operateure im Internet verfügbar. Seit dem Skandal in der Herzchirurgie von Bristol gilt in England ähnliches. Grundsätzlich ist das positiv zu werten, denn das Anliegen ist die Erhebung der Behandlungsqualität im Hinblick auf ihre Verbesserung. Auch soll der Patient die Möglichkeit haben, sich die besten Zentren für sich auszusuchen.

In der Schweiz hat sich das Bundesamt für Gesundheit (BAG) dieser wichtigen Aufgabe angenommen und veröffentlicht, entsprechend den Vorgaben des 1996 eingeführten Krankenversicherungsgesetzes (KVG Art. 58 und 77), seit einigen Jahren Daten zu Sterblichkeit und Infektionen in den Spitälern unseres Landes. Mit beinahe schon gewohnter Regelmässigkeit wird der Zeitungsleser – denn nur in diesen Blättern erhalten die vom BAG generierten Zahlen die ihnen zustehende Aufmerksamkeit – mit Daten konfrontiert, die den Laien zum Staunen und den Fachmann zur Verwunderung bringen. Beide sind schliesslich *«confused at a higher level»*. Dass es die Interpretation des gelieferten Zahlenmaterials in sich hat, haben selbst die Journalisten erfasst, wie die kritischen Artikel der letzten Wochen im Pressewald belegen [4].

Datensammeln

Dass Daten so gut sind, wie sie erhoben werden, gilt als Gemeinplatz, beachtet wird er kaum [5]. Entscheidend ist, dass eine Stichprobe repräsentativ oder vollständig ist. Will man die Gesundheit der Schweizer untersuchen, wäre Vollständigkeit, das heisst die Erfassung eines jeden Bürgers, wohl das Beste, doch lässt sich dies kaum durchführen. Entsprechend muss die kleine Stichprobe repräsentativ sein, sollte alle Aspekte der Schweizer widerspiegeln, zufällig generiert sein, eine genügend hohe Fallzahl aufweisen und jeden Auswahl-Bias meiden. Das gelang und gelingt erfahrenen Forschern von Richard Doll bis Richard Peto.

Für kleinere Populationen – und davon sprechen wir hier – ist Vollständigkeit gefragt: Will man das Outcome von Kolonoperationen in einem bestimmten

Spital bestimmen, so müssen alle Patienten und ihre Daten prospektiv erfasst werden. Werden einige vergessen oder anders kodiert, ist die Kohorte nicht vollständig und die Daten unbrauchbar. Auch eine retrospektive Erfassung kann zur Datenverfälschung führen, vor allem, wenn einige Parameter erst nachträglich erfasst werden.

Vom Segen des Kodierens

Was früher ehrgeizige Assistenten verrichteten, erledigen heute professionelle Kodierer nach ausgeklügelten Algorithmen. Das BAG hat dazu ein angepasstes System der deutschen Privatklinikkette Helios übernommen. Damit vergleichen die Kodierer in den Spitälern die Zahl der Fälle mit der Zahl der Toten. Wie uns versichert wird, handelt es sich bei den Kodierern um Fachkräfte mit medizinischem Hintergrund, die ihre Sache verstehen. Die seitenlangen Listen werden unter dem Titel «Qualitätsfaktoren der Schweizer Akutspitäler» vom BAG jeweils im April veröffentlicht.

Was sagen diese Daten aus? Messen sie wirklich die Qualität von Akutspitälern und wenn nicht, wo liegen die Probleme? Dass es sich bei dieser Statistik in ihrer jetzigen Form bestenfalls um «*damn lies*» handelt, liegt an Folgendem: (1.) referral bias; (2.) Kodierungs-Bias; und (3.) Erfassungsfehlern.

Referral bias

Dass in einem Zentrums- und insbesondere einem Universitätsspital die Fallgewichtung anders ist als in einem Bezirksspital, ist offensichtlich. Der Patient mit Leukämie und einem schweren Infekt wird ebenso ins Zentrum verlegt wie die Patientin mit unbeherrschbarer Sepsis oder der Polytraumatisierte nach einem Verkehrsunfall. Die Mortalität wie auch die Infektraten müssen daher höher sein als in einem kleinen Spital – das Zuweisungsverhalten macht die Statistik, nicht die Qualität der Tätigkeit der Ärzte vor Ort. Die Diskussion um die unterschiedlichen Infektraten in grossen und kleinen Spitälern bei der Veröffentlichung der Daten von 2007 zeigt den Unsinn dieses Tuns – die Öffentlichkeit wird mit der Deutung der Zahlen überfordert und kommt zu unsinnigen Schlüssen.

Ein weiteres Beispiel: In einer städtischen Umgebung mit gutausgebautem Ambulanzsystem werden einem Zentrumsspital mehr reanimierte Patienten zugeführt als in ländlicher Umgebung mit langen Distanzen, denn «*out of hospital arrests*» überleben im ersten Fall eher bis ins Spital als im zweiten. Die Mortalität von Infarkten mit dieser Präsentation ist aber zehnmal höher als bei Patienten, die in stabilem Zustand das Spital erreichen. Je mehr sich ein Spital daher bemüht, instabile Patienten anzunehmen und zu behandeln, desto mehr werden wohl gerettet, aber ein ebenso erheblicher Teil stirbt trotz aller Bemühun-

gen. Das Zentrumsspital wird daher im Mittel eine höhere Infarkt mortalität ausweisen als ein kleines.

Kodierungs-Bias

Die meisten Patienten haben mehrere Diagnosen, was das Kodieren nicht einfach macht. Bleiben wir beim obengenannten Beispiel: Ein «*out of hospital arrest*» kommt nach Reanimation vor Ort intubiert in ein Zentrumsspital. Die akute koronare Intervention verläuft erfolgreich, das Gefäss wird wiedereröffnet, der Patient kommt auf die Intensivstation. Dort entwickelt sich ein systemischer Infekt, eine längere Hospitalisation zeichnet sich ab. Wie wird nun kodiert? Sepsis oder Infarkt? Das hat erhebliche Konsequenzen. Wird aus ökonomischen Gründen – und dies scheint möglich – die komplexere Diagnose gewählt, so ist die Infarktkohorte inkomplett und die Zahlen werden verfälscht. Stirbt der Patient am Infekt, so liegt die Infarkt mortalität zu tief, da der Fall einer anderen Diagnose zugeordnet wird.

Will man also über die Mortalität einer Erkrankung etwas erfahren, so müsste die Eintrittsdiagnose entscheidend sein. Doch die Sache ist komplexer: Abhängig von unserer Fragestellung müssten wir sowohl über die Mortalität bei Sepsis und bei Infarkt etwas erfahren können. Der Patient müsste wohl doppelt kodiert und tabellarisch ausgewiesen werden, wenn die Daten einen Sinn ergeben sollen.

Erfassungsfehler

Wie die Amerikaner sagen: «*bullshit in, bullshit out*» – eine Statistik ist so gut wie die Qualität ihrer Daten. Wenn der Schreibende erfahren muss, dass in seiner Klinik die Infarkt mortalität im Spital bei über 10% liege und zufälligerweise gleichzeitig eine sorgfältige Analyse zweier Assistenten mit Nachkontrollen der Patienten vorliegt, die eine 30-Tage-Mortalität von 2,7% ausweist (Tab. 1) [6], fragt man sich, wie die vielgerühmten Daten entstehen. Ein genauerer Vergleich ergab, dass im Jahre 2008 nur 125 der 412 Infarkte, welche am Universitätsspital behandelt wurden, von den professionellen Kodierern erfasst wurden, also nicht einmal ein Drittel. Da die meisten der tödlich verlaufenen Fälle in die Statistik eingingen, ergab sich eine viermal zu hohe Mortalität.

Wie entstehen solch grobe Fehler bei der Datenerfassung? Zum einen werden die ambulanten Patienten aufgrund der Vorgaben des BAG nicht erfasst. Wird ein Patient aus Wetzikon zur Akutintervention ins Universitätsspital Zürich überwiesen und vor Mitternacht des gleichen Tages aufgrund des Bettenmangels wieder zurückverlegt, wird er in der BAG-Statistik nicht erfasst. Da es sich hierbei meist um stabile Patienten handelt, kommt es bei der Statistik zu einer Überschätzung der Mortalität. Im vorgegeben Fall handelte

Tabelle 1

Vergleich Outcome nach Herzinfarkt 2008: Angegeben sind die vom Universitätsspital mit der vom Bundesamt für Gesundheit vorgegeben Kodierung erhobenen Daten (BAG) sowie die von wissenschaftlichen Assistenten der Klinik für Kardiologie nach Durchsicht aller Berichte des Jahres 2008 sorgfältig erhobenen und durch Verlaufbeobachtungen ergänzten Daten zur 30-Tage-Mortalität (KAR). Beachtenswert sind die unterschiedlichen Fallzahlen in der kodierten und wahren Sterblichkeit.

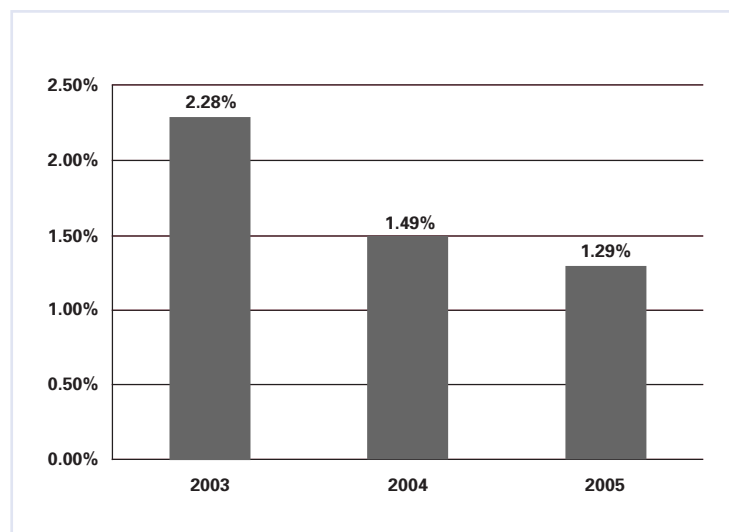
Variable	Daten	Fallzahl	Todesfälle	Sterblichkeit (%)
Herzinfarkte total	BAG	125	14	11,2%
	KAR	415	11	2,7%
Altersgruppe 20–44	BAG	7	0	0,0%
	KAR	30	0	0,0%
Altersgruppe 45–64	BAG	56	3	5,4%
	KAR	211	5	2,3%
Altersgruppe 65–84	BAG	49	5	10,2%
	KAR	150	6	3,8%
Altersgruppe ≥85	BAG	13	6	46,2%
	KAR	13	0	0%
Fälle mit NSTEMI	BAG	14	0	0%
	KAR	135	2	1,5%

es sich immerhin um 69 unerfasste Patienten mit Infarkt (17% aller Patienten). Dazu kommen Kodierungsfehler und möglicherweise schlichtes Übersehen von Fällen mit der Diagnose.

Es bräuchte eine unabhängige Adjudikation der Daten, wie dies in wissenschaftlichen Studien und Trials seit Jahren zwingend vorgegeben ist. Daten, die nur ein Kodierer sieht und nicht die verantwortliche Einheit des Spitals, die die Gesamtzahl der Patienten kontrollieren kann, sowie ein unabhängiger Monitor, sind unbrauchbar. Gewiss, die Datenaquisition ist in der heute praktizierten Form deutlich billiger, doch was billig ist, ist in diesem Falle auch wertlos.

Abbildung 2

Anteil der Patienten mit kardiogenem Schock, welche in den Jahren 2003, 2004 und 2005 im Staate Massachusetts in den USA mit einer akuten perkutanen Koronarintervention behandelt wurden (modifiziert aus: J Am Coll Cardiol. 2009;53:825–30 [7]).



Folgen der Transparenz

Die Veröffentlichung von Outcome-Daten geschieht mit gutem Willen: Man will die Qualität überwachen und dadurch wohl auch verbessern. Daneben hat diese Transparenz aber auch unvorgesehene Folgen: Wie in der Ökonomie beeinflussen die Ergebnisse das Verhalten der Betroffenen. Im Staate Massachusetts beispielsweise hat die öffentliche Verfügbarkeit von Outcome-Daten zu verblüffenden Ergebnissen geführt (Abb. 2) [7]: So haben Patienten, welche wegen eines kardiogenen Schocks mit einer Akutintervention behandelt werden, innerhalb von nur 2 Jahren um 40% abgenommen. Offensichtlich hat

die Sichtbarkeit von Mortalitätsdaten dazu geführt, dass Patienten mit kardiogenem Schock zunächst beobachtet und bei hohem Risiko gar nicht erst wirksam behandelt werden. In New York geschah in der Herzchirurgie ähnliches: Hochrisikopatienten wurden an ein Zentrum ausserhalb des Staates überwiesen. Wer will seine Reputation schon mit Hochrisikopatienten belasten? Und in der Tat, wer auch in hoffnungslosen Situationen noch Infarktpatienten zu retten versucht, wird als Operator die höchsten Mortalitätsraten auf sich nehmen müssen, da die Sterblichkeit bei intubierten und katecholaminbedürftigen Patienten um das Zehnfache steigt. Wir treffen also auf eine sozusagen medizinische Unschärferelation wie sie Heisenberg für die Physik beschrieb: Das beobachtet Objekt verändert sein Verhalten – die Amerikaner sprechen vom *risk avoidance creep* (Abb. 3). Unkorrigierte und nicht separat ausgewiesene Mortalitätsraten sind daher unverantwortlich, weil sie die Versorgung verschlechtern und nicht verbessern.

Was wollen wir erreichen?

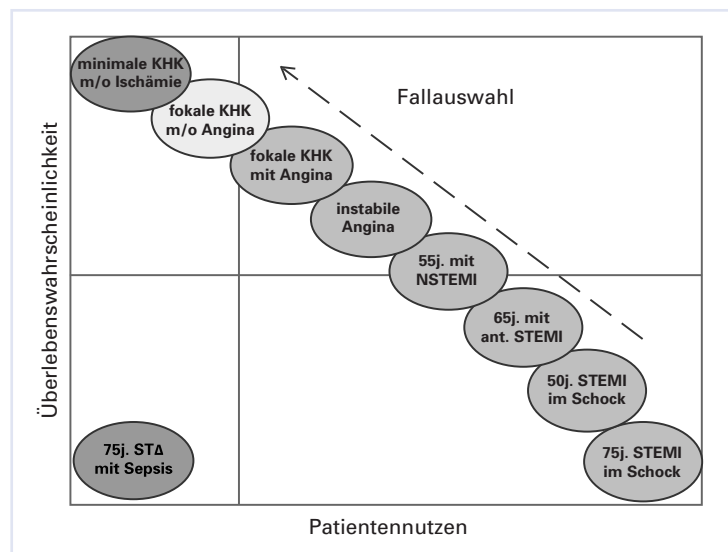
Immanuel Kant ging davon aus, dass nichts ohne Einschränkung als gut erachtet werden kann denn ein guter Wille – und dies können wir dem Gesetzgeber wie dem BAG durchaus zugestehen; doch reicht dies hier nicht aus. Wie wir gesehen haben, sind die vom BAG vorgelegten Outcome-Daten unvollständig, teilweise falsch und irreführend. Sie schaden Spitalern, die ihre Daten offenlegen (es sind bisher nur 29 von 180 Akutspitalern), sie verunsichern die Bevölkerung mit zum Teil überhöhten Sterblichkeitsraten und stellen Zentrumsspitaler bloss, denen komplexe Patienten überwiesen werden. Zudem könnten undifferenzierte Zahlen, die die Komplexität und das Risikoprofil des Patienten nicht berücksichtigen, wie in den USA, zu

unvorhergesehenen Risikovermeidungsstrategien von Spitälern und Ärzten führen, die die Versorgung beeinträchtigen und nicht verbessern.

Was es bräuchte, wären verlässliche Daten, die durch professionelle Fachkräfte vor ihrer Veröffentlichung adjudiziert und geprüft werden, die mit den in-

Abbildung 3

«Risk Avoidance Creep»: Ärzteverhalten angesichts transparenter Outcome-Daten bei Hochrisiko-Patienten. Die Fallauswahl bewegt sich tendenziell von Hochrisiko-Patienten mit dem grössten Nutzen einer perkutanen koronaren Intervention zu den «sichereren» Fällen mit weniger grossem Nutzen (modifiziert aus: J Amer Coll Cardiol. 2009;53:825–30 [7]).



ternen Zahlen der Kliniken abgeglichen und hinsichtlich der Fallkomplexität korrigiert und separat ausgewiesen werden, wie dies in Holland bereits geschieht. Es macht keinen Sinn, eine Infarkt mortalität anzugeben, wenn man die Verteilung stabiler und schockierter Patienten nicht kennt, denn die Sterblichkeit ist nicht normal verteilt. Eine solche Datenerfassung kostet aber, wie der Vergleich mit der klinischen Forschung zeigt: Die Einführung von Adjudication-Committees und Monitoren haben die therapeutischen Trials erheblich verteuert, aber die Qualität der Daten verbessert. Der Gesetzgeber und das BAG aber gehen offenbar davon aus, dass man den gesetzlichen Auftrag auf die billige Weise erledigen kann – das ist irreführend und unverantwortlich. Bis eine Besserung eintritt, sollte man daher auf Benjamin Disreali hören und die Daten unbeachtet zur Seite legen.

Literatur

- 1 Lüscher TF. Gedankenmedizin. Heidelberg: Springer; 2009.
- 2 Taleb NN. The black swan – the impact of the highly improbable. London: Penguin Books; 2008. S. 63 ff.
- 3 Hill AB. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. Br Med J. 1948;769–73.
- 4 Furger M. Sterben verboten. NZZ am Sonntag, 19. Juli 2009.
- 5 Steurer J. Publikation von Qualitätsdaten: Chancen und Risiken. Bulletin SAMW. 2009;2:1–3.
- 6 Toggweiler S, Puck M, Kucher N, Lüscher TF. Unveröffentlichte Daten 2009.
- 7 Resnic FS, Welt FGP. The public health hazards of risk avoidance associated with public reporting of risk-adjusted outcomes in coronary intervention. J Am Coll Cardiol. 2009;53:825–30.