Klinik für Forensische Psychiatrie
Qualitätssicherung und Forschung

Psychiatrische
Universitätsklinik Zürich

Falsche Freude über selbst versteckte Ostereier

Statistische Fehler in der forensischen Forschung

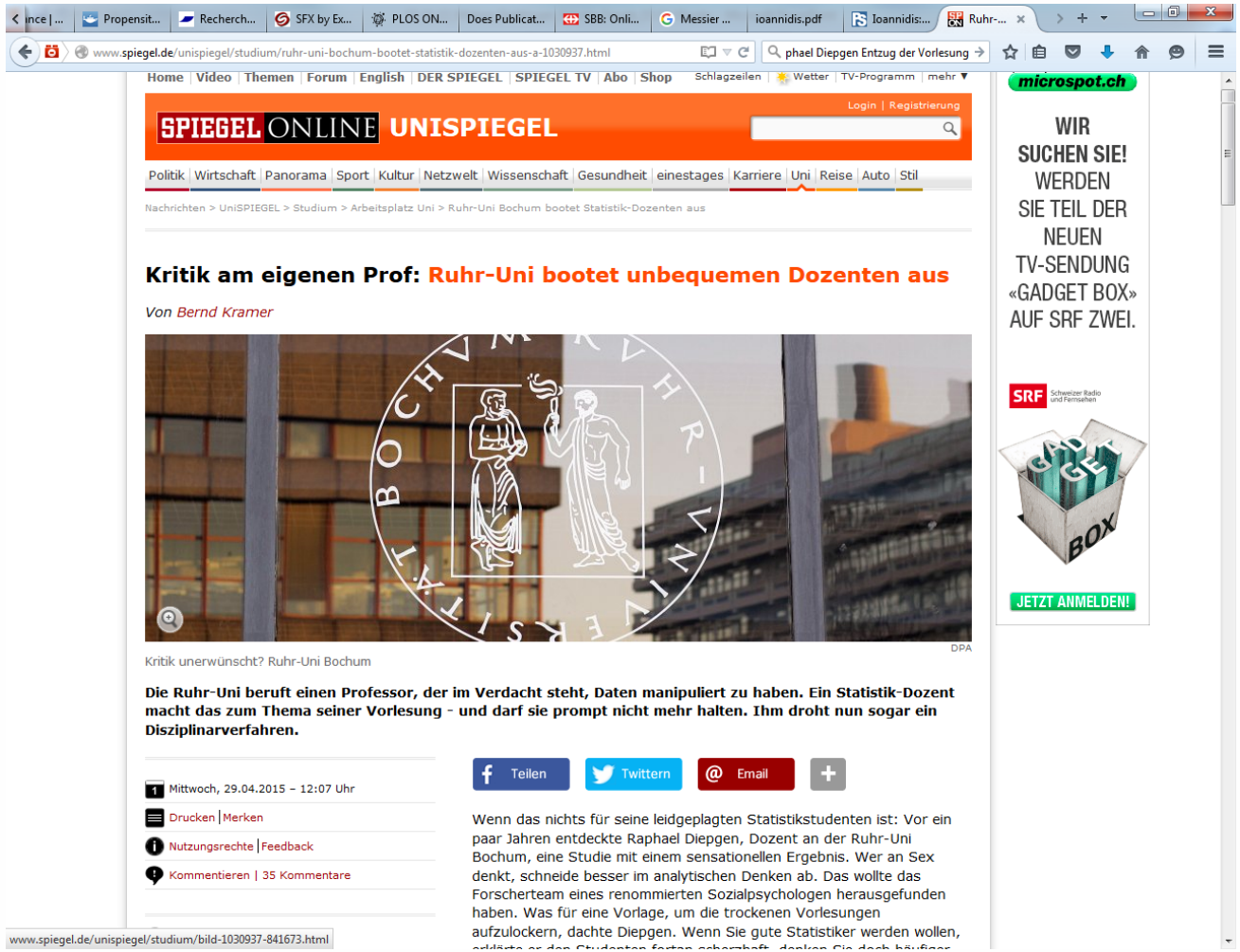
PD Dr. Andreas Mokros

13.11.2015

Universität
Zürich[UZH]

„Gibt es denn niemanden mehr, dem Regeln etwas bedeuten?“ (Walter Sobchak)

Psychiatrische Universitätsklinik Zürich



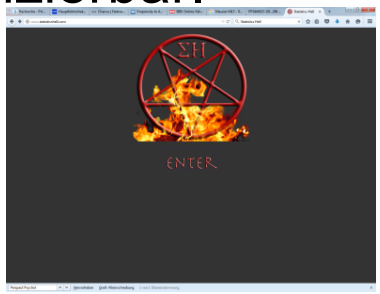
Universität Zürich[UZH]

Themen

Der gegenwärtige Stand

– „Die meisten wissenschaftl. Forschungsergebnisse sind falsch“,

– ... und die Ergebnisse psychol. Experimente kaum replizierbar.



Der Vorhof zur Hölle

– Hypothesentesten als folkloristisches Ritual

– Hypothesenbildung aufgrund Resultatekenntnis (HARK) etc.

Sünden der Forensischen Psychiatrie/Rechtspsychologie

– „Lies, damned lies, and statistics.“ (B. Disraeli)

– Beschränkung auf Extremgruppen

– *N* = ganz wenige etc.

Aus dem Dunkel ins Licht

– Vorschläge zur Verbesserung der Situation

Der gegenwärtige Stand

Eingang zur Hölle

Why most published research findings are false
(John P. A. Ioannidis, 2005)

• *R* = Anzahl der Studien in einem bestimmten Feld, die einen Effekt gezeigt haben, gegenüber solchen, die keinen Effekt gezeigt haben
• *R*/(*R*+1) = WS für das Zutreffen eines Zusammenhangs vor der fraglichen Studie
• WS für das Auffinden eines Zusammenhangs, wenn jener existiert:
$(1 - \beta)\, R/(R - \beta R + \alpha)$ [Quotient wird größer, je kleiner β ist. Der Term ist das Komplement der *False-Discovery Rate*, *FDR*, zu 1.]

$$\frac{R\text{-}\beta R}{R\text{-}\beta R + \alpha} > 0.5, \text{ wenn } R\text{-}\beta R > \alpha, \text{ z.B. bei } \alpha = .05, \beta = .2:$$

$$\frac{(1\text{-}.2)R}{R\text{-}.2 \times R + .05} > 0.5, \text{ wenn } R = .0625 \text{ (bzw. 1 günstige vs. 16 ungünstige Studien)}$$

+ *Bias*, + Anzahl konkurrierender Teams

Theoretische Überlegungen (zu Ioannidis, 2005)

Veranschaulichung

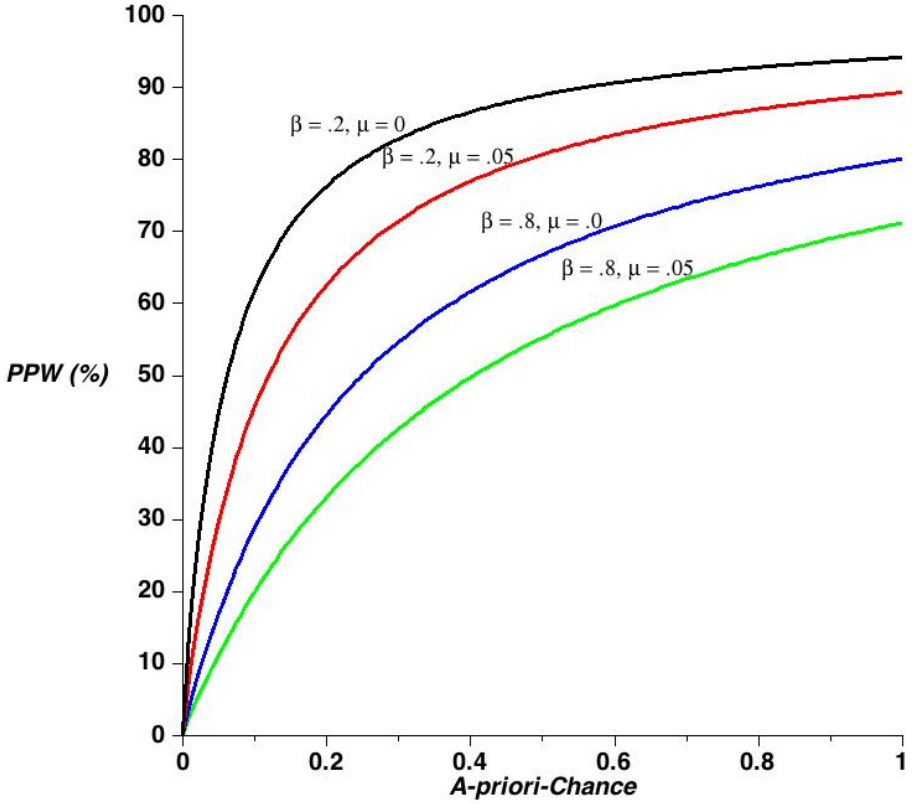
Beispiel:
Teststärke (1- β) = .80

R = 1:10 (zehnmal wahrscheinlicher, dass unzutreffende als dass wahre Zusammenhänge/Effekte gefunden worden sind)

Bias (μ) = 0.3 (Design-, Daten- und Darstellungsfaktoren, die Forschungsergebnisse produzieren, obwohl sie unzutreffend sind)

kritisch hierzu: Jager, L. R. & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics, 15* (1), 1-12. doi: 10.1093/biostatistics/kxt007

= PPW (Positiver Prädiktionswert) 20%
(z.B. explorative epidemiologische Studie mit adäquater Teststärke).
PPW in Abh. von A-Priori-Chance *R* (α = .05)



Kommentar

Hilfreiches Scheitern

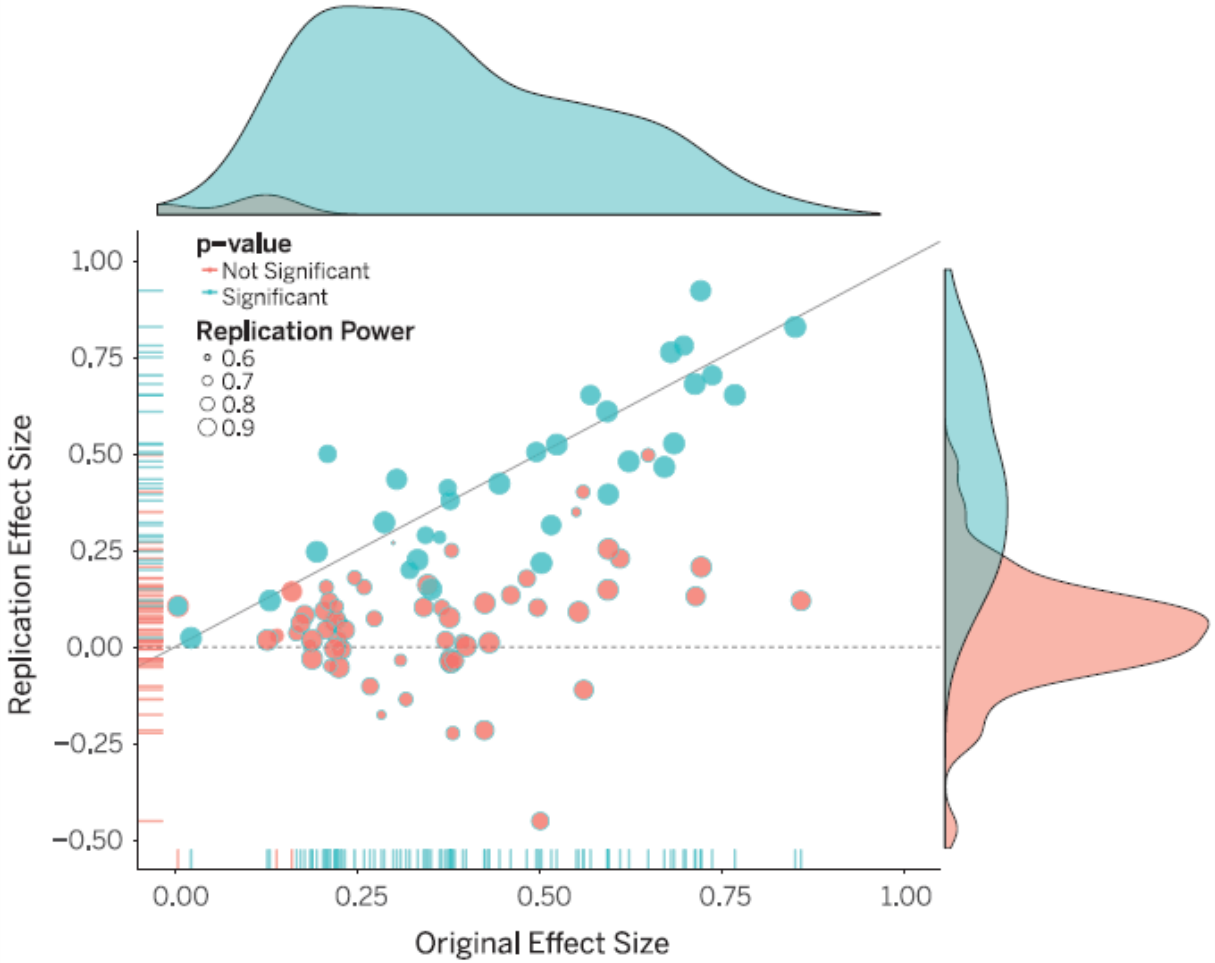
Alle medizinischen Studien müssen veröffentlicht werden.

Die Bilanz war ernüchternd. Mit rund zwei Milliarden Dollar förderte eine Abteilung der amerikanischen National Institutes of Health (NIH) über 200 klinische Studien – doch am Ende wurde mehr als jede dritte dieser Forschungsarbeiten nie oder nur mit extremer Verspätung veröffentlicht. Jetzt ziehen die NIH, die zu den wichtigsten Forschungsförderern der USA zählen, endlich Konsequenzen: Alle von ihnen geförderten Studienergebnisse müssen in Zukunft zumindest im Internet publiziert werden; und in der Herzforschung sollen nur noch solche Projekte unterstützt werden, die gute Chancen haben, den Weg in eine Fachzeitschrift zu finden. Die NIH wollen künftig somit vor allem größere Forschungsvorhaben fördern, die auch wirklich praxisrelevante Fragen untersuchen. Gut so. Es wäre zu hoffen, dass diese Entscheidung weltweit Nachahmer bei anderen Forschungsförderern findet. Auch hierzulande wird permanent Forschungsmüll produziert. Was beispielsweise sollen Allgemeinärzte mit dem Ergebnis anfangen, dass eine neuartige Therapie ein paar nebensächliche Blutwerte senkt? In Tausenden Laboren werden Studien dieser Art angefertigt, deren Erkenntnisgewinn ist nahe null. Immer noch viel zu selten versuchen medizinische Forscher zu klären, ob eine vermeintlich innovative Behandlung tatsächlich Leben retten kann; denn das ist vielen zu aufwendig. Insbesondere könnte die neue Förderpraxis aber auch dazu führen, dass negative Studienergebnisse, mit denen sich Wissenschaftler und Fachzeitschriften bislang ungern rühmen wollten, nicht mehr unveröffentlicht in der Schublade verschwinden. Denn auch das Ergebnis, dass eine Behandlung wirkungslos ist, liefert wertvolle medizinische Hinweise, um Patienten vor therapeutischen Irrwegen zu schützen.

Veronika Hackenbroch
Mail: veronika_hackenbroch@spiegel.de

Estimating the reproducibility of psychological science (Open Science Collaboration, 2015)

Psychiatrische Universitätsklinik Zürich



„Tatsächlich lag die Replikationsrate je nach verwendetem Evaluationskriterium lediglich zwischen 36% und 47%.“ (F. Renkewitz, persönl. Mitteilung, 04.09.2015)

Universität Zürich[UZH]

Süddeutsche Zeitung vom 07.11.2015

Psychiatrische
Universitätsklinik Zürich

PSYCHOTHERAPIE

Schmutzige Geheimnisse

VON CHRISTIAN WEBER

Wer in fremden Schubladen wühlt, findet manchmal schmutzige Geheimnisse. Dummerweise kann es sich dabei um die Wahrheit handeln. Das erfuhr vor Kurzem ein Forscherteam um die klinische Psychologin Ellen Driessen von der Universität Amsterdam. Es war der Frage nachgegangen, was in jenen Studien über Psychotherapie bei Depression stand, die zwischen 1972 und 2008 von den amerikanischen National Institutes of Health finanziert wurden, aber nie in einer Publikation mündeten. Das war bei immerhin 13 von 55 Studien der Fall. Die Forscher baten die damaligen Studienleiter um die Herausgabe der Resultate. Dann verrechneten sie publizierte und unpublizierte Daten in einer Metaanalyse. Das im Wissenschaftsmagazin *Plos One* veröffentlichte Ergebnis ist unerfreulich: Es besagt, dass Psychotherapie bei Depressionen weniger gut hilft als bislang gedacht.

Kontrollierte Studien wären notwendig, doch wie konstruiert man eine Placebo-Therapie?

Nun haben also auch die Psychotherapeuten ihren Schubladen-Skandal, wie man ihn bereits von Antidepressiva und anderen Medikamenten kennt: Wenn nur Studien mit positiven Ergebnissen publiziert, die schlechten Daten aber versteckt werden, dann schleicht sich ein sogenannter Publikationsbias in die Studienlage ein; die Wirkung wird überschätzt. So weit, so schlecht. Doch vielleicht hat die Angelegenheit auch was Gutes: Sie könnte dazu beitragen, die Psychotherapie etwas zu entzaubern.

PLOS ONE

RESEARCH ARTICLE

Does Publication Bias Inflate the Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic Review and Meta-Analysis of US National Institutes of Health-Funded Trials

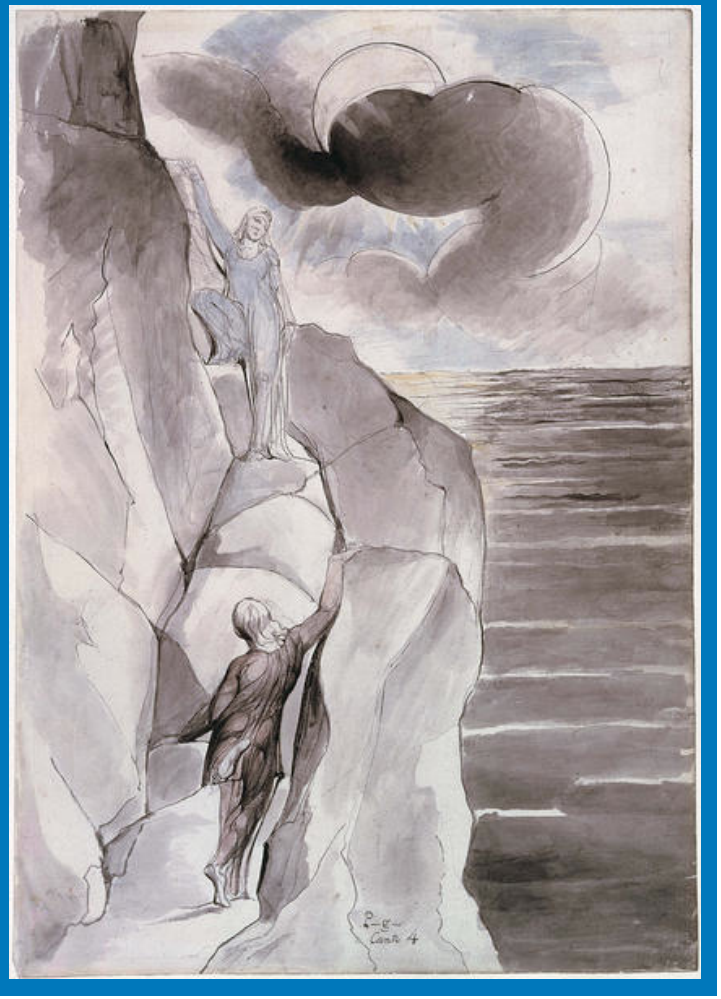
Ellen Driessen[1,2]*, Steven D. Hollon[3], Claudi L. H. Bockting[4,5], Pim Cuijpers[1,2], Erick H. Turner[6,7]

Citation: Driessen E, Hollon SD, Bockting CLH, Cuijpers P, Turner EH (2015) Does Publication Bias Inflate the Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic Review and Meta-Analysis of US National Institutes of Health-Funded Trials. PLoS ONE 10(9): e0137864. doi:10.1371/journal.pone.0137864

Universität Zürich[UZH]

Statistische Missverständnisse

Der Weg zum Purgatorium



Jacob Cohen

„The earth is round ($p < .05$).“

Modus tollens:
„Wenn die Nullhypothese (H_0) zutrifft, kann dieses Ergebnis (D) nicht eintreten.
Das Ergebnis ist eingetreten.
Daher ist die Nullhypothese falsch.“

„Wenn eine Person MarsianerIn ist, ist sie nicht Abgeordnete/r im Bundestag.
Diese Person ist Abgeordnete/r im Bundestag.
Also ist er/sie nicht MarsianerIn.“

(A ➔ B; ¬ B ➔ ¬ A)

In probabilistischer Manier (falsch):
„Wenn die Nullhypothese (H_0) zutrifft, sind diese Daten höchst unwahrscheinlich.
Diese Daten liegen vor.
Deshalb ist die Nullhypothese falsch.“

„Wenn eine Person Deutsche/r ist, ist sie wahrscheinlich nicht Abgeordnete/r im Bundestag.
Diese Person ist Abgeordnete/r im Bundestag.
Also ist er/sie wahrscheinlich nicht Deutsche/r.“

Jacob Cohen

„The earth is round ($p < .05$).“ (Forts.)

Wir interpretieren die Ergebnisse des Nullhypothesen-Signifikanz-Testens (NHST) oft zwar nicht marsianisch, aber Bayesianisch:

p sei die Schätzung für $P(H_0|D)$, also dafür das die Nullhypothese (H_0) korrekt sei, wenn diese Daten D beobachtet worden sind.

Tatsächlich besagt p aber nur:
„Wenn die H_0 zutrifft, wie hoch ist dann die Wahrscheinlichkeit für diese (oder noch extremere) Daten?“

$P(D|H_0) \neq P(H_0|D)$

Diaconis & Freedman (1981; vgl. Eddy, 1982):

Fehler der Transposition der bedingten Wahrscheinlichkeiten

die *Likelihood* $P(D|H)$ mit der Posteriori-WS $P(H|D)$ zu verwechseln, z.B. bei medizinischen Diagnosen:
„Wenn der Test 50% der Kranken tatsächlich erkennt, dann leidet dieser Proband, der ein positives Testergebnis hat, mit 50%-iger Wahrscheinlichkeit an der Krankheit.“

Leider ist die *Likelihood*, Lotto gespielt zu haben, unter der Bedingung, Lottomillionär zu sein, nicht gleich der WS, Lottomillionär zu sein unter der Bedingung, Lotto gespielt zu haben.

Vorschlag

Likelihood und *Bayes*

Dienes (2011, S. 276)
„WS, genau die beobachteten Daten zu erhalten, wenn die Hypothese zutrifft, $P(D|H_1)$."
Likelihood-Quotient bzw. *Bayes*-Faktor (vgl. Goodman, 1999; Wetzels et al., 2011):

$$\frac{P(D \mid H_1)}{P(D \mid H_0)} \qquad \frac{\textit{Sensitivität}}{1 - \textit{Spezifität}} \qquad \frac{\textit{Richtig-Positiv-Rate}}{\textit{Falsch-Positiv-Rate}}$$

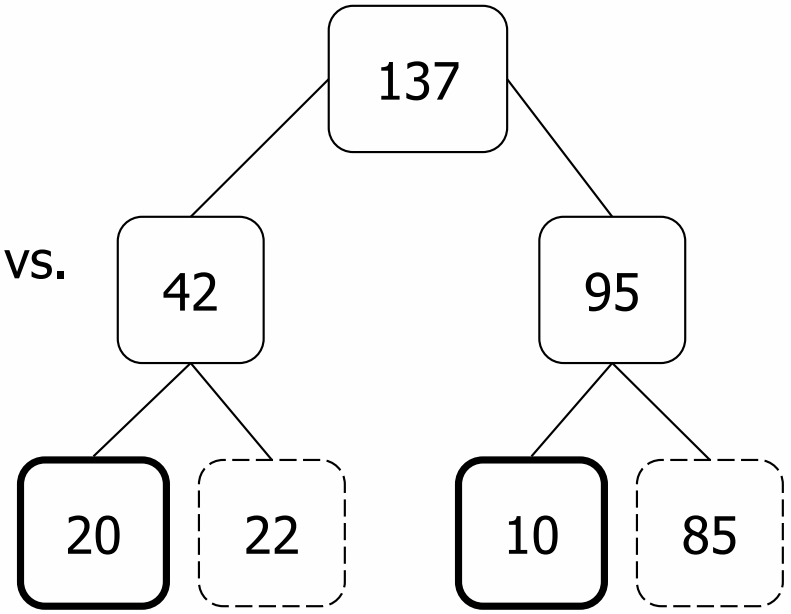
Likelihood-Quotient (LR+) bzw. *Bayes*-Faktor

(Bsp. aus Mokros et al., *Sex Abuse* 2013; 25: 230-258)

Gesamtstichprobe

Kindesmissbraucher (links) vs.
Kontrollprobanden (rechts)

Mit (fett umrandet) vs.
ohne positives/m
Testergebnis (gestrichelt)



LR+: (20/42)/(10/95) = 4,52 (Sensitivität/[1 - Spezifität])

Hintergrund

Das *Bayes*-Theorem, dargestellt über Quoten (engl.: *odds*)

Psychiatrische Universitätsklinik Zürich

$$P(B \mid A) = \frac{P(A \mid B)\,P(B)}{P(A \mid B)\,P(B) + P(A \mid \neg B)\,P(\neg B)}$$

$$\frac{P(B \mid A)}{1 - P(B \mid A)} = \frac{P(A \mid B)\,P(B)}{[P(A \mid B)\,P(B)][1 - P(B \mid A)] + [P(A \mid \neg B)\,P(\neg B)][1 - P(B \mid A)]}$$

$$\frac{P(B \mid A)}{1 - P(B \mid A)} = \frac{P(A \mid B)\,P(B)}{[P(A \mid B)\,P(B)]\left[1 - \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\neg B)P(\neg B)}\right] + [P(A \mid \neg B)\,P(\neg B)]\left[1 - \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\neg B)P(\neg B)}\right]}$$

$$\frac{P(B \mid A)}{1 - P(B \mid A)} = \frac{P(B)}{1 - P(B \mid A)} \times \frac{P(A \mid B)}{P(A \mid \neg B)}$$

Vorschlag

Likelihood und *Bayes*

Dienes (2011, S. 276)
„WS, genau die beobachteten Daten zu erhalten, wenn die Hypothese zutrifft, $P(D|H_1)$."
Likelihood-Quotient bzw. *Bayes*-Faktor:

$$\frac{P(D \mid H_1)}{P(D \mid H_0)} \qquad \frac{\textit{Sensitivität}}{1 - \textit{Spezifität}} \qquad \frac{\textit{Richtig-Positiv-Rate}}{\textit{Falsch-Positiv-Rate}}$$

Posteriori-Chance = LR+ × Priori-Chance

Drei Dilemmata für NHST:

1. Stoppregel: p=.06@N=20, p=.01@N=40
2. Unerwartete Wechselwirkung ➔ Post-Hoc-Interpretation?
3. 5 Tests, einer zu p=.03; Bonferroni?

Bayes-Faktor, Effektstärke und *p*-Wert

(Wetzels et al., 2011, *Perspect Psychol Sci, 6* (3), 291-298)

Psychiatrische Universitätsklinik Zürich

Table 1. Evidence Categories for *p* Values (adapted from Wasserman, 2004, p. 157), for Effect Sizes (as proposed by Cohen, 1988), and for Bayes Factor BF_{A0} (Jeffreys, 1961)

Statistic	Interpretation
p value	
<.001	Decisive evidence against H_0
.001–.01	Substantive evidence against H_0
.01–.05	Positive evidence against H_0
>.05	No evidence against H_0
Effect size	
<0.2	Small effect size
0.2–0.5	Small to medium effect size
0.5–0.8	Medium to large effect size
0.8	Large to very large effect size
Bayes factor	
>100	Decisive evidence for H_A
30–100	Very strong evidence for H_A
10–30	Strong evidence for H_A
3–10	Substantial evidence for H_A
1–3	Anecdotal evidence for H_A
1	No evidence
1/3–1	Anecdotal evidence for H_0
1/10–1/3	Substantial evidence for H_0
1/30–1/10	Strong evidence for H_0
1/100–1/30	Very strong evidence for H_0
<1/100	Decisive evidence for H_0

Note: For the Bayes factor categories, we replaced the label "worth no more than a bare mention" with "anecdotal." Also, in contrast to *p* values, the Bayes factor can quantify evidence in favor of the null hypothesis.

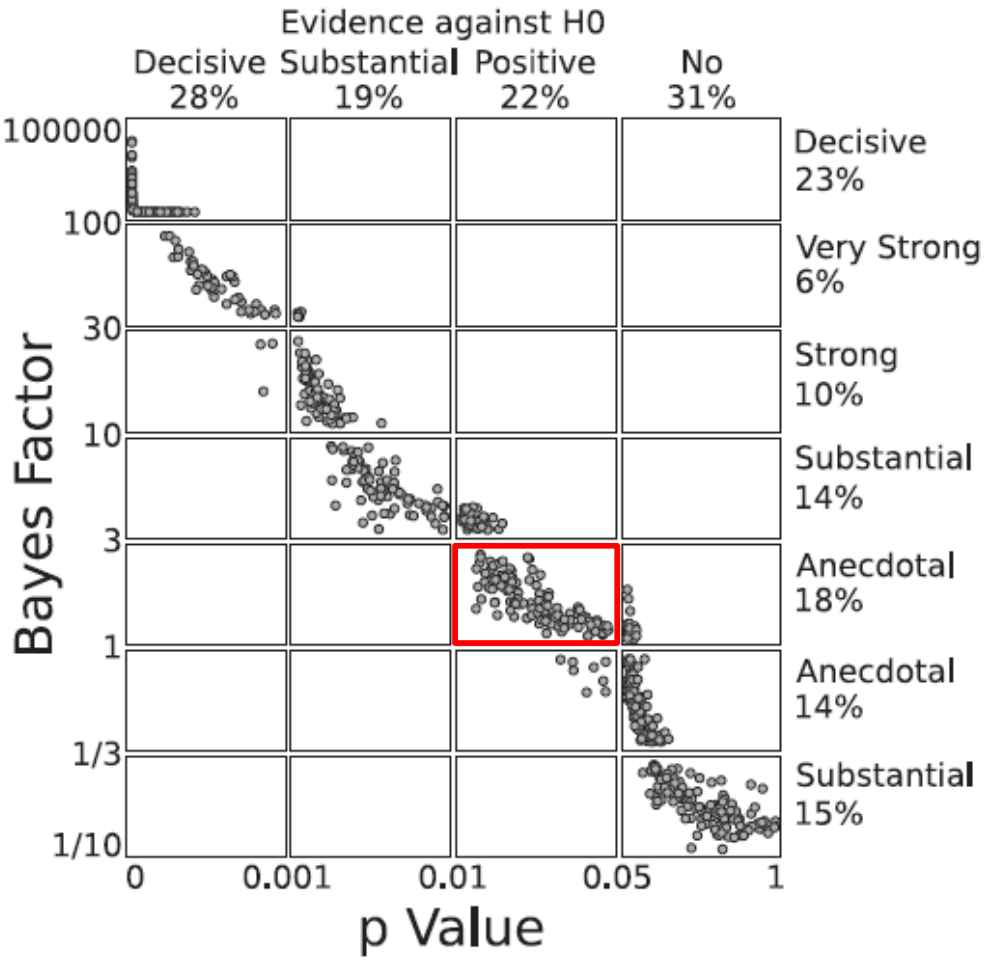


Fig. 3. The relationship between Bayes factor and *p* value. Points denote comparisons (855 in total). The scale of the axes is based on the decision categories, as given in Table 1.

Lindleys Paradox

(aus: Howson & Urbach, 2006, *Scientific Reasoning: The Bayesian Approach* [3rd ed.]. Chicago: Open Court. S. 12*ff*.)

n	**Anteil roter Tulpen an *n*, der die H_0 auf dem 5%-Niveau falsifizieren würde**	**Teststärke bezüglich H_1**
10	.70	.37
20	.60	.50
50	.50	.93
100	.480	.99
1‘000	.426	1.00
10‘000	.4080	1.00
100‘000	.4026	1.00

H_1: Anteil roter Tulpen = .60.
H_0: Anteil roter Tulpen = .40.
Mit zunehmendem *n* nähert sich der kritische Anteil roter Tulpen, bei dessen Überschreitung die H_0 zu verwerfen ist, immer mehr dem in der H_0 spezifizierten Wert an.

„Undisclosed flexibility ... allows presenting anything as significant“ (Simmons et al., 2011)

Psychiatrische Universitätsklinik Zürich

Pbn (*N* = 20) hören entweder „When I‘m sixty-four“ von den Beatles oder „Kalimba“. Dann geben die Pbn ihr Geburtsdatum und das Alter ihres Vaters an. Alter des Vaters wurde als Kovariate berücksichtigt, um Variationen in der Alters-*Baseline* der Pbn zu kontrollieren.

ANCOVA erbringt den erwarteten Effekt: Gemäß Geburtsdaten waren jene, die den Beatles-Song gehört hatten (adj. *M* = 20.1 Jahre), im Schnitt fast 1 ½ Jahre jünger als die übrigen (adj. *M* = 21.5 Jahre; *F*[1,17] = 4.92, *p* = .04).
Der Song machte die Pbn jünger!!!

Auflösung:
-wiederholte Nachtestung nach je 10 Pbn; kein fixes *N* im Vorhinein spezifiziert
- zusätzliche exp. Bedingung (Lied „Hot Potato“) nicht berichtet
-Pbn gaben auch an, wie alt sie sich fühlten (1), wie gerne sie in ein Restaurant gehen würden (2), die Quadratwurzel von 100 (3), ob Computer „komplizierte Maschinen“ seien (4), das Alter ihrer Mütter (5), ob sie Frühbucherrabatte in Anspruch nehmen würden (6), ihre polit. Gesinnung (7), ihre Einschätzung vier kanad. Sportler (8), ob sie die Vergangenheit als „gute alte Zeit“ betrachteten (9) und ihr Geschlecht (10).
-In einer ANOVA (ohne Alter des Vaters als Kovariate) war der Effekt statistisch nicht signifikant (*M* = 20.3 vs. 21.2; *F*[1,18] = 1.01, *p* = .33).

Kugelsternhaufen M13

... und sein Bezug zum multiplen Testen.

Psychiatrische
Universitätsklinik Zürich

z.B. ist es bei 14 durchgeführten Tests (ohne Korrektur) wahrscheinlicher, dass ein signifikantes Ergebnis falsch-positiv ist (Fehler I. Art) als dass es richtig-positiv ist:

$$FWER^* = 1 - (1- \alpha)^k \Rightarrow 1 – (1 - .05)^{14} = .51$$

(vgl. Good, 2005)

* *Family-wise error rate*.

Der tote Lachs im MRT-Scanner

(http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf)

Psychiatrische Universitätsklinik Zürich

Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY; [3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

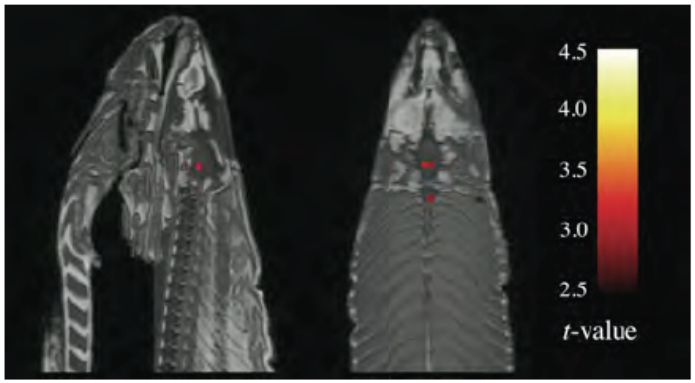
With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we carried out a real experiment that demonstrates the danger of not correcting for chance properly.

METHODS

Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

GLM RESULTS



A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

Several active voxels were discovered in a cluster located within the salmon's

Universität Zürich[UZH]

Der Fall „*psi*“

(Wagenmakers et al., 2011, *JPSP*, 100, 426-432)

Psychiatrische Universitätsklinik Zürich

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 426–432

0022-3514/11/$12.00 DOI: 10.1037/a0022790

Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)

Eric–Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas
University of Amsterdam

Does psi exist? D. J. Bem (2011) conducted 9 studies with over 1,000 participants in an attempt to demonstrate that future events retroactively affect people's responses. Here we discuss several limitations of Bem's experiments on psi; in particular, we show that the data analysis was partly exploratory and that one-sided *p* values may overstate the statistical evidence against the null hypothesis. We reanalyze Bem's data with a default Bayesian *t* test and show that the evidence for psi is weak to nonexistent. We argue that in order to convince a skeptical audience of a controversial claim, one needs to conduct strictly confirmatory studies and analyze the results with statistical tests that are conservative rather than liberal. We conclude that Bem's *p* values do not indicate evidence in favor of precognition; instead, they indicate that experimental psychologists need to change the way they conduct their experiments and analyze their data.

Keywords: confirmatory experiments, Bayesian hypothesis test, ESP

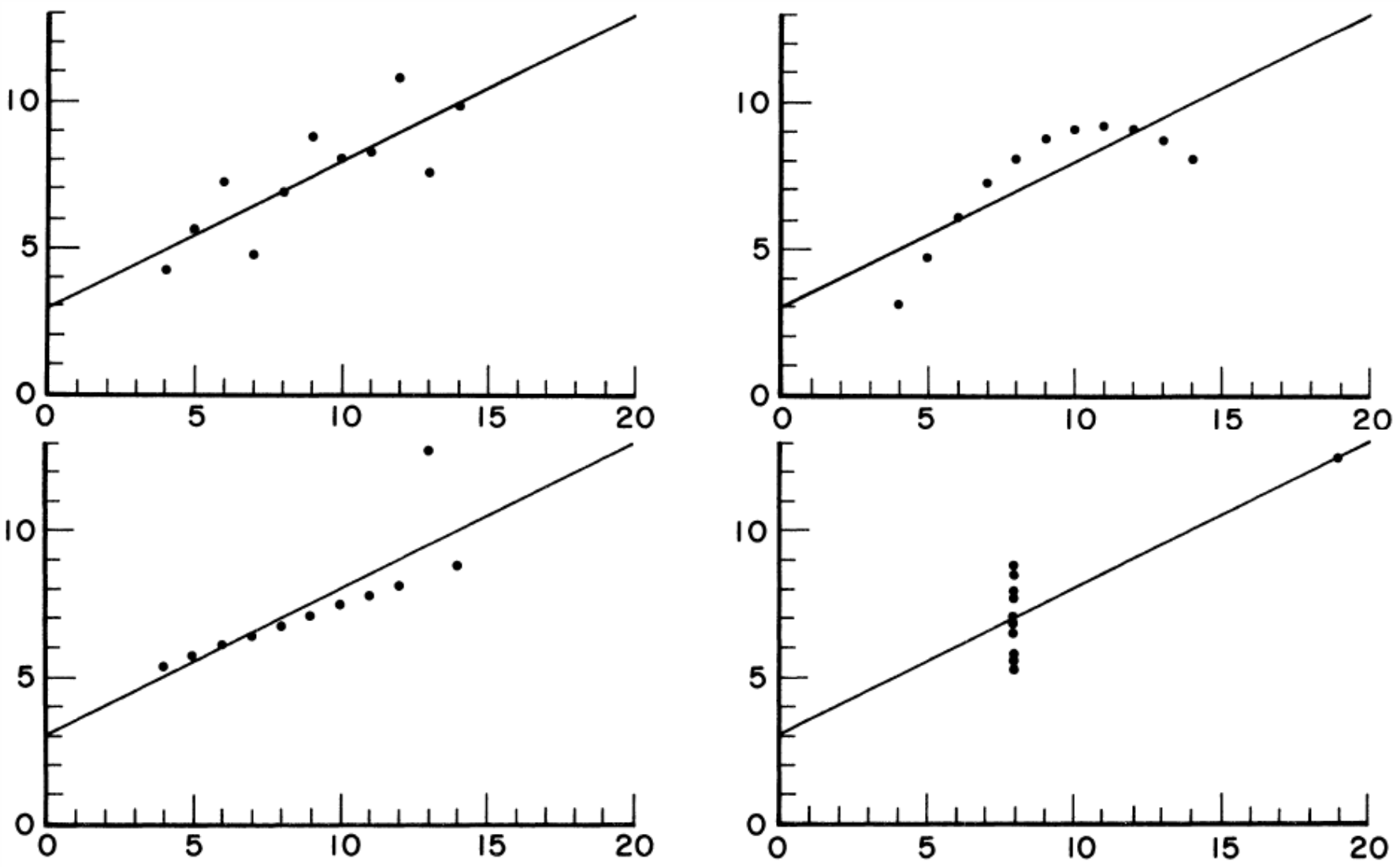
Bem, D.J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*, 407–425.

Universität Zürich[UZH]

Das Quartett von Anscombe (1973, S. 19-20)

r = .82, Mittelwert (*x*) = 9, Mittelwert (*y*) = 7.5, usw.

Psychiatrische
Universitätsklinik Zürich



Vorhersage ≠ Modellanpassung
(Gigerenzer & Brighton, 2009, S. 118)

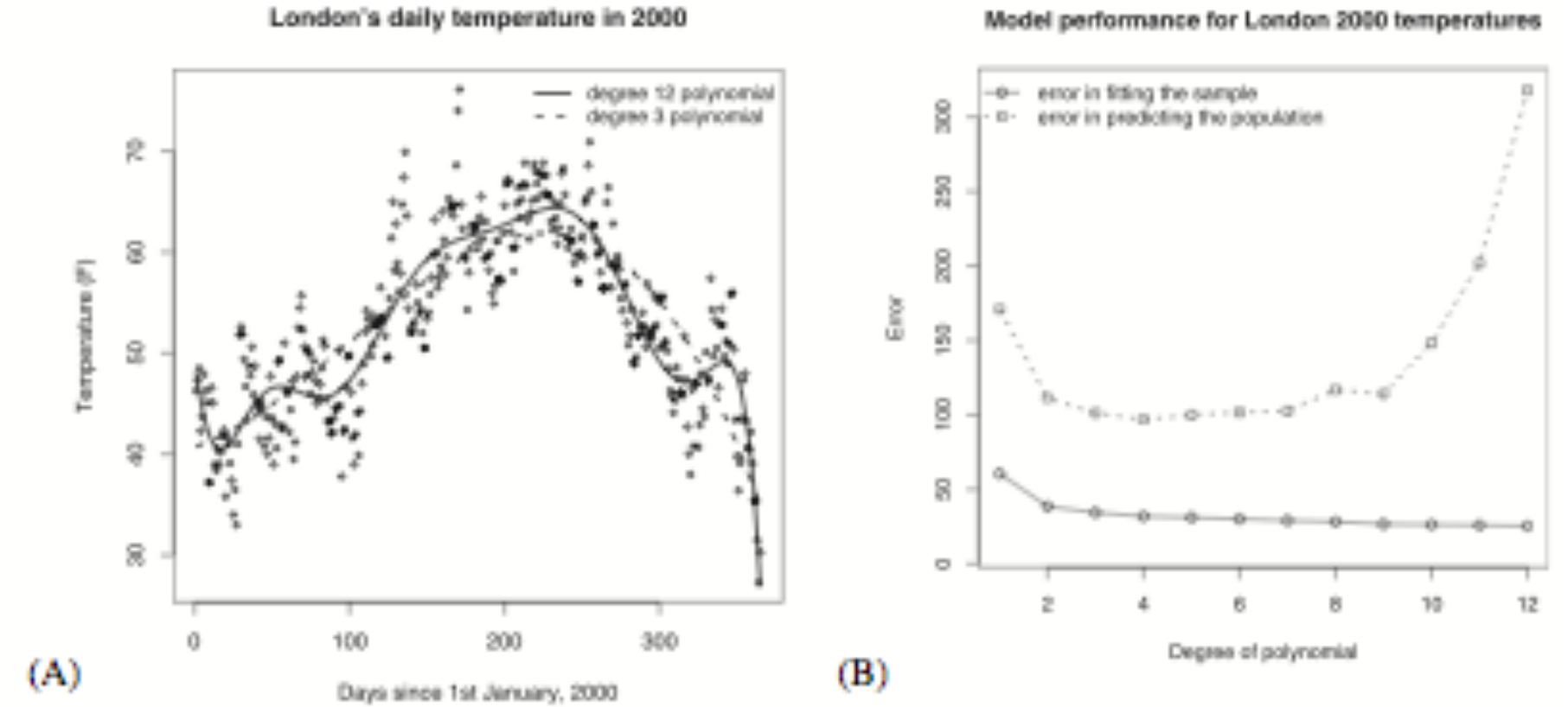


Fig. 3. Plot (A) shows London's mean daily temperature in 2000, along with two polynomial models fitted with using the least squares method. The first is a degree-3 polynomial, and the second is a degree-12 polynomial. Plot (B) shows the mean error in fitting samples of 30 observations and the mean prediction error of the same models, both as a function of degree of polynomial.

Simpsons Paradox

(aus: Senn, 2003, S. 12*ff*.)

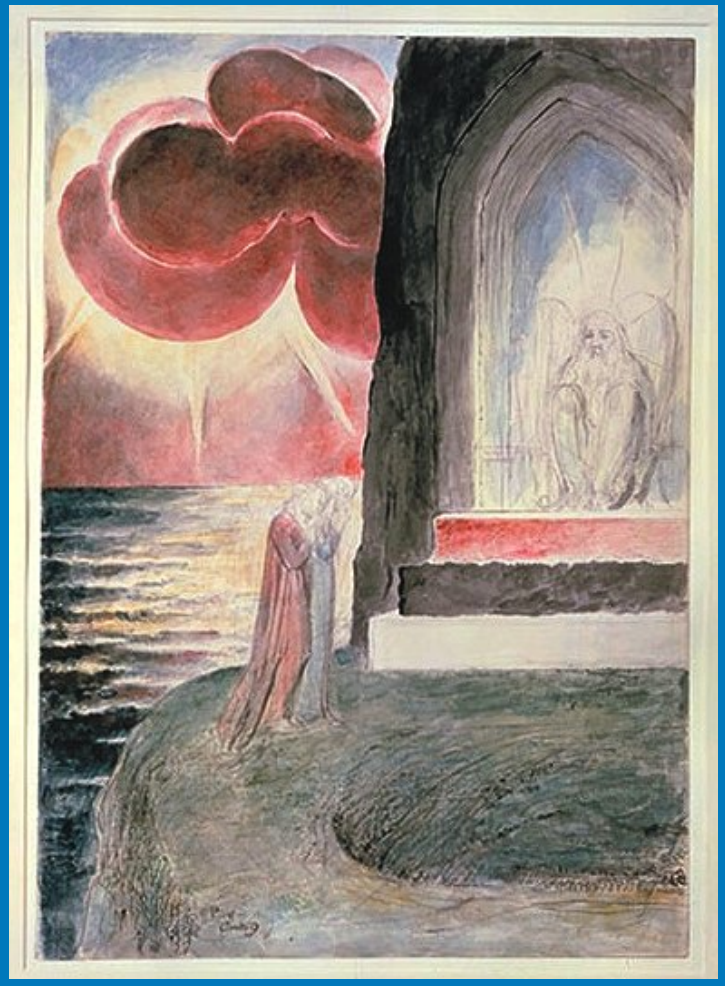
Psychiatrische Universitätsklinik Zürich

	Nicht insulinpflichtig	Insulinpflichtig
Alle Patienten		
Zensoriert	326 (60%)	253 (71%)
Verstorben	218 (40%)	105 (29%)
PatientInnen ≤ 40 Jahre		
Zensoriert	15 (100%)	129 (99%)
Verstorben	0 (0%)	1 (1%)
PatientInnen > 40 Jahre		
Zensoriert	311 (59%)	124 (54%)
Verstorben	218 (41%)	104 (46%)

$\chi_{(1)} = 180.20$, $p = .000$. Nicht insulinpflichtiger Diabetes betrifft vor allem die älteren Pbn. Beachtet man das Alter nicht, wirkt nicht insulinpflichtiger Diabetes gefährlicher.

Sünden der Forensischen Psychiatrie und Rechtspsychologie

Eintritt ins Purgatorium



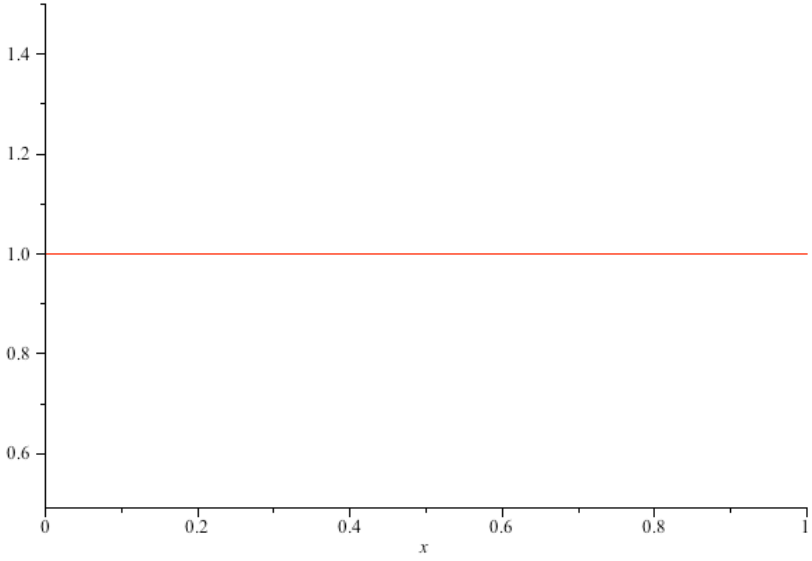
Kredibilitätsintervalle

(Scurich & John, 2012, *Law Human Behav*, 36, 237-246)

Psychiatrische Universitätsklinik Zürich

- frequentistische Konfidenzintervalle sind irrelevant für die Beurteilung des Einzelfalls (wie in der kriminalprognostischen Beurteilung)
  - NICHT: auf lange Sicht enthalten 95% der Konfidenzintervalle den wahren Parameter *p*
  - sondern: Einzelfallbeurteilung (vgl. Meehls Bsp.: Russisches Roulette)

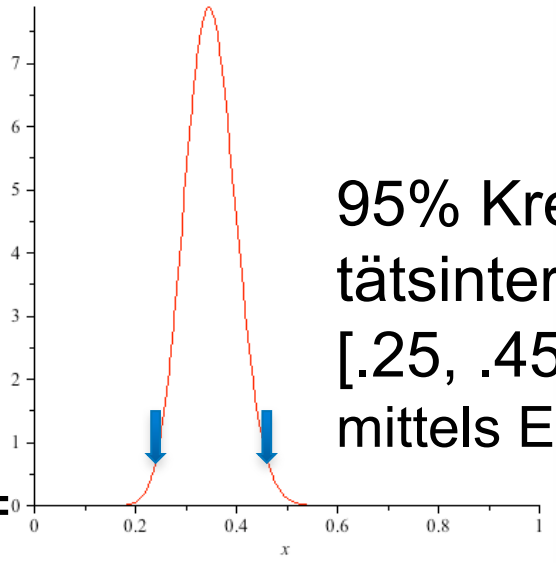
Uniforme Prior-WS-Verteilung



➔„*Update*“ durch *Likelihood*-Funktion*

$L(\theta|x) = f(x|\theta) = p(x|\theta)$

Posteriori-Verteilung



95% Kredibilitätsintervall [.25, .45], z.B. mittels EXCEL

Universität Zürich[UZH]

* WS der beobachteten Daten *x* als Fkt. des festen unbekannten Parameters θ.

Kredibilitätsintervall

(Mossman & Berger, 2001, *Med Decis Making*, 21, 498-507)

Psychiatrische Universitätsklinik Zürich

- Rückfallrate = Basisrate in der Stichprobe (d.h. 24.4%)
- Verwendung einer Monte Carlo-Methode mit Jeffreys-Prior
- *p*(Rückfall | PCL-R Summenwert ≥ 25) = .437
- 95% Kredibilitätsintervall: **[.38, .50]** (Mokros, Vohs & Habermeyer, 2014)

⇒ deutlich schmaleres KI als der Bereich, den Cooke und Michie (2010) für die PCL-R beschrieben haben ([.00, .95])

(s.a. die kritischen Stellungnahmen zu deren Methodik bei Hanson und Howard [2010], Mossman und Sellke [2007] sowie Scurich und John [2012])

Cooke & Michie, 2010, *Law Human Behav*, 34, 259-274
Hanson & Howard, 2010, *Law Human Behav, 34*, 275-281
Hart, Michie, & Cooke 2007, *Brit J Psychiatry, 190* (Suppl. 49), S60-S65
Mossman & Sellke, 2007, *Brit J Psychiatry, 191*, 561
Scurich & John, 2012, *Law Human Behav, 36*, 237-246
Hart & Cooke, 2013, *Behav Sci Law*, *31*, 81-102

Universität Zürich[UZH]

DER SPIEGEL 43/2013:

„Verfehlte Psychotests für Mörder“

Psychiatrische
Universitätsklinik Zürich

BJPsych | The British Journal of Psychiatry (2013)
203, 387–388. doi: 10.1192/bjp.bp.112.118471

Short report

Predicting future violence among individuals with psychopathy

Jeremy W. Coid, Simone Ullrich and Constantinos Kallis

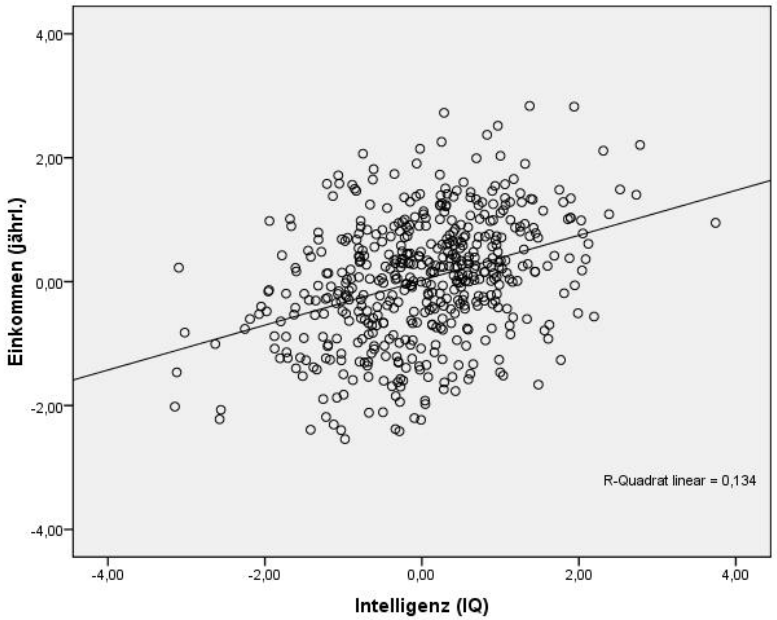
Summary
Structured risk assessment aims to help clinicians classify offenders according to likelihood of future violent and criminal behaviour. We investigated how confident clinicians can be using three commonly used instruments (HCR-20, VRAG, OGRS-II) in individuals with different diagnoses. Moderate to good predictive accuracy for future violence was achieved for released prisoners with no mental disorder, low to moderate for clinical syndromes and personality disorder, but accuracy was no better than chance for individuals with psychopathy. Comprehensive diagnostic assessment should precede an assessment of risk. Risk assessment instruments cannot be relied upon when managing public risk from individuals with psychopathy.

Declaration of interest
None.

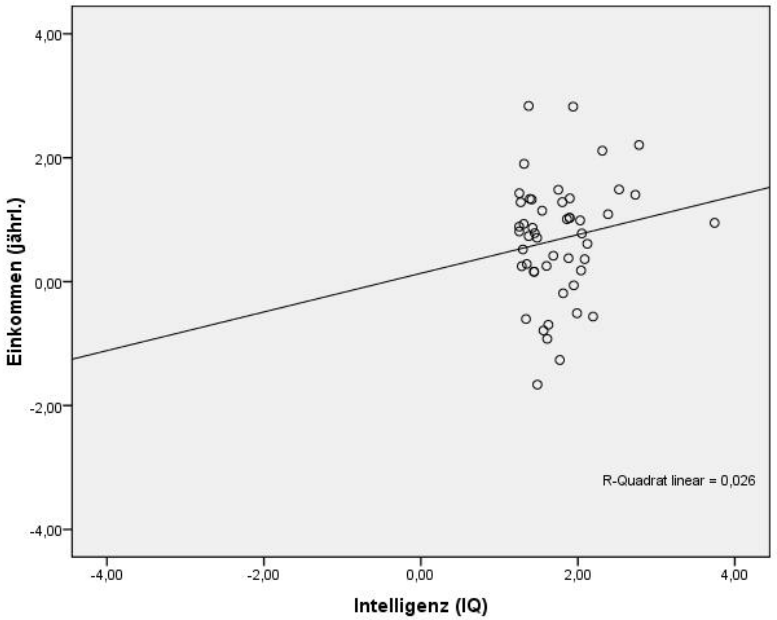
Universität Zürich[UZH]

Warum die Beschränkung auf extreme Sub-gruppen irreführend sein kann

(Bsp. Einkommen und IQ) (vgl. Wottawa, 1980; vgl. Sackett et al., 2007)



alle Probanden
(53 to 156 IQ, *N* = 500)
***r* = .37†** (*p* < .001)
(† simuliert nach Murray, 1998).



nur die schlauesten 10%
(119 to 156 IQ, *n* = 50)
***r* = .16** (*ns*)

Extremgruppen und vermeintliche Behandlungs- oder Remissionseffekte

Psychiatrische Universitätsklinik Zürich

Arch Sex Behav
DOI 10.1007/s10508-015-0652-8

CrossMark

ORIGINAL PAPER

Regression to the Mean Mimicking Changes in Sexual Arousal to Child Stimuli in Pedophiles

Andreas Mokros[1] · Elmar Habermeyer[1]

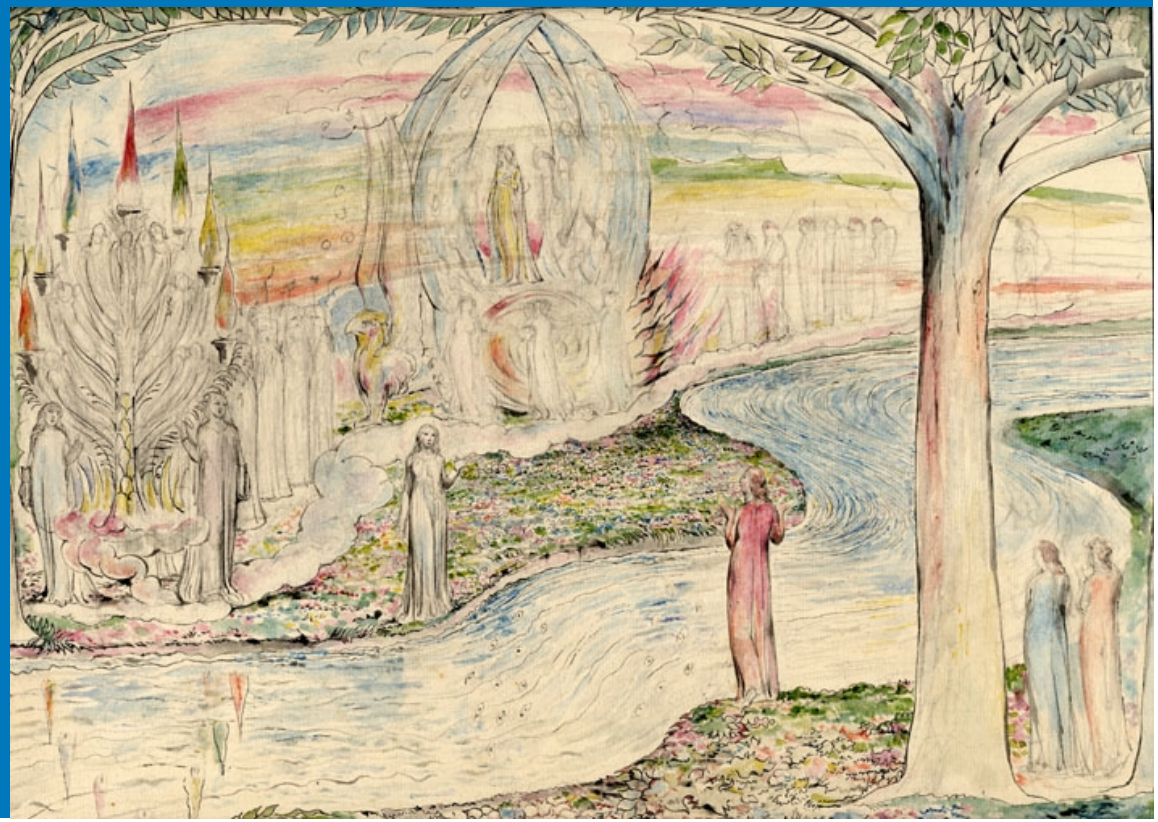
Received: 2 July 2015 / Accepted: 22 October 2015

Abstract The sexual preference for prepubertal children (pedophilia) is generally assumed to be a lifelong condition. Müller et al. (J Sex Med 11:1221–1229, 2014) challenged the notion that pedophilia was stable. Using data from phallometric testing, they found that almost half of 40 adult pedophilic men did not show a corresponding arousal pattern at retest. Critics pointed out that regression to the mean and measurement error might account for these results. Müller et al. contested these

Introduction

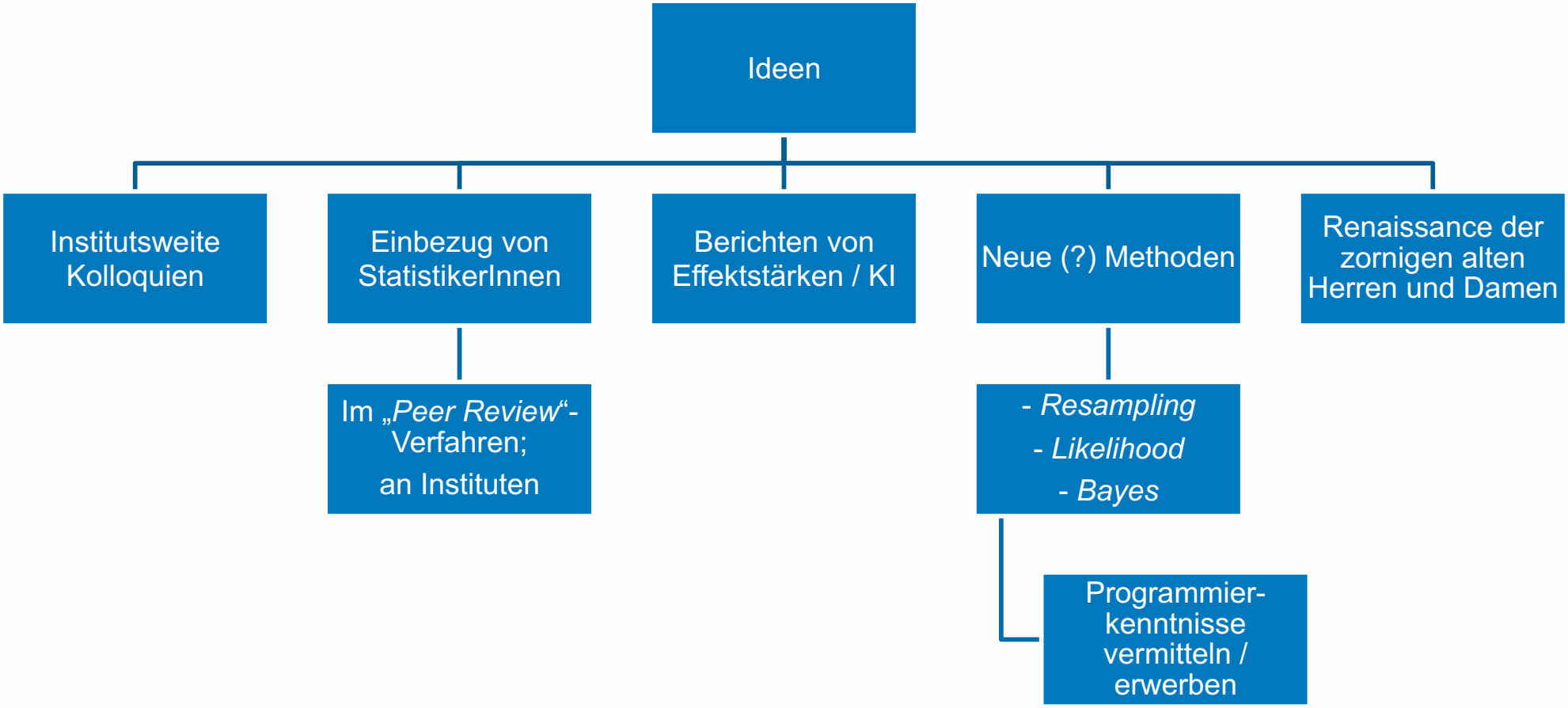
In his book "Thinking, fast and slow," Kahneman (2012) elucidated the phenomenon of regression to the mean with the performance of flight cadets: The ones who showed blunders during one training session will likely perform better (i.e., be closer to the average) on subsequent trials, whereas those who engaged in an excellent training session will probably perform

Aus dem Dunkel ins Licht



Vorschläge zur Verbesserung der Lage

Psychiatrische Universitätsklinik Zürich



Universität Zürich[UZH]

Gängige Effektstärkemaße in der kriminal-prognostischen Forschung (Mokros, 2015)

Psychiatrische Universitätsklinik Zürich

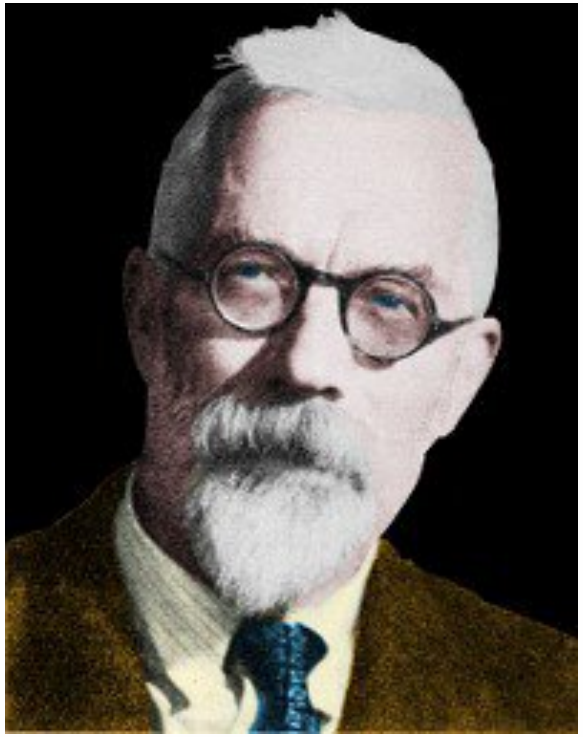
	Einordnung von Effektstärken			Wertebereich	Nulleffekt
	Gering	Mittelgradig	Hoch		
AUC	≥ .56	≥ .64	≥ .71	[0, 1]	.50
d	≥ 0,20	≥ 0,50	≥ 0,80	na [a]	0,00
r	≥ .10	≥ .30	≥ .50	[−1, 1]	.00
r_{pb}	≥ .10	≥ .243	≥ .371	[−1, 1]	.00
LR+	> 2	> 5	> 10	[0, ∞]	1
LR−	≤ 0,5	≤ 0,2	≤ 0,1	[0, ∞]	1
NNT	≤ 8,9	≤ 3,6	≤ 2,3	[1, ∞]	na [b]

vgl. Kraemer et al. (*J Am Acad Child Adolesc Psychiatry* 2003; 42: 1524–29). AUC = *Area under the Curve* (Fläche unter einer ROC-Kurve). *d* = standardisierter Mittelwertunterschied nach Cohen (1992). *r* = Produkt-Moment-Korrelationskoeffizient. r_{pb} = punktbiserialer Korrelationskoeffizient (zwischen einer kontinuierlichen und einer dichotomen Variablen). *LR* = *Likelihood*-Quotient. NNT = *Number needed to treat* (Anzahl der notwendigen Behandlungen). a Keine Mindest- oder Höchstwerte definiert. b Nicht definiert; sehr große Werte würden Nulleffekte nahelegen.

Zitat

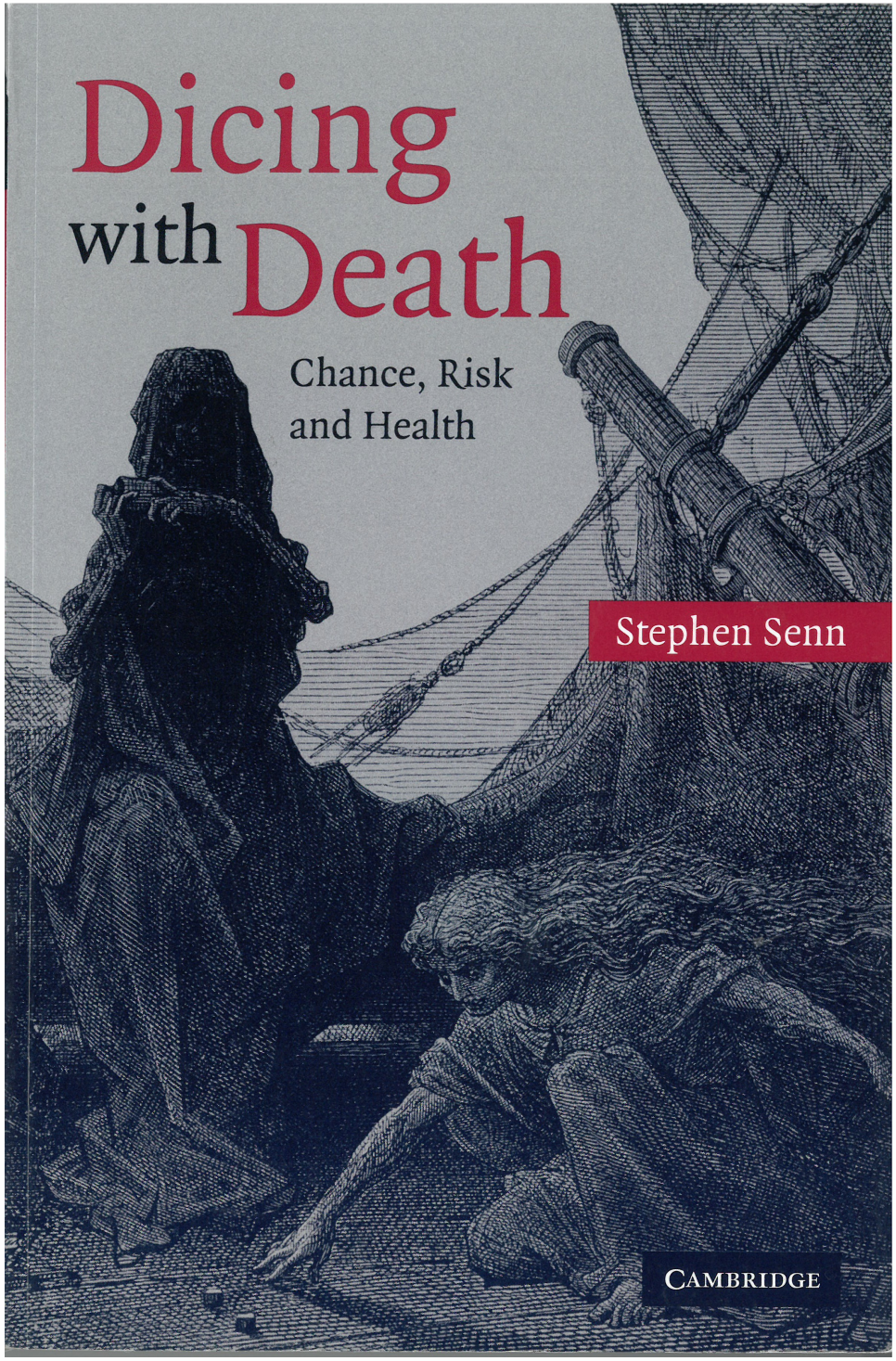
Ronald Aylmer Fisher (1890-1962)

"To consult the statistician *after* an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."



Auswahlbibliografie

- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician, 27* (1), 17-21.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49* (12), 997-1003.
- Diaconis, P. & Freedman, D. (1981). The persistence of cognitive illusions. *Behavioral and Brain Sciences, 4*, 333-334.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6* (3), 274-290.
- Goodman, S. N. (1999). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine, 130*, 1005-1013.
- Guttman, L. (1978). What is not what in statistics. *Statistician, 26* (2), 81-107.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2* (8), e124. doi: 10.1371/journal.pmed.0020124
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349* (6251), 943 (aac4716-1)
- Simmons, J. P., Nelson, L. D. & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22* (11), 1359-1366.



Psychiatrische
Universitätsklinik Zürich

Buchtipp

ISBN: 978-0-521-54023-0

Universität
Zürich[UZH]

Schlusspointe

Weihnachten ist nicht mehr weit ...

Psychiatrische
Universitätsklinik Zürich

BMJ

BMJ 2013;347:f7102 doi: 10.1136/bmj.f7102 (Published 17 December 2013)

RESEARCH

CHRISTMAS 2013: STRANGE NATI

Like a virgin (motl longitudinal, US p survey

OPEN ACCESS

Amy H Herring *professor*[1 2], Sama
William H Joyner *the reverend*[4],

Results

Of 7870 eligible women, 5340 reported a pregnancy, of whom 45 (0.8% of pregnant women) reported a virgin pregnancy (table 1⇓). Perceived importance of religion was associated with virginity but not with virgin pregnancy. The prevalence of abstinence pledges was 15.5%. The virgins who reported pregnancies were more likely to have pledged chastity (30.5%) than the non-virgins who reported pregnancies (15.0%, P=0.01) or the other virgins (21.2%, P=0.007).

Universität Zürich[UZH]

Kontakt

Psychiatrische
Universitätsklinik Zürich