



JACOBS  
UNIVERSITY

Dr. Markus Kühbacher  
Taylorstr. 12a  
14195 Berlin

*Editor's note: We have struck through the word "Fälschungen" -- German for "made up data" -- in this document at the request of Prof. Lerchl. Lerchl has advised us that a court ruled in 2015 that he could not make this claim, and that to avoid financial penalties he should request that anyone who published his claims edit them. While we can't judge the merits of the case in question, and we object in general to Germany's privacy laws, which can inhibit free speech, we are striking through word from this document, first posted by us in 2013, because the risks to Prof. Lerchl seem to outweigh the benefits to our readers of leaving it intact.*

Prof. Dr. Alexander Lerchl  
Professor of Biology

Program Coordinator for  
Biology / Neuroscience

Associate Member, Committee on  
Publication Ethics. London

Editorial Board Member,  
Journal of Algebraic Statistics

Head, Commission on Nonionizing  
Radiation, German Radiation Protection  
Board

School of Engineering and Science  
Campus Ring 6  
D-28759 Bremen, Germany

Phone: +49.421.200-3241  
Fax: +49.421.200-49-3241  
a.lerchl@jacobs-university.de  
www.jacobs-university.de

Bremen, den 19. August 2010

Sehr geehrter Herr Dr. Kühbacher,

gerne komme ich Ihrer Bitte nach, die mir von Ihnen übersandte Excel-Datei auf Auffälligkeiten zu untersuchen. Anmerken möchte ich vorab, dass ich seinerzeit (2008) die ~~Fälschungen~~ an der Medizinischen Universität Wien aufgedeckt habe (REFLEX-Studie), über die anschließend national und international berichtet wurde. Seither beschäftige ich mich intensiv mit Datenfälschungen und ihrem Nachweis.

### I. Allgemeiner Teil:

Zu der Datei (Angaben 2 - 4 entnommen aus „Eigenschaften“):

- 1) Name: 08\_02\_01\_Quantification.xls
- 2) Erstellungsdatum: Freitag, 1. Februar 2008 15:36:14
- 3) Autor: savaskan
- 4) Company: Hifo Uni ZH
- 5) Größe: 30,208 bytes

Die Datei enthält 254 Rohdaten, die in 10 Spalten und bis zu 30 Zeilen angeordnet sind. Die Rohdaten haben 0 (keine) bis maximal 3 Nachkommastellen. Aus den Daten jeder Spalte wurden die Mittelwerte und die Standardabweichungen korrekt mittels der Excel-Funktionen (STABW und MITTELWERT) bestimmt (Zeilen 36 und 37). Weiterhin finden sich Zahlen in den Zeilen 38 und 39, die ohne Berechnung mittels einer Excel-Funktion eingetragen wurden und

offenbar (laut Spalte L) die Mittelwerte, geteilt durch 255 (Zeile 38) bzw. die Standardabweichung (Zeile 39) darstellen sollen.

Aus den Daten in den Zeilen 38 und 39 ist eine in der Datei enthaltene Abbildung hervorgegangen, die grafisch die Werte darstellt. Diese Grafik ist bis auf kleinere Abweichungen nahezu identisch mit der Grafik (Fig. 2 c, S. 21) im Manuskript „Neurochemical imaging of selenium homeostasis in the brain under nutritional deficiency“, welches am 14. Februar 2008 der Zeitschrift „Journal of Neurochemistry“ eingereicht wurde. Laut der im Manuskript enthaltenen Legende sollen die Daten die Mittelwerte + Standardabweichung (S. 18, Zeile 42: „(mean±SD“) darstellen.

## **II. Spezieller Teil**

### **A. Fehlerhafte Transformationen**

Die in der Zeile 38 eingetragenen Mittelwerte, die in der Grafik verwendet wurden, entsprechen nicht den aus den Rohdaten berechneten (Zeile 37), sondern wurden offenbar nicht-linear transformiert. Meine eigenen Berechnungen haben ergeben, dass eine Exponentialfunktion den besten „fit“ für den Vergleich der berechneten und transformierten Daten ergibt (s. Anlage 1). Angewendet wurde tatsächlich die Rechenoperation  $255 / x$  anstelle einer linearen Transformation  $255 - x$ . Der Grund für die Transformation ist mir nicht bekannt.

Es sollten allerdings auch die Standardabweichungen exakt denselben Transformationen unterworfen werden, was aber nicht der Fall ist (s. Anlage 1). Wendet man die Rechenoperation  $255 / x$  auf die zehn Mittelwerte aus Zeile 37 an, so erhält man in acht Fällen die Werte, die auch in Zeile 38 aufgeführt sind. Wendet man die Rechenoperation  $255 / x$  auf alle Rohdaten an und bestimmt anschließend die Standardabweichungen, so stimmen in fünf Fällen die von mir berechneten Werte nicht mit den als Fehlerbalken dargestellten Werten aus Zeile 39 überein (Abweichungen zwischen 5 % und 127 %). Auch bei den Mittelwerten treten in zwei Fällen Abweichungen von immerhin 2,3 % bzw. 7,5% auf. Insofern ist festzuhalten, dass fünf der in der Abbildung gezeigten Standardabweichungen weder der ursprünglichen Berechnung entspringen noch durch Transformationen gewonnen wurden. Diese Werte sind daher als fehlerhaft anzusehen. Außerdem ist festzuhalten, dass von diesen fehlerhaften Werten ein Wert (Zelle I 39) eine als Zahl ungültige Ziffernfolge aufweist und in der Grafik demzufolge keine Standardabweichung eingetragen ist (Säule 9).

Für diese fehlerhaften Werte ist weder ein gemeinsamer systematischer Fehler noch sind „Rundungsfehler“ feststellbar. Diese Unstimmigkeiten stellen bereits starke Indizien für die im folgenden Abschnitt beschriebenen Sachverhalte dar. Weiter ist auffallend, dass die Rohdaten keiner Probennummer zugeordnet worden sind und die Anzahl der Zahlenwerte je Gehirnregion zwischen 22 und 30 Werten variiert.

## B. Hinweise auf Datenfabrikation

Die Rohdaten in der Tabelle enthalten viele Hinweise auf Datenmanipulation bzw. Datenfabrikation (Erfinden von Daten). Es ist wissenschaftlich anerkannt, dass Personen, die Daten erfinden, unbewusst Fehler begehen, indem sie bestimmte Ziffern („Lieblingzzahlen“) unbewusst anderen vorziehen („last digit preference“). Die für die Detektion solcher Ziffernhäufungen angewandte Methode untersucht durch einen Standard-Test ( $\chi^2$ ), ob diese zufällig oder nicht zufällig ist. Hierbei werden die Endziffern ausgewertet. Bei 3 Nachkommastellen wurde die letzte Ziffer gewertet, bei weniger Nachkommastellen eine 0 gezählt.

In der Tabelle sind 254 Rohdaten aufgetragen. Die Endziffern sind wie folgt: 0 (23), **1 (15)**, 2 (37), 3 (31), **4 (15)**, 5 (21), 6 (31), 7 (25), **8 (38)**, 9 (18). Dieses Verteilungsmuster ist hochsignifikant nicht gleichverteilt ( $\chi^2 = 25,7$ ;  $df = 9$ ;  $p = 0.002$ ). Mit anderen Worten: Die Wahrscheinlichkeit für eine zufällige Häufung dieser Art beträgt 0,2%! Dieser Befund ist ein deutlicher Hinweis auf Datenfabrikation.

Zu Ihrer Information habe ich eine Veröffentlichung von einem Kollegen (Statistiker) und mir beigelegt, in dem dieser Sachverhalt ausführlich beschrieben ist; dort finden Sie auch Literaturhinweise.

Dieser wichtige Befund wird durch weitere ergänzt: Die Anzahl Zahlen mit einer 0 am Ende (entspricht 2 anstatt 3 Nachkommastellen) ist 6. Erwartet wird bei 254 Zahlen (zufallsbedingt) jedoch eine viel höhere Zahl von Zahlen mit einer 0 am Ende, nämlich im Durchschnitt 25,4. Diese geringe Anzahl (6 gegenüber 25,4) ist statistisch extrem signifikant ( $p < 0.001$  oder  $< 0,1\%$ ).

Andererseits ist die Anzahl Zahlen mit zwei oder drei Nullen am Ende in der Tabelle sehr hoch, es finden sich 17 Zahlen. Erwartet werden aber nur 2,54 im Schnitt. Auch dieser Unterschied ist stark signifikant ( $p < 0,01$  oder  $< 1\%$ ).

Schließlich finden sich in der Tabelle extrem viele Beispiele für Ziffernwiederholungen, z.B. tauchen „33“ als letzte Endziffern in der 1. Spalte achtmal auf, „17“ taucht sechsmal auf, jeweils in Gruppen. Ähnliche Muster finden sich in Spalte 3: „68“ findet sich viermal, davon dreimal hintereinander, „42“ findet sich dreimal hintereinander. In Spalte 5 ist die „89“ fünfmal zu sehen, dreimal in Folge. Die Wahrscheinlichkeit für ein solches zufälliges Auftreten von identischen Ziffernfolgen, noch dazu in unmittelbarer Folge, ist geringer als 0,0001%.

## C. Zusammenfassung

Die in der Tabelle aufgeführten „Rohdaten“ entstammen mit an Sicherheit grenzender Wahrscheinlichkeit in ihrer Gesamtheit nicht experimentellen Befunden, sondern sind das Ergebnis von Datenfabrikation (Erfinden von Daten). Die hohe

Anzahl an Einzelbefunden schließt jede andere Möglichkeit nach menschlichem Ermessen vollständig aus.

Ich hoffe, Ihnen mit meiner Analyse geholfen zu haben und verbleibe

Mit freundlichen Grüßen

Ihr



(Prof. Dr. Alexander Lerchl)