

# Unbedingte exakte Tests: Eine SAS-basierte Lösung zur effizienten Berechnung für Stichprobengrößen über 1000

Alexander Wagner  
MSD Sharp & Dohme GmbH  
85540 Haar  
Lindenplatz 1  
Deutschland  
alexander.wagner@msd.de

Julia Tewaag  
julia.tewaag@msd.de

Narine Sahakyan  
narine\_sahakyan@msd.de

Monika Scheuringer  
monika\_scheuringer@msd.de

Karl J. Krobot  
karl.krobot@msd.de

## Zusammenfassung

Unbedingte exakte Tests, wie das CSZ Verfahren, werden als "the gold standard for testing association in 2×2 tables" empfohlen, da sie wegen der höheren Testpower aussagekräftiger sind als der exakte Test nach Fisher. Exakte Methoden sollten, entsprechend der vom IQWiG empfohlenen Methoden [4], auch bei der Darstellung der Studienergebnisse in AMNOG-Nutzendossiers bevorzugt werden.

Ein Hindernis für den breiten Einsatz unbedingter exakter Tests sind rechnerische Schwierigkeiten bei der Berechnung der p-Werte. Für Stichproben größer als 1000 Beobachtungen ist derzeit keine SAS-basierte Lösung verfügbar. Dadurch entstammen in AMNOG-Dossiers pharmazeutischer Unternehmen bzw. in Nutzenbewertungen des IQWiG die p-Werte einerseits unbedingten exakten Tests (bei kleinen Stichproben) und andererseits exakten Tests nach Fisher (bei großen Stichproben).

Wir präsentieren eine Lösung zur effizienten Berechnung der p-Werte des unbedingten exakten Tests auch für Stichproben größer als 1000, basierend auf SAS-Makros und den Formeln aus Mehrotra et al.[16]. Die Verfügbarkeit unserer SAS-Makros ermöglicht die Anwendung unbedingter exakter Tests unabhängig von der Stichprobengröße und verbessert damit die Vergleichbarkeit der Testergebnisse innerhalb und zwischen AMNOG-Dossiers oder auch IQWiG-Nutzenbewertungen.

**Schlüsselwörter:** Unbedingte exakte Tests, AutoIt, Makro

## 1 Hintergrund

Unbedingte exakte Tests werden dem exakten Test nach Fisher bei der Berechnung der p-Werte in Kontingenztabellen vorgezogen, da sie wegen der höheren Testpower<sup>1</sup> aussagekräftiger sind ([1]-[3]). Entsprechend der vom IQWiG empfohlenen allgemeinen

---

<sup>1</sup> Testpower: Wahrscheinlichkeit, dass ein Signifikanztest zugunsten einer konkreten Alternativhypothese entscheidet, falls diese richtig ist.

Methoden [4], sollten exakte Methoden auch bei der Darstellung der Studienergebnisse in AMNOG-Nutzendossiers angewandt werden.

Ein Hindernis für die Anwendung unbedingter exakter Methoden, wie das CSZ (Criterion Z for the Chi-squared Test [5]) -Verfahren, ist der aufwendige Rechenprozess zur Ermittlung der p-Werte. Zwar existieren Software-Lösungen für die Berechnung, allerdings gibt es derzeit keine effiziente SAS-basierte Lösung. Eine Übersicht der Funktionalität und der Limitationen der meistverwendeten Software ist in Tabelle 1 dargestellt.

**Tabelle 1:** Vergleich der Softwareprogramme zur Berechnung unbedingt exakter Tests

Software	Testmethodik	Maximale Stichprobengröße (N)	Vorteil	Nachteil
<b>Roger Berger [6]-[8]</b>	<ul style="list-style-type: none"> <li>• Z-pooled exact unconditional test</li> <li>• Z-unpooled exact unconditional test</li> <li>• Boshloo's (1970) exact unconditional test</li> </ul>	1000	kostenlos, kurze Berechnungszeit	nur 3 Tests implementiert. Gesamt N darf nicht größer als 1000 sein
<b>Martin Andres [5], [9]</b>	<ul style="list-style-type: none"> <li>• Barnard real</li> <li>• Barnard's test</li> <li>• Z-pooled exact unconditional test</li> <li>• arcsin Statistics</li> </ul>	400	kostenlos, kurze Berechnungszeit	32Bit Dos Programm, Gesamt N darf nicht größer als 400 sein, Ergebnisse können nicht exportiert werden
<b>R Exact Test [10]</b>	<ul style="list-style-type: none"> <li>• Z-pooled exact unconditional test</li> <li>• Z-unpooled exact unconditional test</li> <li>• Boshloo's (1970) exact unconditional test</li> </ul>	1000	kostenlos	N darf nicht größer als 1000 sein
<b>StatXact 9 ([11]-[13])</b>	<ul style="list-style-type: none"> <li>• Z-pooled exact unconditional test*</li> </ul>	Keine Angabe (eigene Tests: ~1500)	Benutzerfreundliche Oberfläche, Export der Ergebnisse und Input-Daten in MS Excel	kostenpflichtig
*Bezeichnung in StatXact9: Two Binomials: Test of Equivalence				

Durch die Einschränkung der verfügbaren Programme wird bei Stichproben über 1000 auf den exakten Test nach Fischer zurückgegriffen. Um die Vergleichbarkeit von Ergebnissen zu gewährleisten bedarf es einer anwenderfreundlichen Lösung für Stichproben über 1000.

## 2 SAS-basierte Lösung zur effizienten Berechnung von exakten Tests für Stichprobengrößen über 1000

### 2.1 Struktur und Funktionalitäten

Die SAS-basierte Lösung integriert SAS 9.3 (14) und AutoIt 3.3.8 (15) (Betriebssystem: Windows 7.0 64 Bit). AutoIt verwaltet den Prozess, realisiert Input-Output Funktionen, übergibt die Parameter in das SAS-Makro, startet und beendet SAS. SAS berechnet die exakten Tests und liefert die Ergebnisse in RTF-Format.

### 2.2 Programmablauf

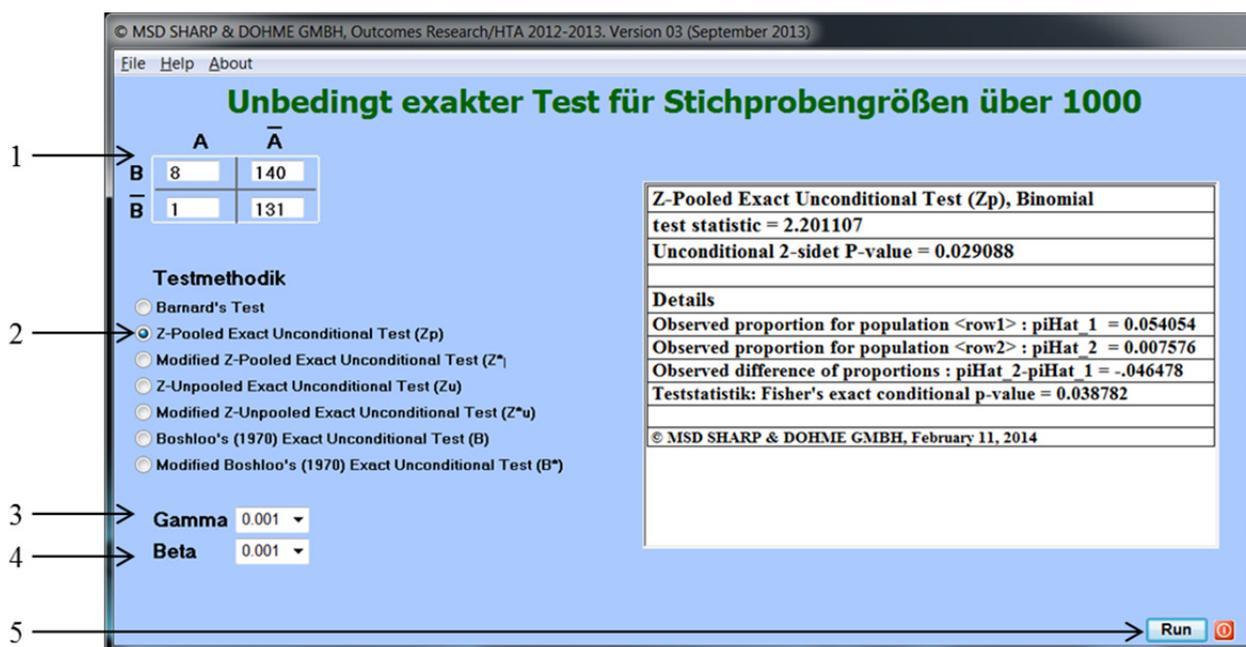
Der Ablauf zur Berechnung des exakten Tests ist wie folgt:

Anwender

- Eingabe der Werte in die 4-Felder-Tafel (1 in Abbildung 1)
- Auswahl des unbedingten exakten Tests (2 in Abbildung 1)
- Wahl des Gamma-Werts (Standardwert: 0.001) (3 in Abbildung 1)
- Wahl des Beta-Werts (Standardwert: 0.001) (4 in Abbildung 1)
- Programmstart durch Wahl des Button „Run“ (5 in Abbildung 1)

Programm

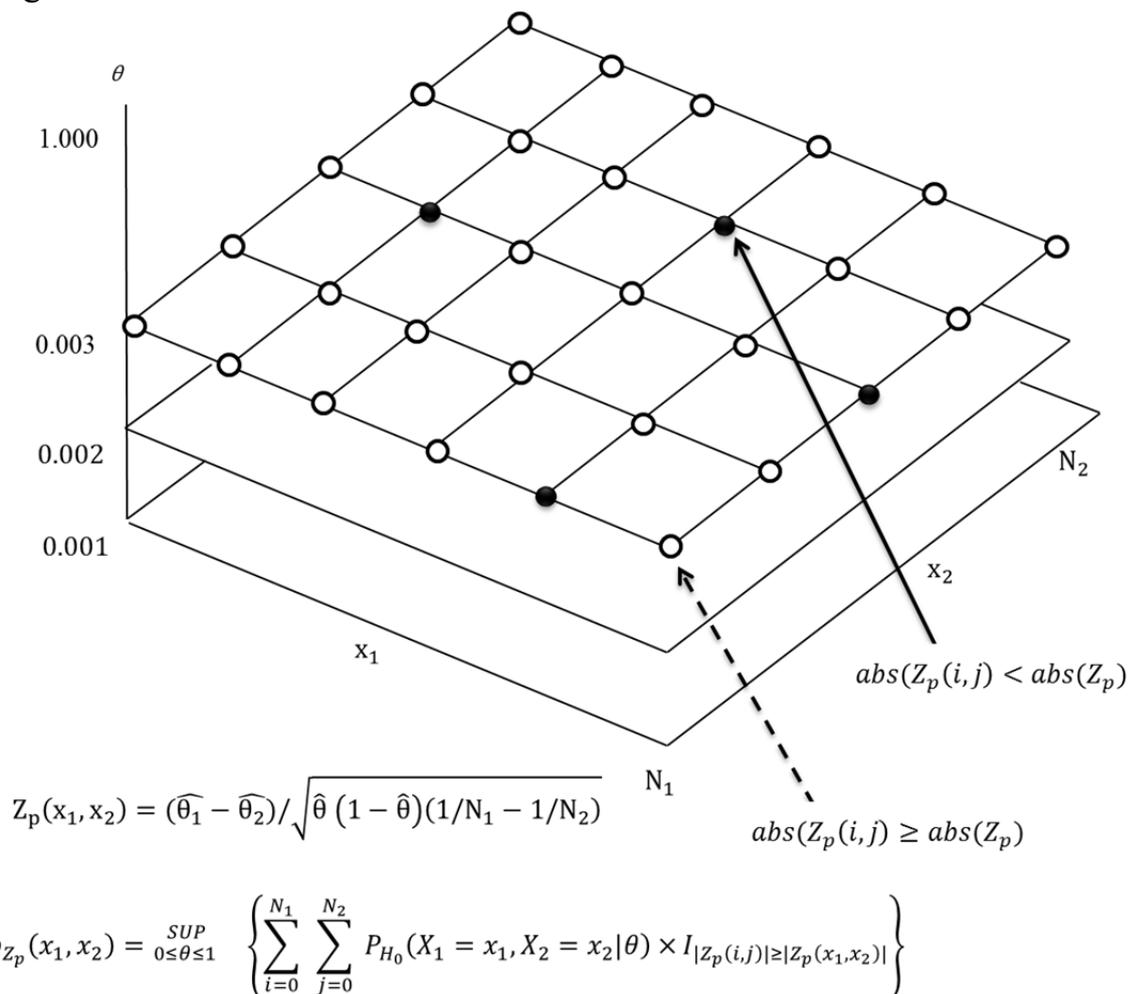
- AutoIt liest die Eingabedaten und definiert Testtyp sowie die Gamma- und Beta-Werte
- AutoIt startet SAS und wählt SAS-Makro aus
- AutoIt übergibt die Parameter an das SAS-Makro und initiiert die Ausführung
- SAS führt den Test durch und liefert die Ergebnisse als externes RTF-Dokument
- AutoIt liest RTF-Dokument in das Menüfenster



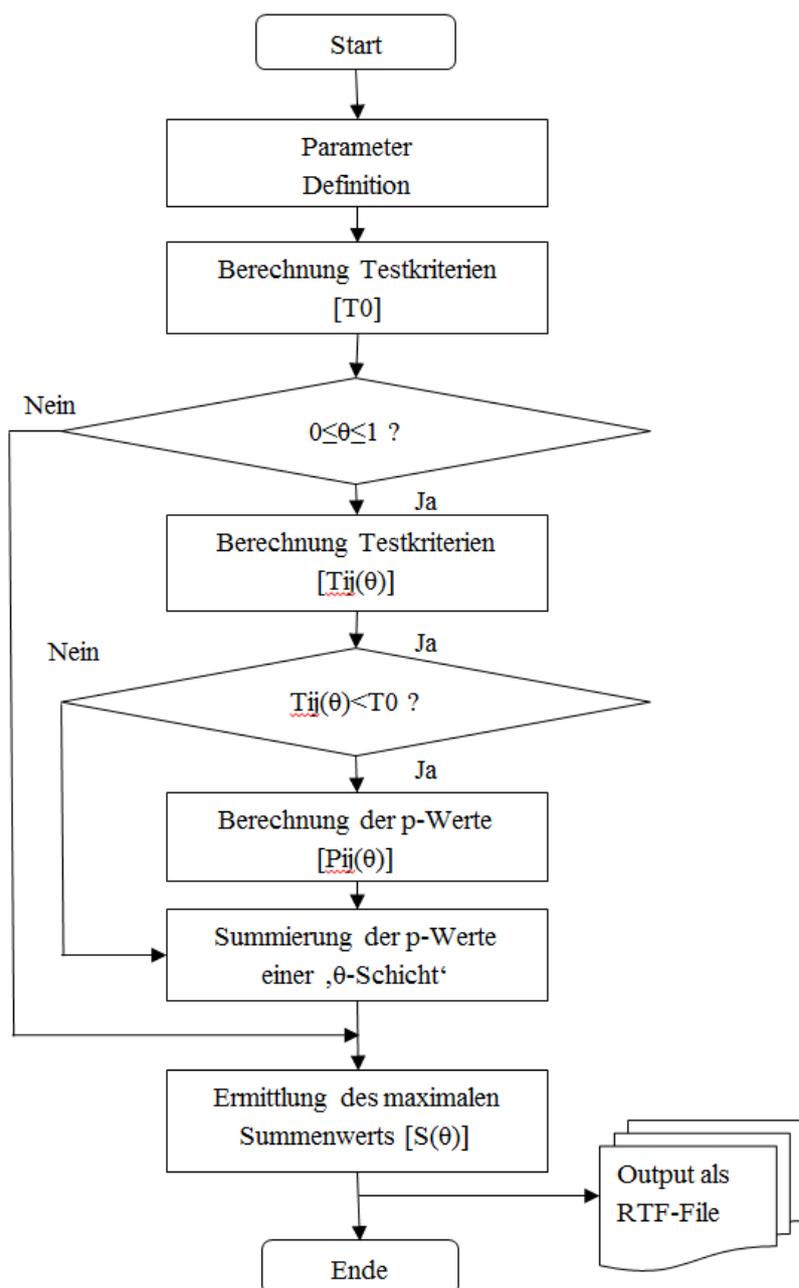
**Abbildung 1:** Eingabemaske für die Parameter zur Berechnung unbedingt exakter Tests sowie Ergebnisse des SAS-Makros

### 2.3 SAS-Makro für exakte Tests

Das SAS-Makro für unbedingte exakte Tests berechnet mit dem durch den Anwender eingegebenen Parameter (siehe Abbildung 1) die Teststatistik. Ziel des SAS-Makro ist es, den maximalen p-Wert für eine gegebene 4-Felder-Tafel mittels eines Optimierungsalgorithmus zu ermitteln. Dazu wird der definierte Bereich von Theta  $[\theta(0,1)]$  in 1000 ‚ $\theta$ -Schichten‘ unterteilt. Anschließend wird für jede Schicht die Wahrscheinlichkeit aller möglichen Variationen der zu untersuchenden 4-Felder-Tafel gegeben dem  $\theta$ -Wert dieser Schicht berechnet (Abbildung 2; eine Kugel entspricht jeweils einer Variation der 4-Felder-Tafel). Die p-Werte jener Punkte, die dem Kriterium  $T_{ij}(\theta) < T_0$  entsprechen, werden für jede der 1000 ‚ $\theta$ -Schichten‘ summiert, um den maximalen Summenwert  $S(\theta)$  zu erhalten. Abbildung 2 veranschaulicht die Ermittlung des p-Werts für den Z-pooled exact unconditional test ( $Z_p$ ): die p-Werte jener Punkte, die dem Kriterium  $abs(Z_p(i,j) < abs(Z_p))$  entsprechen (schwarze Punkte), werden für jede der 1000 ‚ $\theta$ -Schichten‘ summiert. Somit kann der maximale Summenwert  $p_{Z_p}(x_1, x_2)$  ermittelt werden. Der Ablauf des SAS-Makro ist in Abbildung 3 dargestellt; ein Programmausschnitt des SAS-Makro findet sich im Anhang B; die Algorithmen der Teststatistiken in Anhang A.



**Abbildung 2:** Grafische Darstellung der Ermittlung des p-Werts für unbedingte exakte Tests am Beispiel des Z-pooled exact unconditional test (auch CSZ-Methode).



**Abbildung 3:** Ablaufschema der Programmschritte im SAS-Makro

## 2.4 Vergleich mit anderer Software zur Berechnung von exakten Tests

Tabelle 2 zeigt, dass die Berechnungsergebnisse des in diesem Artikel vorgestellten SAS-Makros exakt mit den publizierter Ergebnisse von Berger et al 1996 [6] übereinstimmen.

**Tabelle 2:** Vergleich publizierter Ergebnisse vs. Ergebnisse des in diesem Artikel vorgestellten SAS-Macros

	<b>Methode</b>	<b>Berger et al. 1996</b>	<b>SAS-basierte Lösung 2014</b>
<b>Bedingter exakter Test</b>	Fisher's Test	0.0388	0.038782
<b>Unbedingter exakter Test</b>	Z-Pooled Exact Unconditional Test (CSZ-Methode) ( $Z_p$ )	0.0291	0.02909
	Modified Z-Pooled Exact Unconditional Test ( $Z_p^*$ )	0.0282	0.028158
	Z-Unpooled Exact Unconditional Test ( $Z_U$ )	0.0229	0.022922
	Modified Z-Unpooled Exact Unconditional Test ( $Z_U^*$ )	0.0215	0.02143
	Boshloo's 1970 Exact Unconditional Test (B)	0.0347	0.034695
	Modified Boshloo's 1970 Exact Unconditional Test ( $B^*$ )	0.0325	0.03247
Werte der 4-Felder-Tafel für den Vergleich: A=8, B=148, C=1, D=132			

Tabelle 3 zeigt die Ergebnisse des in diesem Artikel vorgestellten SAS-Makros im Vergleich mit den Ergebnissen von fünf gängigen Softwareprogrammen. Für diese Untersuchung wurden sechs unterschiedliche 4-Felder-Tafeln verwendet und die Methode Z-Pooled Exact Unconditional Test ( $Z_p$ ; CSZ-Methode) angewandt.

### 3 Fazit

Wir präsentieren eine Lösung zur effizienten Berechnung der p-Werte des unbedingten exakten Tests auch für Stichproben größer als 1000, basierend auf SAS-Makros und den Formeln aus Mehrotra et al., 2003(16). Die Testung zeigt, dass die Berechnungsergebnisse des in diesem Artikel vorgestellten SAS-Makros mit den publizierten Ergebnissen übereinstimmen und somit valide ist. Die Verfügbarkeit unserer SAS-Makros ermöglicht die Anwendung unbedingter exakter Tests unabhängig von der Stichprobengröße und verbessert damit die Vergleichbarkeit der Testergebnisse innerhalb und zwischen AMNOG-Dossiers oder auch IQWiG-Nutzenbewertungen.

**Tabelle 3:** Vergleich publizierter Ergebnisse vs. Ergebnisse des in diesem Artikel vorgestellten SAS-Makros sowie der CPU-Laufzeit der Softwareprogramme

<b>4-Felder-Tafel</b>	<b>Berger 1994</b>	<b>Martin 2000</b>	<b>R Aug. 2012</b>	<b>StatXact V9</b>	<b>SAS-basierte Lösung 2014</b>
8, 140 1, 131	0.0291	0.02909	0.0281494	0.02816	0.02815805
25, 400 37, 500	0.6107	*	0.532064	0.5321	0.532063983
18, 740 21, 800	*	*	*	0.835	0.8349454
181, 950 193, 950	*	*	*	0.9294	0.9293936
281, 950 195, 950	*	*	*	0:00142	0.0014198
981, 1950 795, 1950	*	*	*	0.001252	0.001251598
* nicht berechenbar					

## Literatur

- [1] Lydersen S, Fagerland MW, Laake P. Recommended tests for association in 2 x 2 tables. *Statistics in medicine*. 2009 Mar 30;28(7):1159-75. PubMed PMID: 19170020. Epub 2009/01/27. eng.
- [2] Biebler K-E, Wodny M, Jäger B, Kabisch M. Power calculations in SAS of Exact Small-Sample Tests for 2x2-Tables. In: *Simulation, Modeling and Optimization*. 1. Auflage: WSEAS Press; 2009. p. 395-400.
- [3] Galili T. Barnard's exact test - a powerful alternative for Fisher's exact test 2010. Available from: <http://www.r-statistics.com/2010/02/barnards-exact-test-a-powerful-alternative>.
- [4] Allgemeine Methoden: Version 4.1 [Elektronische Ressource] / Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. [https://www.iqwig.de/download/IQWiG\\_Methoden\\_Version\\_4-1.pdf](https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf); 2013.
- [5] Andrés AM, Mato AS. Choosing the optimal unconditioned test for comparing two independent proportions. *Computational Statistics & Data Analysis*. 1994 6//;17(5):555-74.
- [6] Berger RL. Exact Unconditional Homogeneity/Independence Tests for 2X2 Tables 1996. Available from: <http://www4.stat.ncsu.edu/~boos/exact/>.

- [7] Berger RL. More powerful tests from confidence interval p values. *Am Stat.* 1996 Nov;50(4):314-8. PubMed PMID: WOS:A1996VZ17500007. English.
- [8] Berger RL, Boos DD. P Values Maximized Over a Confidence Set for the Nuisance Parameter. *Journal of the American Statistical Association.* 1994;89(427):1012-6.
- [9] Silva Mato, A., Martin Andres, A. SMP. EXE. Version 2.1 2000. Available from: <http://www.ugr.es/~bioest/software.htm>.
- [10] Calhoun P. Package 'Exact'. Version 1.4. Date: 2013-07-21.
- [11] Mehta, C, Patel, N. StatXact 9 User Manual. Cambridge MA: CYTEL Software Corporation; 2010.
- [12] Mehta, C, Patel, N. StatXact 9: Statistical Software for Exact Nonparametric Inference User. Manual. In: Corporation CS, editor. Cambridge MA2010.
- [13] Cytel. StatXact. Version 9 Cambridge2010. Available from: <http://www.cytel.com/software/statxact>.
- [14] SAS Institute. SAS 9.3 Documentation by Title [27.11.2013]. Available from: [http://support.sas.com/documentation/cdl\\_main/93/docindex.html](http://support.sas.com/documentation/cdl_main/93/docindex.html).
- [15] AUTOIT. autoit [27.11.2013]. Available from: <http://www.autoitscript.com>
- [16] Mehrotra DV, Chan IS, Berger RL. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics.* 2003 Jun;59(2):441-50. PubMed PMID: 12926729. Epub 2003/08/21. eng.
- [17] Barnard GA. A new test for 2x2 tables. *Nature.* 1945;156:177.
- [18] Barnard GA. Significance tests for 2x2 tables. *Biometrika.* 1947 January 1, 1947;34(1-2):123-38.
- [19] 19. Gaydyshev IP. Modelling of stochastic and deterministic systems: User Manual 2013 [in russisch], available from: <http://odtdocs.ru/literatura/2305/index.html>.

## Anhang A: Berechnung der p-Werte unbedingt exakter Tests (nach [16])

### Z-Pooled Exact Unconditional Test ( $z_p$ )

Der exakte zweiseitige p-Wert wird wie folgt berechnet:

$$p_{z_p}(x_1, x_2) = \sup_{0 \leq \theta \leq 1} \left\{ \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} P_{H_0}(X_1 = x_1, X_2 = x_2 | \theta) \times I_{|z_p(i,j)| \geq |z_p(x_1, x_2)|} \right\} \quad (4)$$

### Modified Z-Pooled Exact Unconditional Test ( $z_p^*$ )

Der exakte zweiseitige p-Wert wird wie folgt berechnet:

$$p_{z_p^*}(x_1, x_2) = \sup_{\theta_L^y \leq \theta \leq \theta_U^y} \left\{ \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} P_{H_0}(X_1 = x_1, X_2 = x_2 | \theta) \times I_{|z_p(i,j)| \geq |z_p(x_1, x_2)|} \right\} + \gamma \quad (5)$$

$$\theta_L^y = \frac{(x_1 + x_2)}{(x_1 + x_2) + (N_1 - x_1 - x_2 + 1)F\left(2(N_1 - x_1 - x_2 + 1), 2(x_1 + x_2); \frac{\gamma}{2}\right)} \quad (6)$$

$$\theta_U^y = \frac{(x_1 + x_2 + 1)F(2(x_1 + x_2 + 1), 2(N - x_1 - x_2); .5\gamma)}{(x_1 + x_2 + 1)F\left(2(x_1 + x_2 + 1), 2(N - x_1 - x_2); \frac{\gamma}{2}\right) + (N_1 - x_1 - x_2)} \quad (7)$$

### Z-Unpooled Exact Unconditional Test ( $z_U$ )

$$p_{z_U}(x_1, x_2) = \sup_{0 \leq \theta \leq 1} \left\{ \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} P_{H_0}(X_1 = i, X_2 = j | \theta) \times I_{|z_U(i,j)| \geq |z_U(x_1, x_2)|} \right\} \quad (8)$$

$$z_U(x_1, x_2) = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{\hat{\theta}_1(1 - \hat{\theta}_1)}{N_1} + \frac{\hat{\theta}_2(1 - \hat{\theta}_2)}{N_2}}} \quad (9)$$

### Modified Z-Unpooled Exact Unconditional Test ( $z_U^*$ )

$$p_{z_U^*}(x_1, x_2) = \sup_{\theta_L^y \leq \theta \leq \theta_U^y} \left\{ \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} P_{H_0}(X_1 = i, X_2 = j | \theta) \times I_{|z_U(i,j)| \geq |z_U(x_1, x_2)|} \right\} + \gamma \quad (10)$$

### Boschloo's(1970) Exact Unconditional Test (B)

$$p_B(x_1, x_2) = \sup_{0 \leq \theta \leq 1} \{P_{H_0}\{p_F(X_1, X_2) \leq p_F(x_1, x_2)\} | \theta\} \quad (11)$$

*Modified Boschloo's(1970) Exact Unconditional Test (B\*)*

$$p_{B^*}(x_1, x_2) = \sup_{\theta_L^{\gamma} \leq \theta \leq \theta_U^{\gamma}} \{P_{H_0}\{p_F(X_1, X_2) \leq p_F(x_1, x_2)\}|\theta\} + \gamma \quad (12)$$

## Anhang B: Ausschnitt aus dem SAS-Makro zur Berechnung der exakten Tests

```

PROC IML;
  START HAUPT;
  beta=&Beta;
  Data = J(2, 2, 0);
  Data[1, 1] = &A;
  Data[1, 2] = &B;
  Data[2, 1] = &C;
  Data[2, 2] = &D;
  ..... more command line .....

  TXO=TX[data[1,1]+1,data[1,2]+1];
  TX=(I/Cs[1]-J/Cs[2])/sqrt(((I+J)/N)#(1-((I+J)/N))#sum(1/Cs));
  ..... more command line .....
  Par1=sum(data[1,])/ (sum(data[1,])+(N-sum(data[1,])+1)*FINV(1-beta/2,
  2*(N-sum(data[1,])+1), 2*sum(data[1,])));
  Par2=( (sum(data[1,])+1)*FINV(1-beta/2, 2*(sum(data[1,])+1), 2*(N-
  sum(data[1,]))) ) / (N-sum(data[1,])+(sum(data[1,])+1)*FINV(1-
  beta/2, 2*(sum(data[1,])+1), 2*(N-sum(data[1,]))));
  ..... more command line .....
  PROB =J(In,1,0);
  DO I=1 TO In-2;
    DO J=1 TO nTXT;
      X=xInt[I];
      H1=(Tls1[J,1]+Tls2[J,1]-2);
      H2=(N-Tls1[J,1]-Tls2[J,1]+2);
      RP1= LCOMB(X1,X2) + LCOMB(Cs[2], TLS[J,2]-1)+ H1*LOG(X);
      RP2=H2*LOG(1-X);
      xPROB=xPROB+EXP(RP1+RP2);
    END;
    PROB[I]=xPROB;
  END;
  MaxProb=SUM(MAX(PROB), &Gamma);
  RUN HAUPT;
QUIT;

```

Das komplette SAS-Makro kann auf Anfrage vom Erstautor zur Verfügung gestellt werden.