Propensity Score Analyse — eine Fallstudie

Burkhardt Seifert

Biostatistics Unit

University Zürich

Vorspiel auf dem Theater

- **Key Refererenz:**

Rosenbaum, P.R. und D.B. Rubin (**1983**). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*, 41–55.

- 1882 Zitationen (09/09/09)

Vorspiel auf dem Theater

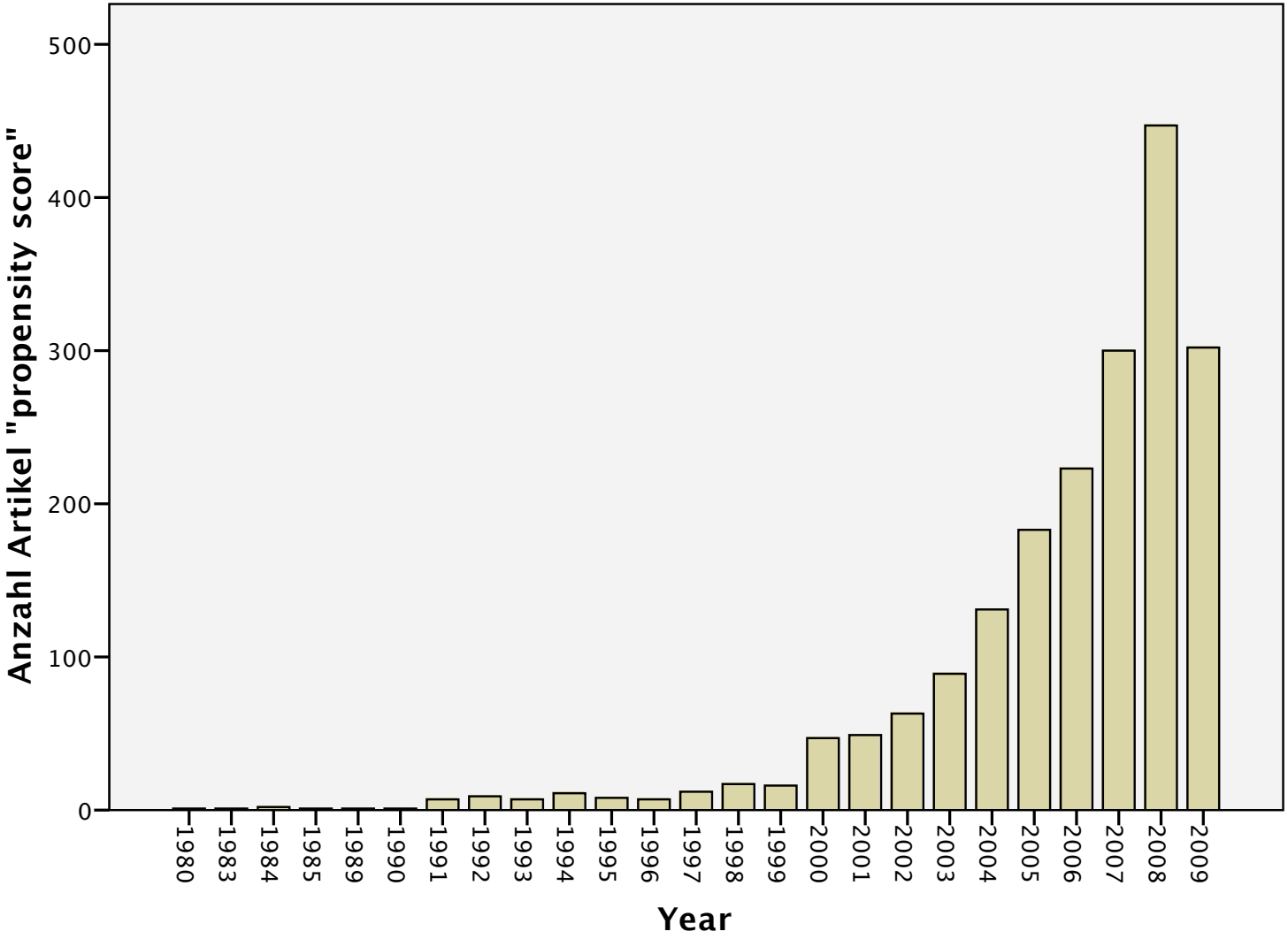
- **Key Refererenz:**

Rosenbaum, P.R. und D.B. Rubin (**1983**). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*, 41–55.

- 1882 Zitationen (09/09/09)

Wieso kennen wir die Propensity-Score Analyse nicht perfekt?

Eine mögliche Antwort:



- mehr als 50% der Artikel seit 2007

Fallstudie

- Analysen November 2007

Received: 16 June 2008

Accepted: 6 August 2008

Primary anastomosis *vs.* Hartmann's procedure in patients undergoing emergency left colectomy for perforated diverticulitis

Urs Zingg, MD,[1,2] Itai Pasternak, MD,[1] Michael Dietrich, MD,[1] Burkhardt Seifert, PhD,[3] Daniel Oertli, MD, FACS,[2] Urs Metzger, MD, FRCS,[1]

- Colorectal Dis. 2008 Oct 1. [Epub ahead of print]

  IF 2.29

Das Problem

- Perforierte Divertikulitis: durchlöcherte entzündete Ausstülpung des Dickdarms (Links-Appendizitis)

→ Notfalloperation

- Standard: Hartmann-OP:
  - Blindverschluss des Enddarms, künstlicher Darmausgang
  - Darmenden in 2. OP nach Abheilung der Erkrankung (ca. 3 Monate) verbunden

- Alternative: direkte Anastomose (Verbindung)
  - Standard bei geplanten (elective) Operationen

- Randomisierte Studie bei Notfalloperation faktisch unmöglich

- 111 aufeinanderfolgende Patienten 2001–2006 Triemli

Postoperative Morbidität

	HP (n=65)	PA (n=46)	P *value*
Overall postoperative morbidity	33 (50.8)	20 (43.5)	0.563
Surgical morbidity	15 (23.1)	17 (37.0)	0.138
rectal stump leak	2 (3.1)	N/A	N/A
anastomotic leak	N/A	13 (28.3)	N/A
intraabdominal abscess	5 (7.7)	7 (15.2)	0.230
wound infection	5 (7.7)	1 (2.2)	0.398
intraabdominal hematoma	0	4 (8.7)	0.027
dehiscence of abdominal fascia	8 (12.3)	3 (6.5)	0.357
Non-surgical morbidity	24 (36.9)	7 (15.2)	0.017

- Mortalität: 19 (29%) vs 8 (17%)

Demographie

	HP (n=65)	PA (n=46)	P *value*
Age (years)	78 (46 - 92)	70 (40-89)	0.002
Female*	40 (61.5)	25 (54.3)	0.558
BMI† (kg/m^2)	25 (15 – 36)	25 (17 – 31)	0.030
Duration of surgery (min)	160 (80-360)	157 (135 – 270)	0.907
Immunosuppression*	22 (33.8)	2 (4.3)	<0.001

- schlechtere Risiken für HP:
  ASA ($p = 0.002$), Hinchey ($p < 0.001$), CCI ($p = 0.03$),
  CR-PPOSSUM ($p = 0.001$), MPI ($p < 0.001$)

Was tun? — Balancing scores

Rosenbaum, P.R. und D.B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*, 41–55.

- 1882 Zitationen (09/09/09)

Blackstone, E.H. (2002). Comparing apples and oranges. *Journal of Thoracic and Cardiovascular Surgery* 123, 8–15.

- "Statistics for the Rest of Us", 193 Zitationen (09/09/09)

Balancing scores

- Wir möchten den Treatment Effekt — die Differenz im erwarteten Outcome zwischen zwei Gruppen — analysieren.
- In randomisierten Studien sind alle klinischen Merkmale auf beide Gruppen gleich verteilt.
- Ein balancing score $b(X)$ ist eine Funktion der Kovariablen X, so dass die bedingte Verteilung von $X|b(X)$ in beiden Gruppen gleich ist.
- Der balancing score erzeugt also lokal eine in den klinischen Merkmalen X balancierte Studie.
- $b(X) = X$ ist ein trivialer balancing score.
- Geschätzte balancing scores verhalten sich wie balancing scores.

Strongly ignorable treatment assignment

- X enthält alle Kovariablen, die zur Entscheidung über die Zuordnung der Gruppen benutzt werden und die möglicherweise mit dem Outcome zusammenhängen:

$$(r_1, r_0) \perp z|v, \quad 0 < pr(z = 1|v) < 1$$

(r: response, z: treatment assignment, v: vector of covariates)

$\rightarrow$ Der erwartete Treatment Effekt von Patienten mit gleichem balancing score ist unabhängig vom balancing score.

$\rightarrow$ Paarweises matching, Stratifizierung und Kovarianzanalyse (covariance adjustment) mit balancing score beseitigen den Selektionsbias.

. . . endlich . . . Propensity Score

- Propensity score: (geschätzte) bedingte Wahrscheinlichkeit, dass ein Patient der Gruppe 1 zugeordnet wird, gegeben X.

Section 2 presents five theorems whose conclusions may be summarized as follows.

(i) The propensity score is a balancing score.

(ii) Any score that is 'finer' than the propensity score is a balancing score; moreover, x is the finest balancing score and the propensity score is the coarsest.

(iii) If treatment assignment is strongly ignorable given x, then it is strongly ignorable given any balancing score.

(iv) At any value of a balancing score, the difference between the treatment and control means is an unbiased estimate of the average treatment effect at that value of the balancing score if treatment assignment is strongly ignorable. Consequently, with strongly ignorable treatment assignment, pair matching on a balancing score, subclassification on a balancing score and covariance adjustment on a balancing score can all produce unbiased estimates of treatment effects.

(v) Using sample estimates of balancing scores can produce sample balance on x.

. . . Hartmann

Multivariate logistic regression using sex, age, BMI, CCI, CR-POSSUM, MPI, ASA score, Hinchey score and immunosuppression was used to determine the propensity score representing the estimated probability that a given individual patient underwent PA based on the variables considered (goodness of fit Hosmer and Lemeshow test chi-square = 7.3 with 8 degrees of freedom, p = 0.51; c-index = 0.91). As CCI has a skew distribution, it was transformed for the computation of the propensity score (log10(CCI+1)). The propensity score then was entered as a covariate into a logistic regression to compare outcomes between groups (HP versus PA). Unadjusted as well as adjusted outcomes (HP versus PA) are presented as odds ratio (OR) with 95 percent confidence interval (CI).

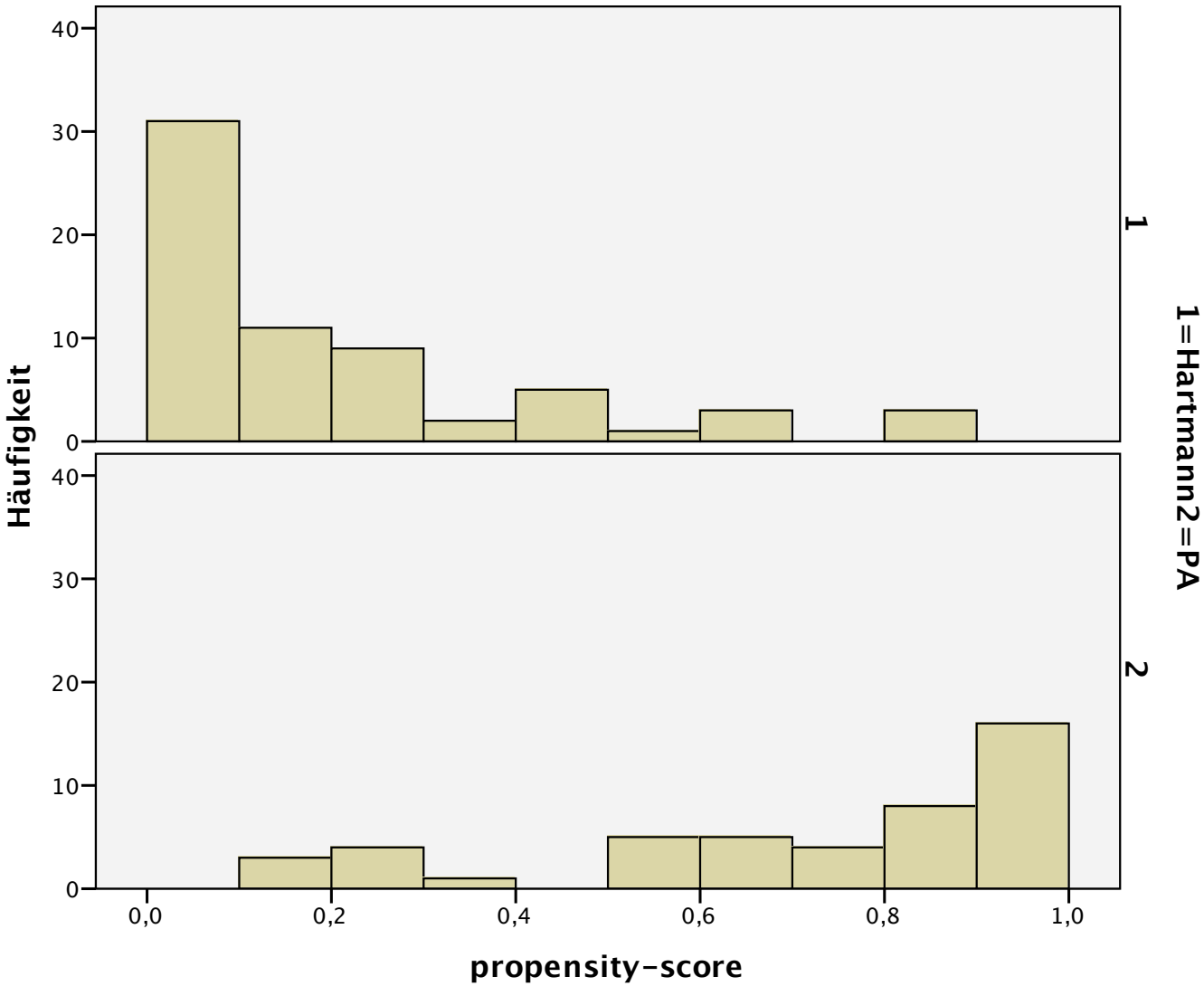
. . . Hartmann

	odds ratio (95% CI)	P *value*
Unadjusted		
mortality		
HP versus PA	1.96 (0.77 - 4.98)	0.156
intraoperative morbidity		
HP versus PA	0.56 (0.19 - 1.67)	0.297
postoperative overall morbidity		
HP versus PA	1.34 (0.63 - 2.86)	0.449
postoperative surgical morbidity		
HP versus PA	0.51 (0.22 - 1.18)	0.114
postoperative non-surgical morbidity		
HP versus PA	3.26 (1.26 - 8.43)	0.015
prolonged hospitalization		
HP versus PA	0.91 (0.43 - 1.94)	0.811

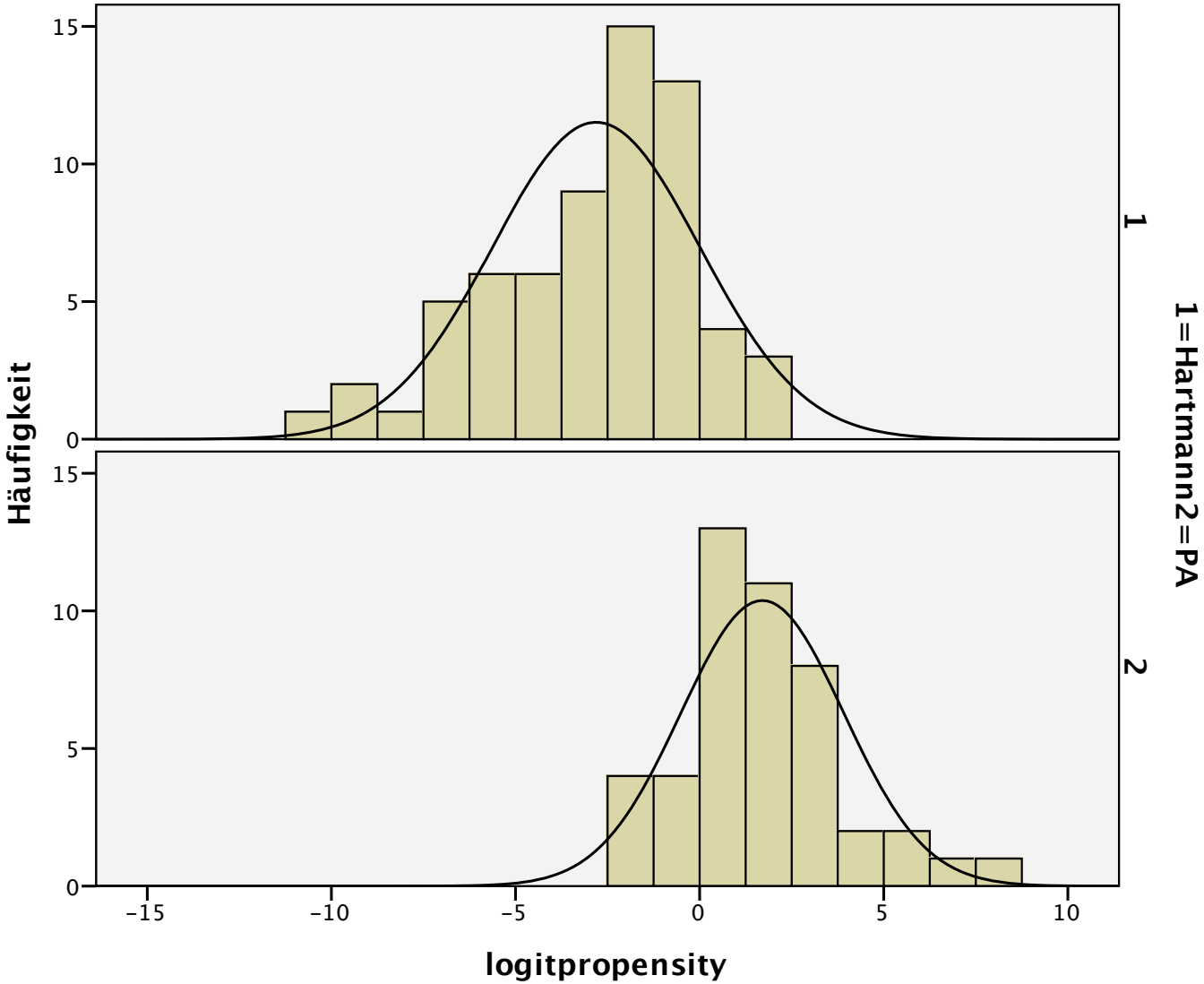
. . . Hartmann

Adjusted		
mortality		
propensity score	0.11 (0.02 - 0.79)	0.028
HP versus PA	0.64 (0.17 - 2.47)	0.522
intraoperative morbidity		
propensity score	4.27 (0.49 - 37.4)	0.190
HP versus PA	1.19 (0.25 - 5.79)	0.826
postoperative overall morbidity		
propensity score	0.47 (0.10 - 2.14)	0.329
HP versus PA	0.90 (0.30 - 2.71)	0.857
postoperative surgical morbidity		
propensity score	1.18 (0.23 - 6.15)	0.841
HP versus PA	0.56 (0.17 - 1.85)	0.342
postoperative non-surgical morbidity		
propensity score	0.20 (0.03 - 1.27)	0.088
HP versus PA	1.46 (0.40 - 5.37)	0.565
prolonged hospitalization		
propensity score	0.06 (0.01 - 0.36)	0.002
HP versus PA	0.19 (0.05 - 0.73)	0.015

. . . Hartmann


Häufigkeit
40
30
20
10
0
1
2
1=Hartmann2=PA
0,0
0,2
0,4
0,6
0,8
1,0
propensity-score

. . . Hartmann


Häufigkeit
15
10
5
0
1
2
1=Hartmann2=PA
-15
-10
-5
0
5
10
logitpropensity

Zerlegung in Quintile

1=Hartmann2=PA * quintile propensity-score Kreuztabelle

Anzahl

		quintile propensity-score						Gesamt
		1	2	3	4	5	6	
1=Hartmann2=PA	1	19	18	12	10	5	1	65
	2	0	1	6	9	13	17	46
Gesamt		19	19	18	19	18	18	111

- Ist der Propensity Score ein balancing score?

$\longrightarrow$ 14 Seiten Output: 4 von 54 Tests signifikant

immunsupp = 1	1	2	3	4	5	6
HP	16	4	0	2	0	0
PA	0	0	1	0	1	0

$\longrightarrow$ unentwirrbares confounding

Ist der odds ratio konstant über die Quintile?

- logistische Regression stetiger propensity score:

  OR = 1.5 (0.40 – 5.4), Hosmer-Lemeshow $p = 0.57$

Anzahl

quintile propensity-score			postopmedicalmorbidity 0	postopmedicalmorbidity 1	Gesamt
1	1=Hartmann2=PA	1	9	10	19
2	1=Hartmann2=PA	1	14	4	18
		2	1	0	1
3	1=Hartmann2=PA	1	7	5	12
		2	4	2	6
4	1=Hartmann2=PA	1	6	4	10
		2	8	1	9
5	1=Hartmann2=PA	1	4	1	5
		2	10	3	13
6	1=Hartmann2=PA	1	1	0	1
		2	16	1	17

- logistische Regression Quintil propensity score:

  OR = 1.9 (0.57 – 6.5), Hosmer-Lemeshow $p = 0.99$

Tests auf Homogenität des Quotenverhältnisses

	Chi-Quadrat	df	Asymptotische Signifikanz (zweiseitig)
Breslow-Day	1,487	4	,829
Tarone	1,486	4	,829

Tests auf bedingte Unabhängigkeit

	Chi-Quadrat	df	Asymptotische Signifikanz (zweiseitig)
Cochran	1,128	1	,288
Mantel-Haenszel	,543	1	,461

Unter Annahme der bedingten Unabhängigkeit ist die Cochran-Statistik nur dann als Chi-Quadrat-Verteilung mit 1 Freiheitsgrad asymptotisch verteilt, wenn die Anzahl der Schichten festgelegt ist. Die Mantel-Haenszel-Statistik ist unter dieser Annahme jedoch immer als Chi-Quadrat-Verteilung mit 1 Freiheitsgrad asymptotisch verteilt. Beachten Sie, daß die Kontinuitätskorrektur aus der Mantel-Haenszel-Statistik entfernt wird, wenn die Summe der Differenzen zwischen den beobachteten und erwarteten Größen 0 ist.

Schätzung des gemeinsamen Quotenverhältnisses nach Mantel-Haenszel

Schätzung			,510
ln(Schätzung)			-,674
Standardfehler von ln(Schätzung)			,630
Asymptotische Signifikanz (zweiseitig)			,285
Asymptotisches 95% Konfidenzintervall	Gemeinsames Quotenverhältnis	Untergrenze	,148
		Obergrenze	1,752
	ln(gemeinsames Quotenverhältnis)	Untergrenze	-1,908
		Obergrenze	,561

Die Schätzung des gemeinsamen Quotenverhältnisses nach Mantel-Haenszel ist unter der Annahme des gemeinsamen Quotenverhältnisses von 1,000 asymptotisch normalverteilt. Dasselbe gilt für den natürlichen Logarithmus der Schätzung.

Der Vorhang fällt und alle Fragen offen . . .

- Wieso ist der Propensity Score ein balancing Score?
  - im linearen Modell bekannt
- Wann ist ein non-parsimonious Modell für den Propensity Score gut?
  - Nicht integrierte Variable können nicht ausbalanciert werden.
  - Ist das ein Vorhersageproblem?
- Welcher c-Index ist optimal?
- Soll man randomisierte Trials mit Propensity Score analysieren?
  - Geschichtet ist besser balanciert als randomisiert.