

Die Berechnung adjustierter NNEs in Kohortenstudien

Ralf Bender
Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)
Dillenburger Str. 27
51105 Köln
Ralf.Bender@iqwig.de

Mandy Hildebrandt
Institut für Medizinische Biometrie,
Epidemiologie und Informatik
(IMBEI)
Johannes-Gutenberg-Universität
Mainz
51101 Mainz
Mandy.Hildebrandt@iqwig.de

Oliver Kuß
Institut für Medizinische Epidemio-
logie, Biometrie und Informatik
(IMEBI)
Martin-Luther-Universität Halle-
Wittenberg
06097 Halle (Saale)
Oliver.Kuss@medizin.uni-halle.de

Zusammenfassung

Das Maß "number needed to treat" (NNT) wird seit einigen Jahren vor allem in der Ergebnisdarstellung randomisierter klinischer Studien immer häufiger verwendet, um die Stärke des Effekts einer neuen Behandlung zu beschreiben. NNT gibt die durchschnittliche Anzahl von Patienten an, die über einen bestimmten Zeitraum die neue Behandlung erhalten müssen, um ein (ungünstiges) Ereignis zu verhindern. Diese Ergebnisdarstellung kann prinzipiell immer dann verwendet werden, wenn die Risiken für das Eintreten eines Ereignisses zwischen zwei Gruppen verglichen werden. Allerdings macht die Bezeichnung "number needed to treat" keinen Sinn, wenn in der Studie gar keine Behandlung untersucht wird. Zum Vergleich der Risiken exponierter und nicht exponierter Personen in epidemiologischen Studien kann die Bezeichnung "number needed to be exposed" (NNE) sinnvoll verwendet werden. Während in randomisierten klinischen Studien mit binärer Zielvariable die Berechnung von NNTs auf der Basis einfacher Vierfeldertafeln erfolgen kann, werden in klassischen epidemiologischen Studien in der Regel Verfahren benötigt, die den Effekt von Confoundern berücksichtigen. Methoden zur Berechnung adjustierter NNEs in epidemiologischen Studien können im Rahmen der multiplen logistischen Regression entwickelt werden. Da sich die geschätzten adjustierten NNEs als Funktionen der logistischen Regressionskoeffizienten schreiben lassen, können approximative Konfidenzintervalle mit Hilfe der multivariaten Delta-Methode hergeleitet werden. Für die Berechnungen steht allerdings keine Standard-Software zur Verfügung. In diesem Beitrag wird ein SAS/IML[®]-Programm vorgestellt, mit dessen Hilfe adjustierte NNEs mit Konfidenzintervallen im Rahmen von Kohortenstudien berechnet werden können.

Schlüsselworte: Confounding, Delta-Methode, Epidemiologie, Konfidenzintervalle, Logistische Regression, Number needed to be exposed (NNE), Number needed to treat (NNT), Odds Ratio (OR).

1 Einleitung

Die Zahl "number needed to treat" (NNT) ist vor allem in randomisierten klinischen Studien ein populäres Maß zur Beschreibung des Effekts einer neuen Behandlung im Vergleich zu einer Standardbehandlung bezüglich des Risikos eines unerwünschten Ereignisses (Cook & Sackett, 1995). Die Zahl NNT gibt die durchschnittliche Anzahl von Patienten an, die über einen bestimmten Zeitraum die neue Behandlung erhalten müssen, um im Vergleich mit der Standardbehandlung ein Ereignis zu verhindern. In epidemiologischen Studien ist es häufig das Ziel, die Risiken zwischen exponierten und nicht exponierten Personen zu vergleichen. Daher haben Bender & Blettner (2002) zur Verwendung von NNT als Effektmaß in der Epidemiologie die Bezeichnung "number needed to be exposed" (NNE) vorgeschlagen. Aufgrund der fehlenden Randomisierung ist in klassischen epidemiologischen Studien die Betrachtung von Confoundern eine wichtige Notwendigkeit. Bender & Blettner (2002) haben auf der Basis der multiplen logistischen Regression eine einfache Methode hergeleitet, um adjustierte NNEs mit approximativen Konfidenzintervallen zu berechnen. Diese Methode ist geeignet, um bei Confoundern mit nicht zu breiter Verteilung den Bias wirksam zu reduzieren und Konfidenzintervalle mit hinreichend guten Überdeckungseigenschaften zu liefern (Bender & Kuss, 2003). Bei sehr breiten oder sehr schiefen Verteilungen liefert der einfache Ansatz nach Bender & Blettner (2002) jedoch immer noch verzerrte Schätzer, so dass für allgemeine Anwendungen eine verbesserte Methode benötigt wird. Eine solche verbesserte Methode kann ebenso im Rahmen der multiplen logistischen Regression hergeleitet werden. In dieser Arbeit wird ein SAS/IML[®]-Programm vorgestellt, mit dessen Hilfe adjustierte NNEs mit Konfidenzintervallen nach der einfachen und nach der verbesserten Methode berechnet werden können.

2 Das Maß "number needed to be exposed" (NNE)

Das NNT-Konzept kann prinzipiell immer dann angewendet werden, wenn zwei Gruppen bezüglich einer binären Zielvariablen verglichen werden. Wird bei diesem Vergleich nicht der Effekt einer neuen Behandlung untersucht, so ergibt der Name "number needed to treat" jedoch keinen Sinn. Zur Anwendung des NNT-Konzepts in epidemiologischen Studien, in denen die Risiken von exponierten und nicht exponierten Personen verglichen werden, haben Bender & Blettner (2002) den Begriff "number needed to be exposed" (NNE) vorgeschlagen. Dies ist die zu NNT analoge Bezeichnung, wenn anstelle einer Behandlung der Effekt einer Exposition untersucht wird. Eine Übersicht über weitere Anwendungen und Bezeichnungen gibt Bender (2005).

Die Bezeichnung NNE ermöglicht auch die Verwendung der erweiterten Terminologie von Altman (1998), um zwischen schädlichen und nützlichen Einflussgrößen zu unterscheiden sowie Konfidenzintervalle sinnvoll darzustellen. Wir verwenden die Begriffe "number needed to be exposed for one additional person to be harmed" (NNEH) bzw. "number needed to be exposed for one additional person to benefit" (NNEB), wenn es notwendig ist, die Richtung des Effekts anzugeben.

3 Berechnung adjustierter NNEs

3.1 Einfache Methode über das adjustierte Odds Ratio

Falls das Design der Studie eine unverzerrte Schätzung der Risiken durch die beobachteten rohen Ereignisraten erlaubt, so kann das NNT-Konzept in der Epidemiologie einfach durch eine entsprechende Namensänderung verwendet werden. Wie in randomisierten klinischen Studien können dann die Häufigkeiten einer einfachen Vierfeldertafel zur Berechnung von NNEs verwendet werden. Falls die Ereignisrate unter den Exponierten (exposed event rate, EER) größer ist als die Ereignisrate unter den nicht Exponierten (unexposed event rate, UER), ist die absolute Risikoerhöhung (absolute risk increase, ARI) gegeben durch $ARI = EER - UER$. Im Fall einer protektiven Exposition berechnet sich die absolute Risikoreduktion (ARR) durch $ARR = UER - EER$. Die Maße NNEH und NNEB sind dann gerade die Kehrwerte der Risikodifferenzen ARI bzw. ARR. Im Fall einer schädlichen Exposition gibt NNEH die durchschnittliche Anzahl von Personen an, die der Exposition ausgesetzt werden müssen, um im Vergleich mit nicht exponierten Personen einen Fall mehr zu beobachten.

In Beobachtungsstudien ist jedoch die Berechnung adäquater Ereignisraten schwieriger als in experimentellen randomisierten Studien, da in der Regel der Effekt von Confoundern berücksichtigt werden muss. Dies geschieht häufig durch Anwendung der multiplen logistischen Regression (Hosmer & Lemeshow, 2000), die als Output adjustierte Odds Ratios (ORs) liefert. Bender & Blettner (2002) haben gezeigt, dass mit Hilfe der Beziehung

$$NNEH = \frac{1}{(OR - 1) \times UER} + \frac{OR}{(OR - 1) \times (1 - UER)} \quad (1)$$

adjustierte NNEs berechnet werden können. Hierbei ist $OR > 1$ das adjustierte Odds Ratio für den Vergleich zwischen exponierten und nicht exponierten Personen. Für den Fall protektiver Expositionen ($OR < 1$) ergibt sich $NNEB = -NNEH$.

Wenn die Verteilung der betrachteten Confounder nicht zu breit ist, genügt es, in Formel (1) für UER die durchschnittliche Ereignisrate der nicht exponierten Personen einzusetzen. Ist die Verteilung der Confounder jedoch sehr breit oder schief, so werden Methoden benötigt, in denen die Verteilung der Confounder besser berücksichtigt wird.

3.2 Einfache Methode für bestimmte Confounder-Profile

Eine Möglichkeit, bei der Berechnung von adjustierten NNEs die Verteilung der Confounder besser zu berücksichtigen, besteht darin, in Formel (1) für UER das Risiko bei Vorliegen von bestimmten interessierenden Werten der Confounder zu verwenden. Auf diese Weise erhält man für jedes Confounder-Profil ein zugehöriges NNE. Falls z.B. Alter ein Confounder ist, so kann man mit Hilfe von Formel (1) NNEs für verschiedene Altersgruppen berechnen (siehe Beispiel in Kap. 5).

3.3 Erweiterte Methode über Mittelung

Die zweite Methode berücksichtigt zwar die Verteilung der betrachteten Confounder, sie kann aber sehr umständlich sein, da es – vor allem bei mehreren Confoundern – sehr viele verschiedene Confounder-Profile gibt. Eine weitere Möglichkeit zur Berechnung adjustierter NNEs stellt die Mittelung aller Risikoreduktionen über die Verteilung der Confounder der nicht exponierten Personen dar. Dies kann im Rahmen eines multiplen logistischen Regressionsmodells durchgeführt werden. Sei p_i das betrachtete Risiko, z_i der Wert der binären Variable $Z=$ Exposition ja/nein und x_{ij} der Wert des j -ten Confounders X_j ($j=1, \dots, k$) der i -ten Person ($i=1, \dots, n$), so lässt sich im Rahmen eines multiplen logistischen Regressionsmodells das Risiko jeder Person darstellen durch

$$p_i = \frac{\exp(\alpha + \gamma z_i + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\alpha + \gamma z_i + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}, \quad i=1, \dots, n \quad (2)$$

Für alle nicht exponierten Personen gilt (da $Z=0$) die Formel

$$p_i = \frac{\exp(\alpha + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\alpha + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}, \quad i=1, \dots, m \quad (3)$$

wobei m die Zahl der nicht exponierten Personen in der Stichprobe ist. Für jede nicht exponierte Person lässt sich nun das hypothetische Risiko berechnen, welches diese Person hätte, wenn sie exponiert wäre, durch

$$p_i^* = \frac{\exp(\alpha + \gamma + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\alpha + \gamma + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}, \quad i=1, \dots, m \quad (4)$$

Für jede nicht exponierte Person erhält man dann durch die Differenz zwischen p_i und p_i^* die individuelle Risikoerhöhung. Die mittlere Risikoerhöhung über die Verteilung der Confounder der nicht exponierten Personen ist dann gegeben durch

$$\overline{\text{ARI}} = \frac{1}{m} \sum_{i=1}^m (p_i^* - p_i) \quad (5)$$

Das adjustierte NNEH erhält man dann durch den Kehrwert des mittleren ARI aus (5).

4 Konfidenzintervalle für adjustierte NNEs

Die Verteilung geschätzter NNTs und NNEs kann häufig nicht durch eine Normalverteilung approximiert werden (Lesaffre & Pledger, 1999). Daher werden Konfidenzintervalle für diese Maße indirekt über Konfidenzintervalle der entsprechenden Risikodifferenzen berechnet (Bender, 2001). Im Rahmen der logistischen Regression können diese Risikodifferenzen als Funktionen der Regressionskoeffizienten dargestellt werden. Daher ist es möglich, approximative Konfidenzintervalle für ARI bzw. ARR mit Hilfe der multivariaten Delta-Methode (Bishop, Fienberg & Holland, 1975) zu berechnen. Die Konfidenzgrenzen für NNEs erhält man dann durch Invertieren und Vertauschen der Konfidenzgrenzen für ARI bzw. ARR (Bender & Blettner, 2002).

5 Beispiele

5.1 Simulierte Daten: Hypothetische Kohortenstudie

Zunächst erläutern wir die vorgestellten Methoden zur Berechnung adjustierter NNEs anhand von simulierten Daten. Angenommen in einer Kohortenstudie werden $n=4000$ Personen über einen festen Zeitraum (z.B. 5 Jahre) beobachtet. Sei Y eine binäre Zielvariable, Z eine binäre Exposition, und X ein stetiger Confounder (z.B. Alter). Werte für diese Variablen wurden wie folgt erzeugt. Die Exposition Z erhält die Werte 1 und 0 jeweils mit Wahrscheinlichkeit 0.5; X ist normalverteilt mit $\sigma=7$ und $\mu=40$ für die exponierten und $\mu=45$ für die nicht exponierten Personen, d.h. die nicht exponierten Personen sind im Mittel 5 Jahre älter als die Exponierten. Dadurch sind die rohen Ergebnisse stark verzerrt. Das Risiko für das Zielereignis $p=P(Y=1)$ wird in Abhängigkeit von Z und X über das logistische Regressionsmodell

$$p = \frac{\exp(\alpha + \gamma z + \beta x)}{1 + \exp(\alpha + \gamma z + \beta x)} \quad (6)$$

mit $\alpha=-10$, $\gamma=1.0986$ und $\beta=0.1823$ erzeugt, d.h. das wahre Odds Ratio für den Expositionseffekt ist $OR = \exp(1.0986) = 3$ und für das Alter gilt $OR = \exp(0.1823) = 1.2$. Die Zielvariable Y wird erzeugt durch Zuweisung des Werts 1 (Ereignis) mit Wahrscheinlichkeit p und des Werts 0 (kein Ereignis) mit Wahrscheinlichkeit $1-p$. Die Programmcodes zur Erzeugung der Daten dieser hypothetischen Kohortenstudie befinden sich am Beginn des SAS/IML[®]-Programms im Anhang.

Im Folgenden wird ein arbiträr ausgewählter simulierter Datensatz ausgewertet. Mit den erzeugten Werten für die Exposition und das Alter erhält man in diesem konkreten Datenbeispiel ein wahres NNE von $NNEH=5.81$.

Betrachtet man nur die einfache Vierfeldertafel zwischen Exposition und Zielvariable (Tab. 1), so erhält man einen nicht signifikanten Effekt der Exposition ($OR = 1.09$, 95% Konfidenzintervall: 0.94 bis 1.27, $p=0.259$). Das geschätzte rohe NNE ist $NNEH=68$ (95% KI: $NNEH$ 25 bis $\pm\infty$ bis $NNEB$ 95). Dieses Ergebnis ist jedoch stark verzerrt, da der Confounder Alter nicht berücksichtigt wurde.

Tabelle 1: Vierfeldertafel Exposition \times Zielvariable von simulierten Daten

		Zielereignis		Gesamt
		Ja	Nein	
Exposition	Ja	434 (21.6%)	1575 (78.4%)	2009
	Nein	401 (20.1%)	1590 (79.9%)	1991
Gesamt		835	3165	4000

Tabelle 2: Ergebnisse der logistischen Regression für die simulierten Daten

	Koeffizient	SE	<i>p</i>	OR	95% KI
Intercept	-10.1743	0.3884	< 0.0001		
Z (Exposition)	1.0801	0.0969	< 0.0001	2.945	2.44 – 3.56
X (Alter)	0.1861	0.0079	< 0.0001	1.205	1.19 – 1.22
Hosmer-Lemeshow Goodness-of-Fit-Test: <i>p</i> =0.4181					

Nach Berücksichtigung des Confounders mit Hilfe des multiplen logistischen Regressionsmodells (2) erhält man dagegen einen signifikanten schädlichen Expositionseffekt (Tab. 2). Das geschätzte adjustierte Odds Ratio von OR=2.945 liegt sehr viel näher am vorgegebenen wahren Odds Ratio von OR=3. Bei Verwendung der durchschnittlichen Ereignisrate der nicht exponierten Personen (UER=0.2014) erhält man mit Hilfe von (1) das adjustierte NNE von NNEH=4.45 (95% KI: 3 bis 6). Das wahre NNEH wird mit dieser einfachen Methode unterschätzt, der Effekt der Exposition damit überschätzt. Die zweite Möglichkeit ist die Berechnung von NNEs für feste Alterswerte. Tabelle 3 enthält neben den Ereignisraten für exponierte (EER) und nicht exponierte Personen (UER) auch das relative Risiko (RR), das Odds Ratio (OR), die absolute Risikoerhöhung (ARI) und den NNEH-Wert in Abhängigkeit vom Alter.

Tabelle 3: Ereignisraten, RR, OR, ARI, und NNEH für verschiedene Altersstufen

Alter (Jahre)	EER	UER	RR	OR	ARI	NNEH
20	0.0046	0.0016	2.936	2.945	0.0030	328.1
30	0.0290	0.0100	2.889	2.945	0.0190	52.8
35	0.0704	0.0251	2.808	2.945	0.0453	22.1
40	0.1611	0.0612	2.632	2.945	0.0999	10.0
45	0.3275	0.1419	2.308	2.945	0.1856	5.4
50	0.5525	0.2954	1.870	2.945	0.2571	3.9
55	0.7579	0.5153	1.471	2.945	0.2426	4.1
60	0.8881	0.7294	1.218	2.945	0.1587	6.3
70	0.9808	0.9455	1.037	2.945	0.0353	28.3

NNEH variiert damit zwischen NNEH=329 für 20-jährige über NNEH=4 für 50-jährige bis zu NNEH=29 für 70-jährige. Diese Ergebnisdarstellung vermittelt einen Eindruck, wie sich der absolute Effekt der Exposition in verschiedenen Altersgruppen auswirkt. Um den mittleren Effekt der Exposition in der betrachteten Population unter Berücksichtigung der Altersverteilung darzustellen, wird die dritte Methode zur Berechnung adjustierter NNEs benötigt. Hiermit erhält man in diesem Beispiel ein adjustiertes NNE von NNEH=5.96 (95% KI: 5 bis 7). Dieser Wert liegt dichter am wahren NNEH=5.81 als das mit der einfachen Methode geschätzte NNEH=4.45. Beide Werte sind aber deutlich besser als das rohe nicht signifikante NNEH=68.

5.2 Reale Daten: Rauchen und diabetische Retinopathie

In einer Studie von Mühlhauser et al. (1996) wurden 613 Typ 1 Diabetiker 6 Jahre lang beobachtet, um den nach weiteren Risikofaktoren adjustierten Effekt des Rauchens (Z) auf die Entwicklung einer diabetischen Retinopathie (Y) zu untersuchen. Als Risikofaktoren wurden Diabetesdauer (X_1), quadratischer Term der Diabetesdauer (X_2), glykiertes Hämoglobin (X_3) und diastolischer Blutdruck (X_4) betrachtet. Zur Datenanalyse wurde die multiple logistische Regression angewendet.

Tabelle 4: Ergebnisse der logistischen Regression für die Diabetes-Daten

	Koeffizient	SE	p	OR	95% KI
Intercept	-13.3203	1.5111	< 0.0001		
Rauchen	0.4154	0.2097	0.0476	1.515	1.004 – 2.285
X_1	0.4083	0.0655	< 0.0001		
X_2	-0.0075	0.0016	< 0.0001		
X_3	0.4292	0.0835	< 0.0001		
X_4	0.0614	0.0152	< 0.0001		
Hosmer-Lemeshow Goodness-of-Fit-Test: $p=0.581$					

Bei Berücksichtigung der betrachteten Risikofaktoren hat Rauchen einen signifikanten Effekt auf die Entwicklung einer diabetischen Retinopathie nach 6 Jahren (Tab. 4). Mit Hilfe von Formel (1) erhält man das adjustierte NNE von $NNEH=10.2$ (95% KI: 5 bis 4169). Mit Hilfe der dritten Methode, die durch Mittelung über alle nicht rauchenden Personen die Verteilung der Confounder besser berücksichtigt, erhält man das adjustierte NNE von $NNEH=14.9$ (95% KI: 7 bis 696). Das bedeutet, dass im Durchschnitt bei 15 rauchenden Diabetikern 1 Fall von Retinopathie mehr auftritt als bei Nichtrauchern. Dieses Ergebnis gilt für Typ 1 Diabetiker mit einer Verteilung der Risikofaktoren wie bei den Nichtrauchern der beobachteten Kohorte.

6 Diskussion

Die Verwendung von NNT als absolutes Effektmaß hat in den letzten Jahren in medizinischen Journals deutlich zugenommen; auf der anderen Seite wird dieses Maß allerdings auch heftig kritisiert (Hutton, 2000). Wie von Walter (2000) beschrieben, sollte eine klare Trennung zwischen der Datenanalyse und der nachfolgenden Risikokommunikation gemacht werden. Im Rahmen der logistischen Regression ist das Odds Ratio das natürliche Effektmaß für den Zweck der Datenauswertung. Die Zahl NNT ist dagegen häufig ein sinnvolles Maß zur Ergebnisdarstellung und spielt eine wichtige Rolle bei der Kommunikation der Ergebnisse mit Klinikern, den Medien und der Öffentlichkeit (Bender, 2005).

In der häufigen Situation des Vergleichs exponierter und nicht exponierter Personen in der Epidemiologie kann das NNT-Konzept als "number needed to be exposed" (NNE) zur Ergebnisdarstellung verwendet werden. Die Angabe, wie viele Personen exponiert sein müssen, um einen zusätzlichen Fall zu erzeugen, ist häufig anschaulicher als die Angabe eines Odds Ratios. Insbesondere ist NNE ein sinnvolles Maß bei modifizier-

baren Expositionen, da NNE die Frage beantwortet, wie viele Personen vor der Exposition geschützt werden müssen, um einen zusätzlichen Fall zu verhindern.

Zur Berechnung von NNEs in Kohortenstudien werden Methoden zur Berücksichtigung von Confoundern benötigt. Diese Arbeit stellt ein SAS/IML[®]-Programm zur Berechnung adjustierter NNEs mit approximativen 95% Konfidenzintervallen zur Verfügung. Dieses Programm kann in allen Situationen verwendet werden, in denen der adjustierte Effekt einer binären Exposition mit Hilfe eines multiplen logistischen Regressionsmodells geschätzt werden kann.

Literatur

- [1] Altman, D.G. (1998): *Confidence intervals for the number needed to treat*. BMJ Vol. 317, 1309-1312.
- [2] Bender, R. (2001): *Calculating confidence intervals for the number needed to treat*. Control. Clin. Trials Vol. 22, 102-110.
- [3] Bender, R. (2005): *Number needed to treat (NNT)*. In: Armitage, P. & Colton, T. (Eds.): *Encyclopedia of Biostatistics (2nd Ed.)* Chichester: Wiley, 3752-3761.
- [4] Bender, R. & Blettner, M. (2002). *Calculating the "number needed to be exposed" with adjustment for confounding variables in epidemiological studies*. J. Clin. Epidemiol. Vol. 55, 525-530.
- [5] Bender, R. & Kuss, O. (2003): *Confidence intervals for adjusted NNEs: A simulation study (Letter)*. J. Clin. Epidemiol. Vol. 56, 205-206.
- [6] Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975): *Discrete Multivariate Analysis: Theory and Methods*. Cambridge, MA: MIT Press.
- [7] Cook, R.J. & Sackett, D.L. (1995): *The number needed to treat: A clinically useful measure of treatment effect*. BMJ Vol. 310, 452-454.
- [8] Hosmer, D.W. & Lemeshow, S (2000): *Applied Logistic Regression (2nd Ed.)*. New York: Wiley.
- [9] Hutton, J.L. (2000): *Number needed to treat: Properties and problems*. J. R. Stat. Soc. A Vol. 163, 403-415.
- [10] Lesaffre, E. & Pledger, G. (1999): *A note on the number needed to treat*. Control. Clin. Trials Vol. 20, 439-447.
- [11] Mühlhauser, I., Bender, R., Bott, U., Jörgens, V., Grüsser, M., Wagener, W., Overmann, H. & Berger, M. (1996): *Cigarette smoking and progression of retinopathy and nephropathy in type 1 diabetes*. Diabet. Med. Vol. 13, 536-543.
- [12] Walter, S.D. (2000): *Choice of effect measure for epidemiological data*. J. Clin. Epidemiol. Vol. 53, 931-939.

Anhang: SAS/IML-Macro zur Berechnung adjustierter NNEs

Im Folgenden ist der SAS Code des SAS/IML-Macros "nne_v8b.sas" zur Berechnung adjustierter NNEs mit Konfidenzintervallen aufgeführt. Das Programm ist zur Verwendung unter SAS 8.02 oder höher gedacht. Falls die Version SAS 6.12 oder darunter benutzt wird, muss "Intercept" ersetzt werden durch "INTERCEP". Das Programm ist auch im Internet erhältlich unter "<http://www.rbsd.de/softw.html>".

```

1 *****
2 *           CALCULATION OF THE NUMBER NEEDED TO BE EXPOSED (NNE)           *
3 *           AND 95% CONFIDENCE INTERVALS                                   *
4 *           WITH ADJUSTMENT FOR CONFOUNDING VARIABLES                       *
5 * REF.: BENDER & BLETTNER (2002), J. CLIN. EPIDEMIOLOG. 55: 525-530 *
6 *****;
7
8 *-----*
9 |                               QUESTIONS ?                               |
10 | |                               |                                       |
11 | ---> Dr. Ralf BENDER  Phone:  +49 221 35685-451  |
12 | |                               Fax:    +49 221 35685-891  |
13 | |                               E-Mail:  Ralf@rbsd.de  |
14 *-----*;
15
16 *-----*
17 !!                               NOTE:                               !!
18 !!                               !!                               !!
19 !! This code is appropriate for the SAS Version 8 or higher.        !!
20 !! You have to replace the simulated data set below with your      !!
21 !! own data. The SAS name of your data set should be "cohort".    !!
22 !!                               !!                               !!
23 !! This data set must contain the following variables:            !!
24 !! 1) RESPONSE = adverse event of interest (1=yes, 0=no)          !!
25 !! 2) EXPOSURE = binary exposure variable (1=yes, 0=no)          !!
26 !! 3) COVAR_1 = the first confounding variable                    !!
27 !! 4) If there are more than 1 confounding variables these        !!
28 !! variables must be named COVAR_2, COVAR_3, and so on.          !!
29 !!                               !!                               !!
30 !! In line 39 you have to enter the number of confounding         !!
31 !! variables (e.g. %LET K=3) !                                     !!
32 !!                               !!                               !!
33 !! It is necessary to investigate the goodness-of-fit of the      !!
34 !! logistic regression model to get useful NNE estimates !        !!
35 *-----*;
36
37 options linesize=105;
38
39 *---->; %LET K=1 ; *** K = Number of covariables (excluding exposure);
40
41 *-----*
42 |                               Simulation of Cohort Data                               |
43 *-----*;
44 title1 'Calculation of Adjusted NNEs with 95% Confidence Intervals';
45 title4 'Simulated Cohort Data (n=4000, OR=3, 1 Covariate)';
46 title5 ' ';

```

```
47 data sim;
48   do i=1 to 4000;
49     No=i;
50     alpha=-10;
51     gamma=1.0986123;
52     beta1=0.1823216;
53     Z=RANBIN(0,1,0.5);
54     if Z=0 then X1=7*RANNOR(0)+45;
55     if Z=1 then X1=7*RANNOR(0)+40;
56     prob=exp(alpha+gamma*Z+beta1*X1)/(1+exp(alpha+gamma*Z+beta1*X1));
57     Y=RANBIN(0,1,prob);
58     output;
59   end;
60   drop i alpha gamma beta1;
61 run;
62
63 *-----*
64 |           True Parameters of Simulated Cohort Data           |
65 *-----*;
66 title4 'True Mean UER, Hypothetical EER, True OR, True ARI, and True NNEH';
67 title5 ' ';
68 data UE;  set sim;  if Z=0;
69   alpha=-10;
70   gamma=1.0986123;
71   beta1=0.1823216;
72   UER=exp(alpha+beta1*X1)/(1+exp(alpha+beta1*X1));
73   EER=exp(alpha+beta1*X1+gamma)/(1+exp(alpha+beta1*X1+gamma));
74   ARI=EER-UER;
75   keep UER EER ARI;
76 run;
77
78 proc means data=UE noprint;
79   var UER EER ARI;
80   output out=UEmean mean=UER_True EER_Hyp ARI_True;
81 run;
82
83 data TRUE;  set UEmean;
84   OR_True = exp(1.098612);
85   NNEH_True = 1/ARI_True;
86   format NNEH_True 16.8;
87 run;
88
89 proc print data=TRUE;
90 var UER_True EER_Hyp OR_True ARI_True NNEH_True;
91 run;
92
93 *-----*
94 |           Descriptive Anlysis of Generated Cohort Data           |
95 *-----*;
96 title4 'Descriptive Anlysis of Generated Cohort Data';
97 title5 ' ';
98 data cohort;  set sim;
99   RESPONSE = Y;
100  EXPOSURE = Z;
101  COVAR_1 = X1;
102  drop Y Z X1 PROB;
103 run;
```

```

104 proc means data=cohort n min max mean std sum maxdec=3 fw=8;
105 proc means data=cohort n min max mean std sum maxdec=3 fw=8;
106 class EXPOSURE;
107 run;
108
109 proc corr data=cohort spearman;
110     var EXPOSURE COVAR_1;
111 run;
112
113 *-----*
114 |                               Simple 2x2 Table                               |
115 *-----*
116 title4 'Simple 2x2 Table with Crude Estimation of OR';
117 title5 ' ';
118 proc freq data=cohort;
119     tables EXPOSURE*RESPONSE / nopercnt nocol cmh exact;
120 run;
121
122 *-----*
123 |                               Multiple Logistic Regression Model                               |
124 *-----*
125 title4 'Estimation of Adjusted Odds Ratio (Logistic Regression)';
126 title5 ' ';
127 proc logistic data=cohort descending outest=est covout;
128     model RESPONSE = EXPOSURE COVAR_1-COVAR_&K / rl lackfit rsq;
129     output out=risks p=risk l=low u=up xbeta=linpred;
130 run;
131
132 *-----*
133 |                               Data Sets Used for SAS/IML                               |
134 *-----*
135 data COVARIAT;    set cohort;
136     keep COVAR_1-COVAR_&K;
137     if EXPOSURE=0;
138 run;
139
140 data beta;    set est;    if _type_='PARMS';
141     drop _LINK_ _TYPE_ _NAME_ _LNLIKE_;
142 run;
143
144 data beta2_K;    set beta;
145     keep COVAR_1-COVAR_&K;
146 run;
147
148 data COVB;    set est;    if _type_='COV';
149     drop _LINK_ _TYPE_ _NAME_ _LNLIKE_;
150 run;
151
152 data covb2_k;    set covb;
153     keep COVAR_1-COVAR_&K;
154 run;

```

```

155 *-----*
156 |                PROC IML                |
157 *-----*
158 title4 'Crude and Adjusted Results Calculated with SAS/IML';
159 title5 ' ';
160 proc IML;
161 use cohort;    read all var{RESPONSE}                into RESPONSE;
162 use cohort;    read all var{RESPONSE} where (EXPOSURE=0) into RESP_UNE;
163 use cohort;    read all var{RESPONSE} where (EXPOSURE=1) into RESP_EXP;
164 use cohort;    read all var{EXPOSURE}                into EXPOSURE;
165 use covariat;  read all var _all_                    into COVARIAT;
166 use beta;      read all var{Intercept}              into BETA0;
167 use beta;      read all var{EXPOSURE}              into BETA1;
168 use beta2_k;   read all var _all_                    into BETA2_k;
169 use covb;      read all var{Intercept EXPOSURE}     into COV1;
170 use covb2_k;   read all var _all_                    into COV2;
171
172 COVB = (COV1 || COV2);
173 N     = NROW(RESPONSE);
174 EVENT = SUM(RESPONSE);
175 Ne    = SUM(EXPOSURE);
176 Nu    = N-NE;
177 Ee    = SUM(RESP_EXP);
178 Eu    = SUM(RESP_UNE);
179 K     = NCOL(BETA2_k);
180 UER   = SUM(RESP_UNE)/NU;
181 EER   = SUM(RESP_EXP)/NE;
182 ARI   = EER-UER;
183 NNEH  = 1/ARI;
184
185 print "GENERAL INFORMATION";
186 print "Sample size and number of events: "
187       N " " EVENT;
188 print "Number of unexposed and exposed subjects with events: "
189       NU EU " " NE EE , , , , ;
190
191 print "CRUDE (UNADJUSTED) RESULTS";
192 if ARI>0 then do;
193 print "Crude event rates, ARI and NNEH: "
194       UER[format=6.4] " " EER[format=6.4] " "
195       ARI[format=6.4] " " NNEH[format=8.2] , , , , ;
196 end;
197
198 if ARI<0 then do;
199 ARR = -ARI;
200 NNEB = 1/ARR;
201 print "Crude event rates, ARR and NNEB: "
202       UER[format=6.4] " " EER[format=6.4] " "
203       ARR[format=6.4] " " NNEB[format=8.2] , , , , ;
204 end;
205
206 if ARI=0 then do;
207 NNEH="infinity";
208 print "Crude event rates, ARI and NNEH: "
209       UER[format=6.4] " " EER[format=6.4] " "
210       ARI[format=6.4] " " NNEH[format=8.2] , , , , ;
211 end;

```

```

212 D10 = 0;
213 D11 = exp(BETA1);
214 D12_k = 0#BETA2_k;
215 D1 = (D10 || D11 || D12_k);
216
217 LINPRED = BETA0 + COVARIAT*BETA2_k`;
218 LOGIST2 = EXP(LINPRED) / ((1+EXP(LINPRED))##2);
219 DO J=1 to K;
220 LOGISTJ = COVARIAT[,J] # EXP(LINPRED) / ((1+EXP(LINPRED))##2);
221 LOGIST = (LOGIST || LOGISTJ);
222 END;
223
224 D20 = SUM(LOGIST2)/NU;
225 D21 = 0;
226 D22_k = LOGIST[+,]/NU;
227 D2 = (D20 || D21 || D22_k);
228 D = (D1 // D2);
229
230 COVF = D*COVB*D`;
231
232 OR = exp(BETA1);
233 RISK = EXP(LINPRED) / (1+EXP(LINPRED));
234 UER = SUM(RISK)/Nu;
235
236 dv1 = UER#(1-UER) / ((1+UER#(OR-1))##2);
237 dv2 = (OR-1)#(1-2#UER-(OR-1)#UER##2) / ((1+UER#(OR-1))##2);
238 dv = (dv1 || dv2);
239
240 ARI = (OR-1)#UER#(1-UER) / (1+UER#(OR-1));
241 EER = ARI+UER;
242
243 print "ADJUSTED RESULTS (SIMPLE METHOD VIA ODDS RATIO)";
244 print "Attention: The simple method is only valid if the confounder
245 distribution is not too wide!";
246
247 if ARI=0 then do;
248 NNEH="infinity";
249 print "Absolute risk increase (ARI):" ARI[format=6.4];
250 print "Number needed to be exposed to be harmed (NNEH):"
251 NNEH[format=8.2] , , , ;
252 print "Sorry, but because ARI=0 the delta method is not applicable
253 to calculate the standard error of ARI!";
254 abort;
255 end;
256
257 else NNEH = 1/ARI;
258
259 VAR = dv*COVF*dv`;
260 SE = SQRT(VAR);
261
262 sn = 0.05;
263 Zsn = PROBIT(1-SN/2);
264 ARI_L = ARI - Zsn#SE;
265 ARI_U = ARI + Zsn#SE;
266 NNEH_L = 1/ARI_U;
267 NNEH_U = 1/ARI_L;

```

```
268 print "OR, observed unexposed (UER) and hypothetical
269   exposed event rate (EER): "
270   OR[format=7.3] " " UER[format=6.4] " " EER[format=6.4];
271
272 if ARI>0 then do;
273 print "Absolute risk increase (ARI) with SE and 95% CI: "
274     ARI[format=6.4] " " SE[format=8.6] " "
275     ARI_L[format=6.4] ARI_U[format=6.4];
276 print "Number needed to be exposed to be harmed (NNEH) with 95% CI: "
277     NNEH[format=8.2] " " NNEH_L[format=8.2] NNEH_U[format=8.2] , , , , ;
278 end;
279
280 if ARI<0 then do;
281   ARR = -ARI;
282   NNEB = 1/ARR;
283   ARR_L = ARR - Zsn#SE;
284   ARR_U = ARR + Zsn#SE;
285   NNEB_L = 1/ARR_U;
286   NNEB_U = 1/ARR_L;
287 print "Absolute risk reduction (ARR) with SE and 95% CI: "
288     ARR[format=6.4] " " SE[format=8.6] " "
289     ARR_L[format=6.4] ARR_U[format=6.4];
290 print "Number needed to be exposed to benefit (NNEB) with 95% CI: "
291     NNEB[format=8.2] " " NNEB_L[format=8.2] NNEB_U[format=8.2] , , , , ;
292 end;
293
294 print "ADJUSTED RESULTS (GENERAL METHOD VIA AVERAGING)";
295 print "Note: The general method is valid for narrow and wide
296   confounder distributions";
297
298 LINPRED1 = BETA0 + COVARIAT*BETA2_k`;
299 LINPRED2 = BETA0 + BETA1 + COVARIAT*BETA2_k`;
300 LOGIST1 = EXP(LINPRED1) / ((1+EXP(LINPRED1))##2);
301 LOGIST2 = EXP(LINPRED2) / ((1+EXP(LINPRED2))##2);
302 d1 = SUM(LOGIST2-LOGIST1)/Nu;
303 d2 = SUM(LOGIST2)/Nu;
304
305 DO J=1 to K;
306 LOGCOVJ = COVARIAT[,J] # (LOGIST2-LOGIST1);
307 LOGCOV = (LOGCOV || LOGCOVJ);
308 END;
309
310 d3 = LOGCOV[+,]/Nu;
311 d = (d1 || d2 || d3);
312
313 RISK1 = EXP(LINPRED1)/(1+EXP(LINPRED1));
314 RISK2 = EXP(LINPRED2)/(1+EXP(LINPRED2));
315 UER = SUM(RISK1)/Nu;
316 EER = SUM(RISK2)/Nu;
317 ARI = EER-UER;
318
```

```

319 if ARI=0 then do;
320   NNEH="infinity";
321   print "Absolute risk increase (ARI):" ARI[format=6.4];
322   print "Number needed to be exposed to be harmed (NNEH):"
323     NNEH[format=8.2] , , , ;
324 print "Sorry, but because ARI=0 the delta method is not applicable
325 to calculate the standard error of ARI!";
326 abort;
327 end;
328
329 else NNEH = 1/ARI;
330 VAR  = d*COVB*d`;
331 SE   = SQRT(VAR);
332
333 sn    = 0.05;
334 Zsn   = PROBIT(1-SN/2);
335 ARI_L = ARI - Zsn#SE;
336 ARI_U = ARI + Zsn#SE;
337 NNEH_L = 1/ARI_U;
338 NNEH_U = 1/ARI_L;
339
340 print "Observed unexposed (UER) and hypothetical
341 exposed event rate (EER): "
342   UER[format=6.4] " " EER[format=6.4];
343
344 if ARI>0 then do;
345 print "Absolute risk increase (ARI) with SE and 95% CI: "
346   ARI[format=6.4] " " SE[format=8.6] " "
347   ARI_L[format=6.4] ARI_U[format=6.4];
348 print "Number needed to be exposed to be harmed (NNEH) with 95% CI: "
349   NNEH[format=8.2] " " NNEH_L[format=8.2] NNEH_U[format=8.2];
350 end;
351
352 if ARI<0 then do;
353   ARR  = -ARI;
354   NNEB = 1/ARR;
355   ARR_L = ARR - Zsn#SE;
356   ARR_U = ARR + Zsn#SE;
357   NNEB_L = 1/ARR_U;
358   NNEB_U = 1/ARR_L;
359 print "Absolute risk reduction (ARR) with SE and 95% CI: "
360   ARR[format=6.4] " " SE[format=8.6] " "
361   ARR_L[format=6.4] ARR_U[format=6.4];
362 print "Number needed to be exposed to benefit (NNEB) with 95% CI: "
363   NNEB[format=8.2] " " NNEB_L[format=8.2] NNEB_U[format=8.2];
364 end;
365
366 quit;
367 run;

```