

Vergleich verschiedener Verfahren der Confounderselektion Eine Simulationsstudie mit 4 Kovariablen

N. Mühlberger⁽¹⁾, U. Siebert⁽²⁾, A. Wulff⁽¹⁾, P. Reitmeir⁽¹⁾

⁽¹⁾ GSF-Forschungszentrum für Umwelt und Gesundheit, Medis Institut, Neuherberg

⁽²⁾ Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, LMU München

Einleitung

Die Schätzung unverzerrter Expositionseffekte aus Beobachtungsdaten ist ein Hauptziel epidemiologischer Forschung. Dabei ist Kontrolle von Confounding eine wesentliche Voraussetzung für valide Ergebnisse. Wegen der Komplexität biologischer Prozesse ist das Vorwissen über confundierende Zusammenhänge in der Grundgesamtheit jedoch meist unzureichend, so daß man auf datengestützte Verfahren zur Confounderselektion zurückgreifen muß. Seit längerem befinden sich mehrere Selektionsverfahren in Anwendung, deren Eignung kontrovers diskutiert wird. Frühere Simulationen mit nur einer Kovariablen [1, 2] untersuchten die Güte verschiedener Verfahren für Mantel-Haenszel und Poisson-Regressionsanalysen. Unser Ziel war es die Verfahren in realistischeren Situationen mit mehreren Kovariablen und bei logistischer Regressionanalyse zu vergleichen.

Methoden

Prinzip der Simulation war, den wahren Effekt einer Exposition in der Grundgesamtheit (β) aus einem volladjustierten Superpopulationsmodell ($n=10^7$) abzuleiten und diesen mit der Effektschätzung aus Stichproben (b) zu vergleichen. Die Untersuchung beschränkte sich auf 12 gezielt gewählte Szenarien mit binärer Ziel- und Expositionsgröße und vier binären, untereinander unkorrelierten Kovariablen (C1-C4).

Datenerzeugung. Zur Erzeugung der Superpopulationen und Stichproben wurde zunächst eine, die rohe Beziehung von Exposition (E) und Zielgröße (D) beschreibende 2x2 Tafel, generiert. Anschließend wurde jedem nach Expositions- und Ereignisstatus klassifizierten Subjekt mit Hilfe eines Zufallsgenerators ein Kovariablenstatus zugewiesen. Vorgaben zu Expositionsprävalenz ($P[E=1]=0.3$), Ereignisprävalenz unter Exposition ($P[D|E=1]=0.1$), roher Expositions-Ereignis Assoziation ($OR_{ED}=3$) und Kovariablenprävalenz unter Exposition ($P[C=1|E=1]=0.5$) waren für alle Simulationen identisch. Variiert wurden die klassischen Confounderkriterien - Kovariablenassoziation zur Exposition (OR_{CE}) und Kovariablenassoziation zur Zielgröße bei Nichtexposition (OR_{CD}).

Tabelle 1. Parametervorgaben bei der Datenerzeugung

<u>Rohe 2x2 Tafel</u>	<u>Kovariablenstatus</u>
<ul style="list-style-type: none"> • n • $P[E=1] = 0,3^*$ • $P[D=1 E=1] = 0,1^*$ • $OR[ExD] = 3^*$ 	<ul style="list-style-type: none"> • $OR[C_i \times E]$ • $OR[C_i \times D E=0]$ • $P[C_i=1 E=1] = 0,5^*$ • Effektmodifikationsfaktor = 1^*

* Im Rahmen der Simulation nicht variiert

Simulationsszenarien. Folgende Auswertungsszenarien wurden simuliert - (S1a) vier Noise-Variablen; (S2a-S2b) vier ausschließlich expositionsassoziierte Größen; (S3a-S3b) vier ausschließlich outcomeassoziierte Größen; (S4a-S4c) zwei Confounder, zwei outcomeassoziierte

Größen; (S5a-S5c) vier Confounder; (S6a) ein Confounder, je eine outcome- und expositionsassoziierte Größe und eine Noise Variable; Tabelle 2 zeigt die Assoziationsvorgaben der 12 untersuchten Kovariablenkonstellationen mit der daraus resultierenden marginalen, prozentualen OR-Abweichung (CIE%). Jede Konstellation wurde unter vier verschiedenen Fallzahlbedingungen analysiert, die sich durch die Zahl der Ereignisse (E) pro Freiheitsgrad (df) des vollen Modells unterschieden (5, 10, 20 und 40 E/df → Stichprobenumfänge 455, 910, 1820, 3640). Bei Einbeziehung der Fallzahlvarianten ergaben sich 48 Simulationsszenarien.

Tabelle 2. Kovariablenvorgabe der Simulationsszenarien mit daraus resultierender prozentualer Odds Ratio Abweichung (CIE% beobachtete in der Superpopulation)

Szenarien	C1			C2			C3			C4		
	OR _{CE}	OR _{CD}	CIE%									
S1a	1	1	0,0	1	1	0,0	1	1	0,0	1	1	0,0
S2a	1,1	1	0,0	1,2	1	0,0	1,5	1	0,0	2	1	0,1
S2b	2	1	0,0	4	1	0,0	6	1	0,0	8	1	0,1
S3a	1	1,1	0,0	1	1,2	0,0	1	1,5	0,3	1	2	0,6
S3b	1	2	0,7	1	4	2,4	1	6	3,7	1	8	4,5
S4a	2	2	11,7	4	4	53,7	1	2	0,7	1	4	2,5
S4b	1,5	2,0	6,4	2	1,5	6,9	1	2	0,7	1	3	1,7
S4c	3	5	46,9	5	3	48,0	1	2	0,7	1	3	1,7
S5a	1,5	3	9,5	3	1,5	10,8	3	5	46,9	5	3	48,0
S5b	1,2	3	3,1	3	1,2	4,8	1,5	2	6,3	2	1,5	6,9
S5c	1,2	10	3,1	10	1,2	7,9	1,5	8	14,1	8	1,5	17,8
S6a	1	3	1,8	2	2	11,8	5	1	0,0	1	1	0,0

Methoden der Confounderselektion.

1. rohes Modell
2. volles Modell
3. Kovariablenwahl nach Wald Statistik (p-Wert)
4. Kovariablenwahl nach relativem Change-in-Estimate (CIE%) [1, 2]

$$\text{CIE\%} = \left[\exp\left(|b_{adj} - b_{crude}|\right) - 1 \right] \times 100$$

5. Kovariablenwahl nach Signifikanz des Change-in-Estimate (p-CIE) [1, 2]

$$W = \frac{b_{adj} - b_{crude}}{\sqrt{\text{Var}(b_{adj}) - \text{Var}(b_{crude})}}, \text{ wobei } W^2 \text{ Chi}^2\text{-verteilt ist. } W = 0, \text{ wenn } \text{Var}(b_{crude}) \geq \text{Var}(b_{adj})$$

Die Güte der schrittweisen Selektionsverfahren (3-5) wurde für mehrere α -Niveaus bzw. CIE Schranken beurteilt (5%, 10%, 20%), wodurch sich die Gesamtzahl der zu beurteilenden Auswertungsstrategien auf 11 erhöhte.

Auswertungsstrategie. Zu jedem der 48 Simulationsszenarien wurden 500 Datensätze erzeugt, die mit allen 11 Verfahren ausgewertet wurden. Die Güte der Verfahren wurde anhand der absoluten Abweichungen vom wahren Wert der Superpopulation ($|\Delta\beta|=|b-\beta|$) beurteilt. Durch Verwendung der Absolutwerte wurden lage- und streuungsbedingte Schätzfehler gleichermaßen berücksichtigt und eine auf Change in Estimate basierende Interpretation der Verteilungsparameter ermöglicht. So zeigt beispielsweise ein Median von 0,2, daß die geschätzte OR in 50% der Analysen um mehr als 22% vom wahren Wert abweicht ($[\exp(0,2)-1]*100=22\%$).

Der Methodenvergleich erfolgte für alle Szenarien getrennt. Zunächst wurden die für das jeweilige Szenario besten Schrankenwerte gesucht. Danach wurde unter Verwendung der insgesamt am besten abschneidenden Schrankenwerte das optimale Selektionsverfahren ermittelt. Der Nachweis von Verfahrensunterschiede erfolgte über paarweise Wilcoxon-Vorzeichen-Rang Tests mit Bonferroni-Holm adjustierten p-Werten zum multiplen Testniveau von 0,05. Zur komprimierten Ergebnisdarstellung wurde jedem Verfahren ein Rang zugeordnet, der das Abschneiden in der multiplen Testprozedur erkennen läßt. Höhere Rangzahl weist dabei auf ein signifikant besseres Abschneiden hin. Verfahrensunterschiede, die das multiple Testniveau nicht einhielten, wurden durch Unterstreichung gekennzeichnet.

Ergebnisse

Vergleich der Schrankenwerte. Insgesamt zeigte sich eine Überlegenheit liberaler Schrankenwahl ($\alpha=0,2$ bzw. CIE=5%). Bei geringer Confoundingvorgabe oder kleiner Fallzahl (<10E/df) erwiesen sich jedoch restriktivere Schranken zum Teil als besser. Besonders deutlich zeigte sich der Vorteil restriktiver Modellierung bei Szenarien mit ausschließlich expositionsassoziierten Kovariablen (S2a, S2b). Tabelle 3-5 zeigt eine detaillierte Übersicht des Schrankenwertvergleichs unter den simulierten Kovariablenkonstellationen (S1a-S6a) und Fallzahlbedingungen (40E/df-5E(df)). Wegen der deutlichen Überlegenheit liberaler Schranken wurden diese für den nachfolgenden Vergleich zwischen den Selektionsverfahren, dargestellt in Tabelle 6, beibehalten.

Rohes Modell. In den Szenarien mit Confoundervorgabe (S4a-S6a) liefert das rohe Modell zumeist die schlechteste Effektschätzung. Allerdings zeigte sich auch hier bei weniger als vier Confoundern und schwächerer Confoundingvorgabe (S4a, S4b, S6a) eine mit abnehmender Fallzahl zunehmende Verbesserung der leeren Modelle. In den Szenarien ohne Confoundervorgabe (S1a-S3b) erwies sich das rohe Modell erwartungsgemäß häufig überlegen. Dabei zeigte sich der Vorteil restriktiver Modellierung erneut am deutlichsten im Falle ausschließlich expositionsassoziierten Kovariablen (S2a, S2b).

Volles Modell. Erwartungsgemäß zeigte sich Volladjustierung in allen „vier Confounder“ Szenarien (S5a-S5c) überlegen. In den übrigen Szenarien mit geringerer Confoundervorgabe führte sie jedoch insbesondere bei geringer Fallzahl, häufig zu vergleichsweise schlechten Effektschätzern. Bei höchster Fallzahl (40E/df) lieferte sie, abgesehen von den Szenarien mit ausschließlich expositionsassoziierten Kovariablen (S2a, S2b), aber auch hier gute Ergebnisse.

P-Wert Selektion. Ein auf sämtlichen Datensätzen basierender Vergleich ergab eine Unterlegenheit des p-Wert Verfahrens gegenüber p-CIE gesteuerter Variablenselektion. Wie die nach Szenarien getrennte Auswertung zeigte, lagen die Schwächen gegenüber dem p-CIE Verfahren hauptsächlich im Bereich von Konstellation mit outcomeassoziierte Kovariablen (S3b, S4a, S4b). Umgekehrt schnitt P-Wert gesteuerte Variablenselektion lediglich in einem der „vier Confounder“ Szenarien (S5a) besser ab.

CIE Selektion. Insgesamt zeigte sich auch dieses Verfahren weniger robust als p-CIE Selektion. Bemerkenswert war das schlechte Abschneiden in Szenario S5b mit vier schwachen Confoundern und in Szenarien mit starker einseitiger Expositionsassoziation (S2b, S6a). Im ersten Fall neigte das Verfahren trotz der sensitiv gewählten Schranke zur Unteradjustierung.

Im zweiten Fall dürfte das schlechte Abschneiden auf zufällig auftretende Assoziationen mit der Zielgröße zurückzuführen sein, die dazu führen könnten, daß kollinearen Kovariablen fälschlich als Confounder ausgewählt werden. Bemerkenswert scheint weiterhin, daß sich in Szenario S5c trotz der für das Verfahren günstig gewählten Vorgaben (geringe CD-Assoziation mit starker CIE-Vorgabe, bzw. starke CD-Assoziation mit geringer CIE-Vorgabe) kein signifikanter Vorteil gegenüber dem p-Wert bzw. p-CIE Verfahren nachweisen ließ.

P-CIE Selektion. Im Vergleich über alle Datensätzen, schnitt dieses Verfahren am besten ab. Ähnlich wie das p-Wert und CIE-Verfahren zeigte sich die Methode in „vier Confounder“ Situationen (S5a-S5c) dem vollen Modell und in einigen Situationen mit ausschließlich einseitiger Assoziationsvorgabe (S2a-S3a) dem rohen Modell unterlegen. Gerade in Szenarien mit einseitig outcome- bzw. expositionsassoziierten Kovariablen (S2a-S4c, S6a) kam das Verfahren jedoch häufiger zu besseren Effektschätzern als die beiden anderen schrittweisen Selektionsverfahren. Dabei zeigte sich der Vorteil gegenüber dem CIE-Verfahren vor allem bei einseitiger Expositionsassoziation (S2b), der Vorteil gegenüber dem p-Wert Verfahren vor allem bei einseitiger Outcomeassoziation (S3b).

Tabelle 3. Schrankenwertvergleich p-Wert Verfahren

Szenario	p 0.05			p 0.1			p 0.2		
	Med. $\Delta\beta$	P 95%	Rank	Med. $\Delta\beta$	P 95%	Rank	Med. $\Delta\beta$	P 95%	Rank
S1a 40E/df	0,066	0,201	1	0,067	0,201	1	0,067	0,198	1
20E/df	0,100	0,316	1	0,100	0,307	1	0,100	0,304	1
10E/df	0,148	0,471	1	0,148	0,471	1	0,150	0,471	1
5E/df	0,218	0,757	1	0,220	0,764	1	0,215	0,769	1
S2a 40E/df§	0,069	0,219	2	0,070	0,224	<u>2</u>	0,070	0,227	<u>1</u>
20E/df	0,099	0,309	3	0,101	0,314	2	0,103	0,320	1
10E/df	0,141	0,431	1	0,142	0,431	1	0,142	0,444	1
5E/df	0,206	0,759	1	0,206	0,763	1	0,219	0,779	1
S2b 40E/df	0,080	0,243	3	0,082	0,261	2	0,092	0,281	1
20E/df	0,107	0,346	3	0,115	0,351	2	0,120	0,379	1
10E/df	0,159	0,474	3	0,171	0,503	2	0,187	0,575	1
5E/df	0,233	0,775	3	0,262	0,797	2	0,266	0,855	1
S3a 40E/df	0,071	0,211	1	0,072	0,213	1	0,074	0,213	1
20E/df	0,104	0,307	1	0,104	0,309	1	0,104	0,307	1
10E/df	0,148	0,437	1	0,150	0,442	1	0,150	0,445	1
5E/df	0,223	0,727	<u>2</u>	0,224	0,727	<u>1</u>	0,227	0,747	<u>1</u>
S3b 40E/df	0,088	0,253	1	0,088	0,253	1	0,088	0,253	1
20E/df	0,135	0,371	1	0,134	0,373	1	0,134	0,373	1
10E/df	0,169	0,563	1	0,172	0,565	1	0,174	0,565	1
5E/df	0,234	0,705	1	0,237	0,698	1	0,239	0,695	1
S4a 40E/df	0,081	0,257	1	0,081	0,257	1	0,081	0,257	1
20E/df	0,129	0,382	1	0,127	0,375	2	0,127	0,369	3
10E/df	0,200	0,593	1	0,189	0,580	2	0,184	0,580	3
5E/df	0,338	0,968	1	0,337	0,961	1	0,342	0,961	1
S4b 40E/df§	0,084	0,238	<u>1</u>	0,082	0,236	<u>2</u>	0,082	0,236	2
20E/df	0,120	0,380	1	0,118	0,376	1	0,118	0,376	1
10E/df§	0,149	0,545	<u>1</u>	0,146	0,536	<u>1</u>	0,144	0,528	<u>2</u>
5E/df§	0,214	0,842	<u>1</u>	0,213	0,842	<u>1</u>	0,212	0,826	<u>2</u>
S4c 40E/df	0,100	0,276	1	0,100	0,276	1	0,100	0,276	1
20E/df	0,131	0,396	1	0,131	0,402	1	0,131	0,402	1
10E/df	0,195	0,647	1	0,194	0,637	2	0,192	0,620	2
5E/df	0,328	0,984	1	0,308	0,940	2	0,296	0,924	3
S5a 40E/df	0,107	0,317	1	0,105	0,302	2	0,100	0,303	3
20E/df	0,149	0,432	1	0,145	0,430	1	0,145	0,432	1
10E/df	0,239	0,691	1	0,222	0,678	1	0,215	0,656	1
5E/df	0,338	1,076	1	0,314	0,992	1	0,289	0,979	1
S5b 40E/df	0,091	0,282	1	0,084	0,265	2	0,081	0,255	3
20E/df§	0,125	0,377	1	0,124	0,362	<u>1</u>	0,122	0,350	<u>2</u>
10E/df	0,164	0,543	1	0,171	0,519	1	0,169	0,498	1
5E/df§	0,215	0,878	<u>1</u>	0,220	0,878	<u>2</u>	0,220	0,833	<u>3</u>
S5c 40E/df	0,128	0,391	1	0,120	0,388	2	0,109	0,368	3
20E/df	0,206	0,590	1	0,194	0,560	1	0,174	0,540	1
10E/df	0,248	0,714	1	0,261	0,722	1	0,262	0,743	1
5E/df	0,325	0,998	1	0,336	0,998	1	0,305	1,050	1
S6a 40E/df	0,072	0,235	2	0,076	0,237	1	0,076	0,236	1
20E/df	0,103	0,343	1	0,103	0,352	1	0,103	0,352	1
10E/df	0,166	0,509	1	0,164	0,511	1	0,164	0,520	1
5E/df§	0,221	0,859	<u>2</u>	0,241	0,856	1	0,241	0,850	<u>1</u>
GESAMT	0,140	0,591	1	0,140	0,581	2	0,140	0,580	2

Rangordnung aus multipler Testprozedur. Höhere Rangzahlen verdeutlichen besseres Abschneiden.

§ Unterstrichene Rangunterschiede halten das multiple Testniveau nicht ein

Tabelle 4. Schrankenwertvergleich CIE Verfahren

Szenario	CIE 0.2			CIE 0.1			CIE 0.05		
	Med. $\Delta\beta$	P 95%	Rank	Med. $\Delta\beta$	P 95%	Rank	Med. $\Delta\beta$	P 95%	Rank
S1a 40E/df	0,068	0,198	1	0,068	0,198	1	0,068	0,198	1
20E/df	0,100	0,315	1	0,100	0,315	1	0,100	0,315	1
10E/df	0,148	0,465	1	0,148	0,465	1	0,148	0,471	1
5E/df§	0,212	0,749	<u>2</u>	0,216	0,757	<u>1</u>	0,218	0,765	<u>1</u>
S2a 40E/df	0,070	0,213	1	0,070	0,213	1	0,070	0,213	1
20E/df	0,094	0,303	3	0,097	0,303	2	0,103	0,321	1
10E/df	0,142	0,420	1	0,140	0,436	1	0,141	0,438	1
5E/df	0,205	0,760	1	0,214	0,779	1	0,212	0,800	1
S2b 40E/df	0,074	0,219	3	0,085	0,262	2	0,095	0,313	1
20E/df	0,106	0,352	3	0,122	0,386	2	0,135	0,412	1
10E/df	0,178	0,533	3	0,192	0,622	2	0,209	0,627	1
5E/df	0,272	0,887	2	0,295	0,877	1	0,293	0,895	1
S3a 40E/df	0,073	0,210	1	0,073	0,210	1	0,073	0,210	1
20E/df	0,107	0,295	1	0,107	0,295	1	0,108	0,297	1
10E/df	0,148	0,410	2	0,148	0,410	2	0,151	0,414	1
5E/df	0,214	0,679	3	0,216	0,720	2	0,224	0,734	1
S3b 40E/df	0,112	0,273	1	0,114	0,270	1	0,097	0,272	2
20E/df	0,139	0,337	1	0,136	0,340	1	0,132	0,356	1
10E/df	0,185	0,466	1	0,177	0,507	1	0,170	0,557	1
5E/df§	0,206	0,651	<u>2</u>	0,217	0,692	<u>1</u>	0,230	0,680	1
S4a 40E/df§	0,091	0,283	<u>1</u>	0,085	0,264	<u>2</u>	0,082	0,237	<u>3</u>
20E/df§	0,130	0,415	1	0,131	0,371	<u>1</u>	0,122	0,368	<u>2</u>
10E/df	0,186	0,596	1	0,189	0,561	1	0,173	0,561	1
5E/df§	0,351	0,924	2	0,343	0,894	<u>2</u>	0,355	0,961	<u>1</u>
S4b 40E/df	0,112	0,325	1	0,104	0,311	2	0,085	0,237	3
20E/df	0,139	0,445	1	0,126	0,374	2	0,115	0,367	3
10E/df	0,135	0,529	1	0,147	0,529	1	0,141	0,503	1
5E/df	0,204	0,813	1	0,210	0,821	1	0,215	0,792	1
S4c 40E/df	0,092	0,263	1	0,092	0,263	1	0,093	0,269	1
20E/df	0,131	0,377	1	0,130	0,378	1	0,131	0,392	1
10E/df§	0,189	0,603	<u>2</u>	0,189	0,610	<u>2</u>	0,188	0,633	<u>1</u>
5E/df	0,286	0,897	1	0,287	0,862	1	0,278	0,829	1
S5a 40E/df	0,190	0,432	1	0,128	0,370	2	0,098	0,308	3
20E/df	0,204	0,488	1	0,160	0,448	2	0,152	0,431	3
10E/df	0,257	0,714	1	0,216	0,683	2	0,208	0,653	3
5E/df	0,306	0,955	1	0,292	0,900	2	0,282	0,900	3
S5b 40E/df	0,218	0,414	1	0,192	0,400	2	0,095	0,288	3
20E/df	0,200	0,492	1	0,170	0,458	2	0,131	0,372	3
10E/df	0,196	0,614	1	0,170	0,592	2	0,170	0,518	2
5E/df	0,246	0,943	1	0,229	0,863	2	0,233	0,834	3
S5c 40E/df	0,300	0,551	1	0,121	0,410	2	0,108	0,347	3
20E/df	0,269	0,673	1	0,189	0,519	2	0,169	0,489	3
10E/df§	0,267	0,747	<u>1</u>	0,249	0,700	<u>2</u>	0,247	0,714	<u>2</u>
5E/df	0,335	0,975	1	0,317	1,005	1	0,315	1,005	1
S6a 40E/df	0,099	0,298	1	0,086	0,256	2	0,082	0,242	2
20E/df	0,121	0,388	1	0,118	0,361	1	0,117	0,354	1
10E/df	0,164	0,540	1	0,156	0,529	1	0,165	0,538	1
5E/df	0,232	0,821	2	0,243	0,870	1	0,243	0,859	1
GESAMT	0,155	0,588	1	0,147	0,576	2	0,141	0,579	3

Rangordnung aus multipler Testprozedur. Höhere Rangzahlen verdeutlichen besseres Abschneiden.

§ Unterstrichene Rangunterschiede halten das multiple Testniveau nicht ein

Tabelle 5. Schrankenwertvergleich p-CIE Verfahren

Szenario	p-CIE 0.05			p-CIE 0.1			p-CIE 0.2		
	Med. $\Delta\beta$	P 95%	Rank	Med. $\Delta\beta$	P 95%	Rank	Med. $\Delta\beta$	P 95%	Rank
S1a 40E/df	0,068	0,198	1	0,068	0,198	1	0,068	0,198	1
20E/df	0,100	0,315	1	0,100	0,315	1	0,099	0,315	1
10E/df	0,148	0,465	1	0,148	0,471	1	0,148	0,471	1
5E/df	0,212	0,749	1	0,212	0,749	1	0,218	0,757	1
S2a 40E/df	0,070	0,213	1	0,070	0,224	1	0,069	0,227	1
20E/df	0,100	0,305	3	0,102	0,320	2	0,103	0,320	1
10E/df	0,140	0,425	1	0,140	0,429	1	0,142	0,431	1
5E/df	0,205	0,757	1	0,200	0,757	1	0,213	0,757	1
S2b 40E/df	0,080	0,243	3	0,083	0,255	2	0,091	0,290	1
20E/df	0,107	0,348	3	0,112	0,358	2	0,124	0,376	1
10E/df	0,160	0,490	3	0,171	0,503	2	0,186	0,571	1
5E/df	0,236	0,845	2	0,260	0,861	1	0,261	0,862	1
S3a 40E/df	0,072	0,211	1	0,071	0,211	1	0,071	0,210	1
20E/df	0,107	0,295	1	0,107	0,301	1	0,108	0,304	1
10E/df§	0,148	0,411	<u>2</u>	0,148	0,411	2	0,15	0,411	<u>1</u>
5E/df	0,212	0,679	2	0,211	0,720	2	0,219	0,734	1
S3b 40E/df§	0,093	0,255	<u>1</u>	0,095	0,254	1	0,090	0,256	<u>2</u>
20E/df	0,128	0,423	1	0,127	0,447	1	0,125	0,489	1
10E/df	0,181	0,485	1	0,178	0,485	1	0,173	0,502	1
5E/df	0,212	0,648	1	0,212	0,650	1	0,219	0,672	1
S4a 40E/df	0,082	0,240	1	0,084	0,244	1	0,083	0,246	1
20E/df	0,122	0,371	1	0,123	0,371	1	0,123	0,368	1
10E/df§	0,198	0,575	<u>1</u>	0,190	0,561	<u>2</u>	0,177	0,555	2
5E/df	0,344	0,928	1	0,331	0,903	1	0,32	0,895	1
S4b 40E/df	0,079	0,234	1	0,081	0,233	1	0,08	0,234	1
20E/df	0,118	0,357	1	0,112	0,354	1	0,112	0,359	1
10E/df	0,142	0,512	1	0,141	0,527	1	0,141	0,513	1
5E/df	0,206	0,806	1	0,207	0,806	1	0,212	0,806	1
S4c 40E/df	0,093	0,274	1	0,098	0,271	1	0,098	0,274	1
20E/df	0,128	0,387	1	0,129	0,392	1	0,131	0,395	1
10E/df	0,196	0,626	1	0,191	0,632	2	0,189	0,627	2
5E/df	0,338	0,970	1	0,299	0,931	2	0,288	0,931	3
S5a 40E/df	0,108	0,318	1	0,105	0,303	2	0,099	0,303	3
20E/df	0,156	0,438	1	0,154	0,430	2	0,144	0,431	3
10E/df	0,249	0,745	1	0,236	0,695	2	0,221	0,674	3
5E/df	0,353	1,134	1	0,301	1,009	2	0,289	0,955	3
S5b 40E/df	0,093	0,284	1	0,087	0,268	2	0,084	0,263	3
20E/df	0,135	0,408	1	0,126	0,371	2	0,128	0,350	3
10E/df	0,183	0,587	1	0,178	0,560	2	0,167	0,505	3
5E/df	0,245	0,938	1	0,236	0,887	2	0,215	0,872	3
S5c 40E/df	0,143	0,408	1	0,130	0,394	2	0,111	0,382	3
20E/df	0,218	0,587	1	0,199	0,571	2	0,179	0,533	3
10E/df	0,274	0,755	1	0,259	0,730	2	0,254	0,704	2
5E/df	0,335	0,959	1	0,313	0,967	2	0,306	0,969	2
S5a 40E/df	0,075	0,234	1	0,079	0,240	1	0,078	0,242	1
20E/df	0,105	0,331	<u>2</u>	0,107	0,345	<u>2</u>	0,109	0,349	<u>1</u>
10E/df	0,160	0,506	1	0,160	0,497	1	0,164	0,520	1
5E/df	0,219	0,807	2	0,231	0,807	1	0,235	0,825	1
GESAMT	0,141	0,591	1	0,140	0,576	2	0,139	0,568	3

Rangordnung aus multipler Testprozedur. Höhere Rangzahlen verdeutlichen besseres Abschneiden.

§ Unterstrichene Rangunterschiede halten das multiple Testniveau nicht ein

Tabelle 6. Intermethodenvergleich

Szenario	rohes Modell		p 0.2		p-CIE 0.2		CIE 0.05		volles Modell	
	Med. $\Delta\beta$	Rank								
S1a 40E/df	0.068	1	0.067	1	0.068	1	0.068	1	0.067	1
20E/df	0.100	1	0.100	1	0.099	1	0.100	1	0.101	1
10E/df§	0.148	<u>2</u>	0.150	<u>2</u>	0.148	<u>2</u>	0.148	<u>2</u>	0.152	<u>1</u>
5E/df§*	0.212	<u>2</u>	0.215	<u>1</u>	0.218	<u>2</u>	0.218	<u>1</u>	0.216	<u>1</u>
S2a 40E/df§*	0.070	<u>2</u>	0.070	<u>1</u>	0.069	<u>1</u>	0.070	<u>2</u>	0.070	<u>1</u>
20E/df*	0.094	2	0.103	1	0.103	1	0.103	1	0.101	1
10E/df	0.142	1	0.142	1	0.142	1	0.141	1	0.143	1
5E/df	0.206	1	0.219	1	0.213	1	0.212	1	0.219	1
S2b 40E/df*	0.075	3	0.092	2	0.091	2	0.095	2	0.102	1
20E/df*	0.099	3	0.120	2	0.124	2	0.135	1	0.134	1
10E/df*	0.160	3	0.187	2	0.186	2	0.209	1	0.359	1
5E/df*	0.208	3	0.266	2	0.261	2	0.293	1	0.299	1
S3a 40E/df	0,073	1	0,074	1	0,071	1	0,073	1	0,073	1
20E/df	0,107	1	0,104	1	0,108	1	0,108	1	0,104	1
10E/df*	0,148	3	0,15	2	0,15	2	0,151	2	0,153	1
5E/df*	0,211	2	0,227	1	0,219	1	0,224	1	0,225	1
S3b 40E/df	0.112	1	0.088	3	0.090	3	0.097	2	0.088	3
20E/df§	0.140	<u>1</u>	0.134	1	0.125	<u>2</u>	0.132	<u>1</u>	0.134	1
10E/df§	0.185	<u>1</u>	0.174	1	0.173	<u>2</u>	0.170	1	0.174	1
5E/df	0.213	2	0.239	1	0.219	2	0.230	1	0.246	1
S4a 40E/df	0,507	1	0,081	3	0,083	2	0,082	2	0,081	3
20E/df	0,515	1	0,127	2	0,123	2	0,122	2	0,126	2
10E/df	0,516	1	0,184	2	0,177	4	0,173	4	0,184	3
5E/df*§	0,338	<u>1</u>	0,342	1	0,32	<u>2</u>	0,355	1	0,368	1
S4b 40E/df	0,112	1	0,082	2	0,08	3	0,085	2	0,082	2
20E/df§	0,139	1	0,118	<u>2</u>	0,112	3	0,115	<u>3</u>	0,116	3
10E/df	0,137	1	0,144	1	0,141	1	0,141	1	0,142	1
5E/df	0,2	1	0,212	1	0,212	1	0,215	1	0,211	1
S4c 40E/df	0,749	1	0,1	2	0,098	2	0,093	2	0,1	2
20E/df	0,762	1	0,131	2	0,131	2	0,131	2	0,132	2
10E/df	0,739	1	0,192	2	0,189	2	0,188	2	0,188	2
5E/df§	0,742	1	0,296	2	0,288	<u>2</u>	0,278	<u>3</u>	0,283	<u>3</u>
S5a 40E/df§	0,969	1	0,1	2	0,099	<u>2</u>	0,098	2	0,098	<u>3</u>
20E/df	0,964	1	0,145	2	0,144	2	0,152	2	0,152	2
10E/df	1,007	1	0,215	3	0,221	2	0,208	3	0,209	3
5E/df§	0,954	1	0,289	2	0,289	<u>2</u>	0,282	<u>3</u>	0,275	<u>3</u>
S5b 40E/df	0,219	1	0,081	3	0,084	3	0,095	2	0,078	4
20E/df	0,201	1	0,122	3	0,128	3	0,131	2	0,129	3
10E/df	0,199	1	0,169	2	0,167	2	0,17	2	0,164	2
5E/df	0,247	1	0,22	2	0,215	2	0,233	2	0,233	2
S5c 40E/df	0,401	1	0,109	2	0,111	2	0,108	2	0,107	2
20E/df	0,424	1	0,174	2	0,179	2	0,169	2	0,171	2
10E/df	0,41	1	0,262	2	0,254	2	0,247	2	0,247	2
5E/df	0,388	1	0,305	2	0,306	2	0,315	2	0,32	2
S6a 40E/df*	0,098	1	0,076	2	0,078	2	0,082	2	0,075	2
20E/df§	0,12	<u>1</u>	0,103	<u>2</u>	0,109	<u>2</u>	0,117	<u>1</u>	0,112	1
10E/df	0,143	1	0,164	1	0,164	1	0,165	1	0,16	1
5E/df*	0,198	4	0,241	3	0,235	3	0,243	2	0,252	1
GESAMT	0.227	1	0.140	2	0.139	3	0.141	2	0.140	2

Rangordnung aus multipler Testprozedur. Höhere Rangzahlen verdeutlichen besseres Abschneiden.

§ Unterstrichene Rangunterschiede halten das multiple Testniveau nicht ein

* Im Innermethodenvergleich zeigte sich ein restriktiverer Schrankenwert überlegen (siehe Tab. 3-5)

Schlußfolgerungen

Keines der Verfahren erwies sich in allen Szenarien überlegen. Es lassen sich folgende Erkenntnisse für die Auswertungspraxis ableiten. Bei apriori Wissen über den Confounderstatus der Kandidatenvariablen gewährleistete Volladjustierung, selbst bei schwachen Confoundern und geringer Stichprobengröße, die beste Effektschätzung. Bei unsicherem Confounderstatus bestimmt dagegen die Fallzahl die Wahl des Auswertungsverfahrens.

Zwar lieferte das volle Modell bei Stichprobenumfängen mit 40 Ereignissen/df auch hier gute Ergebnissen. Bei geringer Fallzahl überwogen jedoch die Präzisions- und Validitätsvorteile gezielter Confounderselektion. Dabei offenbarte das p-CIE Verfahren eine Überlegenheit gegenüber p-Wert bzw. CIE gesteuerter Variablenselektion. Letztere erwies sich vor allem bei mehreren schwachen Confoundern bzw. einseitig mit der Exposition assoziierten Kovariablen als nachteilig. Ersteres könnte auf eine Addition von Schätzfehlern zurückzuführen sein, die dann erfolgt, wenn mehrere gleichgerichtete Confounder die vorgegebene CIE%-Schranke verfehlen. Letzteres dürfte auf mangelnde Robustheit gegenüber zufällig auftretende Assoziationen mit der Zielgröße zurückzuführen sein, die dazu führen, daß kollinearen Kovariablen fälschlich als Confounder ausgewählt werden. Dies zeigt auch, daß Change in Estimate zwar eine notwendige, aber keine hinreichende Bedingung für Confounding ist. Der Vorteil gegenüber der p-Wert Selektion zeigte sich vor allem bei Vorgabe nichtconfundierender Risikofaktoren, was ein geringeres Maß an Überadjustierungen vermuten läßt.

In den meisten Fällen schrittweiser Variablenselektion führte die Wahl liberaler Schrankenwerte ($\alpha=0,2$, CIE=5%) zu signifikant besseren Ergebnissen als das in Statistikpaketen voreingestellte α -Niveau von 0,05 oder die für das CIE Kriterium empfohlen 10% Schranke. Bei kleinen Stichproben und geringer Confoundingvorgabe erwiesen sich allerdings restriktivere Schranken als vorteilhaft.

Für weitergehendere Empfehlungen zur Wahl von Selektionsverfahren, sollten in zukünftigen Untersuchungen auch die Prävalenzparameter variiert werden.

Literatur

- 1) Greenland S.: Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340-349.
- 2) Maldonado G, Greenland S.: Simulation study of confounder-selection strategies. Am J Epidemiol 1993;138:923-936.