

ÜBERSICHTSARBEIT

Systematische Übersichtsarbeiten und Metaanalysen

Teil 6 der Serie zur Bewertung wissenschaftlicher Publikationen

Meike Rassing, Maria Blettner, Stefanie J. Klug

ZUSAMMENFASSUNG

Hintergrund: Aufgrund der zunehmenden Zahl an wissenschaftlichen Publikationen ist es wichtig, einzelne Studien zu einem Thema übersichtlich zusammenzufassen und gemeinsam zu bewerten. Immer häufiger werden daher systematische Übersichtsarbeiten, Metaanalysen publizierter Daten und Metaanalysen mit Individualdaten (gepoolte Reanalysen) publiziert. Im Folgenden werden die wesentlichen Methoden solcher Übersichtsarbeiten umrissen und Stärken und Probleme dargestellt.

Methoden: Diese Arbeit basiert auf einer selektiven Literaturrecherche. Es werden die verschiedenen Arten von Übersichtsarbeiten sowie ihre Gemeinsamkeiten und Unterschiede beschrieben. Ferner werden die jeweiligen angewendeten Methoden skizziert, um eine Bewertung zu ermöglichen. Zudem wird eine Checkliste zur Beurteilung von Zusammenfassungen wissenschaftlicher Artikel zur Verfügung gestellt.

Ergebnisse: Systematische Übersichtsarbeiten können einen Überblick über den Stand der Forschung zu einem bestimmten Thema geben. Außerdem kann anhand systematischer Übersichtsarbeiten die Qualität der einzelnen Studien bewertet werden. Sie erlauben eine Beurteilung der Ergebnisse bei inkonsistenter Datenlage. Metaanalysen ermöglichen zusätzlich die Berechnung von gepoolten Effektschätzern. Anhand aktueller Anwendungsbeispiele zu einem Thema (hier: Veröffentlichungen zur Assoziation zwischen der Einnahme von oralen Kontrazeptiva und der Entstehung des Zervixkarzinoms) werden die verschiedenen Formen von Zusammenfassungen dargestellt.

Schlussfolgerung: Systematische Übersichtsarbeiten ermöglichen die Zusammenfassung und Bewertung von Forschungsergebnissen oder Therapieeffekten aus verschiedenen Einzelstudien.

Dtsch Arztebl Int 2009; 106(27): 456–63
DOI: 10.3238/arztebl.2009.0456

Schlüsselwörter: Literaturrecherche, Übersichtsarbeiten, Metaanalysen, klinische Forschung, Epidemiologie

Institut für Medizinische Biometrie, Epidemiologie und Informatik, Universitätsmedizin der Johannes Gutenberg-Universität Mainz: Dr. med. Rassing, Prof. Dr. rer. nat. Blettner, PD Dr. rer. nat. et med. habil. Klug

Die Zahl an verfügbaren wissenschaftlichen Veröffentlichungen nimmt jedes Jahr stark zu. Während im Jahr 1987 noch 361 000 neue Publikationen in der Literaturdatenbank PubMed registriert wurden, waren es 1997 schon 448 000 und 2007 stieg die Anzahl auf 766 000 (Recherche in Medline, zuletzt aktualisiert Januar 2009). Diese Zahlen verdeutlichen die zunehmende Herausforderung für niedergelassene Ärzte, Kliniker und Wissenschaftler, sich über ein medizinisches Themengebiet umfassend und aktuell zu informieren. Daher ist es notwendig, einzelne Studien zur gleichen Thematik systematisch zusammenzufassen und kritisch zu bewerten.

Zusammenfassungen einzelner Studien werden meistens dann durchgeführt, wenn die Ergebnisse von Einzelstudien unübersichtlich oder inkonsistent sind. Sie werden außerdem eingesetzt, um Zusammenhänge zu untersuchen, für die die vorliegenden Einzelstudien zu einem Thema aufgrund geringer Fallzahlen keine ausreichende statistische Power aufweisen (1).

Die Cochrane Collaboration unternimmt für viele therapeutische Fragestellungen eine systematische Aufarbeitung und Zusammenfassung der Primärliteratur, vor allem für randomisierte klinische Studien (www.cochrane.org). Sie hat ein Handbuch für die Durchführung von systematischen Reviews und Metaanalysen randomisierter klinischer Studien herausgegeben (2). Von Cook und Kollegen wurden methodische Leitlinien zu diesem Thema publiziert (3). Solche Anleitungen helfen, Standards bezüglich der Durchführung einer Zusammenfassung von Einzelstudien darzulegen. Des Weiteren wurden Leitlinien für das Publizieren von Metaanalysen zu randomisierten klinischen Studien (4) und zu Beobachtungsstudien (5) erstellt.

Verschiedene Formen der Zusammenfassung von Publikationen zu Einzelstudien sind möglich (1, 6–10):

- narrative Reviews,
- systematische Übersichtsarbeiten,
- Metaanalysen publizierter Daten und
- gepoolte Reanalysen (Metaanalysen mit Individualdaten).

Diese Begriffe werden in der Literatur oft nicht eindeutig zugeordnet. Das Ziel dieses Artikels ist es, die

TABELLE 1

Verschiedene Formen der Zusammenfassung von Einzelstudien

	A Narrativer Review	B Systematische Übersichtsarbeit	C Metaanalyse publizierter Daten	D Gepoolte Reanalyse	E Prospektiv geplante Metaanalyse
Erstellung eines detaillierten Studienprotokolls und Auswertepans	-	+	+	+	+
Literaturrecherche geeigneter Studien nach a priori definierten Ein- und Ausschlusskriterien	-	+	+	+	+
quantitative Zusammenfassung der Ergebnisse (Berechnung gepoolter Schätzer, Untersuchung der Heterogenität, Sensitivitätsanalysen)	-	-	+	+	+
Analyse von Individualdaten	-	-	-	+	+
gemeinsames Studienprotokoll der Einzelstudien und prospektiv geplante Auswertung	-	-	-	-	+

einzelnen Formen zu beschreiben und voneinander abzugrenzen sowie dem Leser die Möglichkeit zu geben, die Ergebnisse der Einzelstudie in der Zusammenschau und die Qualität der Zusammenfassung kritisch zu beurteilen.

Methoden

Im Folgenden sollen die verschiedenen Formen der Zusammenfassung wissenschaftlicher Artikel definiert und das Vorgehen erläutert werden. Hierfür wurde eine selektive Literaturrecherche durchgeführt.

Unter einem Review versteht man eine qualitative Zusammenfassung der Ergebnisse einzelner Studien (1). Man unterscheidet narrative Reviews und systematische Übersichtsarbeiten (Synonym: systematischer Literaturreview, „systematic review“) (Tabelle 1). Narrative Reviews (A) bieten meist einen breiten Überblick zu einem bestimmten Thema (1, 11). Sie sind daher gut geeignet, sich schnell über den aktuellen Forschungsstand zu einem Thema zu informieren.

Die Auswahl der berücksichtigten Artikel erfolgt jedoch subjektiv und unsystematisch (1, 11). Im Deutschen Ärzteblatt wird seit einiger Zeit für diese Arbeiten der Begriff „selektive Literaturübersicht“ verwendet. Auf narrative Reviews wird im Folgenden nicht weiter eingegangen.

Systematische Übersichtsarbeiten (B) hingegen haben den Anspruch – unter Anwendung vorher definierter Ein- und Ausschlusskriterien – nach Möglichkeit alle publizierten Studien zu einem bestimmten Thema zu berücksichtigen (11). Es sollen relevante Informationen systematisch aus den Publikationen extrahiert werden. Wichtig sind eine Beurteilung der methodischen Qualität der eingeschlossenen Veröffentlichungen und eine Untersuchung der Gründe für eventuelle Unterschiede in den Ergebnissen der einzelnen Studien. Die Resultate jeder Studie werden dargestellt und nach definierten Kriterien wie Studiendesign und Rekrutierungsmodus bewertet.

Gleiches gilt für eine Metaanalyse publizierter Daten (C). Zusätzlich werden hier die Ergebnisse mittels statistischer Methoden quantitativ zusammengefasst, und es werden gepoolte Effektschätzer (Glossar) berechnet (1).

Eine gepoolte Reanalyse (D) ist eine quantitative Zusammenführung von Originaldaten (Glossar) einzelner Studien zur gemeinsamen Auswertung (1). Die Autoren jeder in die Arbeit eingeschlossenen Studie stellen hierfür Individualdaten (Glossar) zur Verfügung. Diese werden zu einer gemeinsamen Datenbank zusammengeführt und nach einheitlichen, a priori festgelegten Kriterien ausgewertet. Für diese Form der gepoolten Reanalyse werden auch die Begriffe „gepoolte Auswertung“ oder „Metaanalyse von Individualdaten“ benutzt.

Bei einer prospektiv geplanten Metaanalyse (E) sind eine Zusammenfassung der Einzelstudien und eine gemeinsame Auswertung schon bei der Planung der Einzelstudien vorgesehen. Entsprechend werden

KASTEN

Checkliste zur Beurteilung einer systematischen Zusammenfassung

- Lag ein a priori angefertigtes Studienprotokoll vor?
- Lag eine a priori festgelegte Hypothese vor?
- Wurde die Literaturrecherche detailliert beschrieben?
- Wurden a priori festgelegte Ein- und Ausschlusskriterien angewendet und klar beschrieben?
- Wurde eine möglicherweise vorhandene Heterogenität zwischen den Studien berücksichtigt?
- Wurden die verwendeten statistischen Methoden klar beschrieben?
- Wurden die Limitationen der jeweiligen Zusammenfassung diskutiert?

alle Einzelstudien einheitlich durchgeführt. Auf prospektiv geplante Metaanalysen wird im Folgenden nicht näher eingegangen.

Ein a priori erstelltes, detailliertes Studienprotokoll, in dem die Fragestellung, die Hypothesen, die Ein- und Ausschlusskriterien, die Auswahl der Studien, die Auswertung und gegebenenfalls die Datenzusammenführung und die Rekodierung der Individualdaten (nur bei der gepoolten Reanalyse) beschrieben werden, ist für alle Formen der Zusammenfassung außer bei einem narrativen Review unerlässlich.

Formen der Zusammenfassung von Studien

Im Folgenden wird die Vorgehensweise bei Zusammenfassungen von Studien vorgestellt (modifiziert nach [7, 10, 12, 13]). Dies soll dem Leser die Beurteilung ermöglichen, ob eine Zusammenfassung bestimmte Kriterien erfüllt (*Kasten*).

1. Wurde die Fragestellung a priori festgelegt?

Die Fragestellung der Zusammenfassung und die Hypothesen sollen klar definiert und in einem Studienprotokoll a priori schriftlich fixiert werden.

2. Wurden Ein- und Ausschlusskriterien a priori festgelegt?

Anhand von Ein- und Ausschlusskriterien wird darüber entschieden, ob die in der Literatursuche gefundenen Studien (siehe 3.) in die Zusammenfassung eingeschlossen werden.

3. Wurden Vorkehrungen getroffen, alle Studien aufzufinden, die zu der speziellen Fragestellung durchgeführt worden sind?

Es soll eine umfangreiche Literatursuche, möglichst in verschiedenen Literaturdatenbanken, nach Studien zum Thema vorgenommen werden. Zur Vermeidung einer Verzerrung sind alle relevanten Artikel unabhängig von der Sprache der Publikation zu berücksichtigen. Des Weiteren soll in den Literaturverzeichnissen der gefundenen Artikel und nach unveröffentlichten Studien in Kongressbänden sowie mittels Suchmaschinen im Internet recherchiert werden.

4. Wurden die relevanten Informationen aus den publizierten Artikeln extrahiert beziehungsweise wurden die Originaldaten zusammengeführt?

Bei einer systematischen Übersichtsarbeit (B) und bei einer Metaanalyse publizierter Daten (C) sollen relevante Informationen aus den Publikationen extrahiert werden.

Bei einer gepoolten Reanalyse (D) sind die Autoren aller identifizierten Studien anzuschreiben und um Bereitstellung der Individualdaten zu bitten. Diese Individualdaten sollen nach einheitlicher Festlegung kodiert und zu einer gemeinsamen Datenbank zusammengefasst und ausgewertet werden.

5. Wurde eine deskriptive Analyse der Daten durchgeführt?

Üblicherweise sollen in allen Formen der Zusammenfassung die wichtigsten Charakteristika der Einzelstu-

TABELLE 2

Wichtigste Charakteristika der eingeschlossenen Studien zum Zusammenhang zwischen oralen Kontrazeptiva und dem Zervixkarzinom (14)

Studie (Land)	Studien-design	Grad der zervikalen Neoplasie	HPV-Nachweis	Status Fälle/Kontrollen	Jahr der Diagnose (Fälle)	Einnahme hormoneller Kontrazeptiva bei Kontrollen	
						% jemals Einnahme oraler Kontrazeptiva (% Einnahme > 5 Jahre lang)	% jemals Einnahme injizierbarer Kontrazeptiva (% Einnahme > 5 Jahre lang)
Fall-Kontroll-Studien, einschließlich populations- (pop) und/oder klinikbasierter (hosp) Kontrollen							
Brinton/Jones, 1986 (USA) (e1, e2)	pop	invasiv/in situ	keiner	772 / 801	1982–1984	51 (18)	k. A.
Peters, 1986 (USA) (e3)	pop	invasiv*1	keiner	200 / 200	1980–1981	26 (k. A.)*2	k. A.
Ebeling, 1987 (Deutschland) (e4)	hosp	invasiv	keiner	129 / 275	1983–1985	66 (46)	k. A.
Brinton, 1990 (4 Länder*4) (e5, e6)	pop/hosp	invasiv	FISH	759 / 1 429	1986–1987	25 (11)	6 (1)
WHO, 1993 (9 Länder*5) (e7–e12)	hosp	invasiv/in situ	keiner	3 848 / 13 644	1979–1988	41 (8)	15 (4)*3
Ursin, 1994 (USA) (e13)	pop	invasiv*6	keiner	195 / 386	1977–1991	81 (36)	k. A.
Cuzick, 1996 (GB) (e14)	pop	invasiv	keiner	121 / 241	1985–1991	92 (62)	k. A.
Madeleine, 2001 (USA) (e15)	pop	in situ*6	PCR/Serologie	132 / 478	1990–1996	84 (29)	k. A.
Berrington, 2002 (GB) (e16)	pop	invasiv	Serologie	221 / 393	1984–1988	88 (47)	k. A.
Moreno, 2002 (8 Studien*7) (e17)	pop/hosp	invasiv/in situ*1	PCR	2 171 / 2 299	1985–1997	36 (11)	K. A.

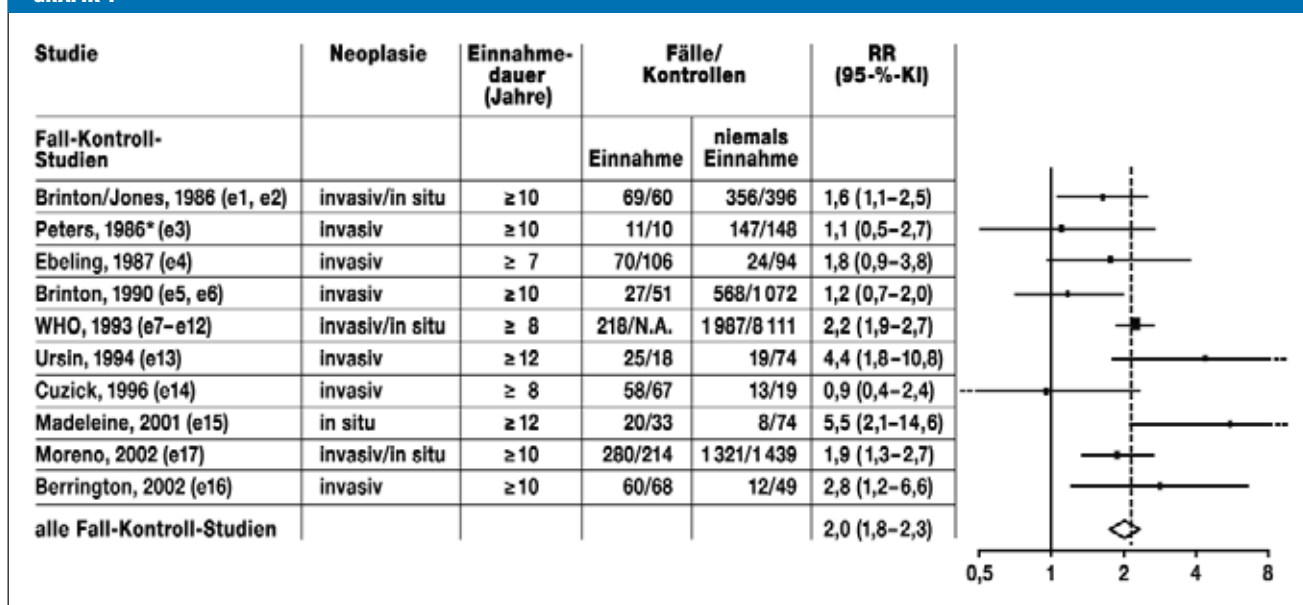
k. A., keine Angabe; FISH, Fluoreszenz in situ Hybridisierung; *1 nur Plattenepithelkarzinome; *2 jemals Einnahme → Einnahme mehr als 2 Jahre lang;

*3 relative Risiken für injizierbare Kontrazeptiva adjustiert für orale Kontrazeptiva-Einnahme; *4 Costa Rica, Kolumbien, Mexiko, Panama;

*5 Australien, Chile, Kolumbien, Israel, Kenia, Mexiko, Nigeria, Philippinen, Thailand; *6 nur Adenokarzinome der Zervix;

*7 Brasilien, Kolumbien, Marokko, Paraguay, Peru, Philippinen, Spanien, Thailand (gekürzter Nachdruck aus: Smith J, Green J, Berrington de Gonzalez A et al.: Cervical cancer and use of hormonal contraceptives: a systematic review. Lancet 2003; 361: 1159–67. Mit freundlicher Genehmigung des Elsevier Verlags.)

GRAFIK 1



Die Ergebnisse der Einzelstudien und der gemeinsamen Schätzer, dargestellt als „forest plots“, zum Zusammenhang zwischen oralen Kontrazeptiva und dem Zervixkarzinom als Beispiel für eine Metaanalyse publizierter Daten (14); N.A. = nicht vorhanden; * niemals Einnahme < 2 Jahre; KI = Konfidenzintervall (gekürzter Nachdruck aus: Smith J, Green J, Berrington de Gonzalez A et al.: Cervical cancer and use of hormonal contraceptives: a systematic review. Lancet 2003; 361: 1159–67. Mit freundlicher Genehmigung des Elsevier Verlags).

dien in Übersichtstabellen dargestellt werden. *Tabelle 2* steht beispielhaft für eine solche Tabelle aus einer Metaanalyse mit publizierten Daten (C) (14). Auf diese Weise können die Unterschiede zwischen den Studien bezüglich der untersuchten Daten deutlich gemacht werden.

6. Werden die Berechnung der Effektschätzer der Einzelstudien und die Berechnung der gepoolten Effektschätzer dargestellt?

Wie wurden die Effektschätzer der Einzelstudien berechnet? – Systematische Übersichtsarbeiten (B) beinhalten in der Regel Tabellen mit den Effektschätzern der Einzelstudien. Bei einer Metaanalyse publizierter Daten (C) werden die Effektschätzer der Einzelstudien (zum Beispiel Odds Ratio oder Relatives Risiko, siehe *Glossar*) entweder direkt aus den Publikationen extrahiert oder aus den Angaben in den jeweiligen Publikationen einheitlich neu berechnet (*Grafik 1*). Je nach Ausprägung von Einfluss- und Zielgrößen (binäre, kategorielle oder stetige Merkmale) wird sowohl bei der Metaanalyse publizierter Daten (C) als auch bei der gepoolten Reanalyse (D) meist ein logistisches oder lineares Regressionsmodell zur Berechnung der Effektschätzer der Einzelstudien angewendet.

Wie wurde der gepoolte Effektschätzer berechnet? – Die Effektschätzer der Einzelstudien werden mittels statistischer Verfahren zu einem gemeinsamen, gepoolten Effektschätzer zusammengefasst (9) (*Grafik 1*). Bei Metaanalysen mit publizierten Daten (C) wendet man meist zwei Methoden zur Berechnung eines gepoolten Effektschätzers an: das „fixed

effect“-Modell und das „random effect“-Modell (15, 16). Sie unterscheiden sich durch unterschiedliche Annahmen zur Heterogenität der Schätzer (siehe 7) zwischen den Einzelstudien. Die angewendete Methode soll in der Publikation genannt und begründet werden. Die Effektschätzer der Einzelstudien und die gepoolten Effektschätzer können in Form von sogenannten „forest plots“ (*Glossar*) grafisch dargestellt werden (*Grafik 1*) (14).

In gepoolten Reanalysen (D) wird der gepoolte Effektschätzer meist mittels logistischer oder linearer Regression berechnet. In der statistischen Auswertung muss allerdings die Herkunft der Datensätze aus verschiedenen Studien adäquat berücksichtigt werden. Die Ergebnisse von gepoolten Reanalysen können wie die Ergebnisse einer einzigen gemeinsamen Studie dargestellt werden (*Tabelle 3*).

7. Wurden Probleme bei der Interpretation von gepoolten Schätzern berücksichtigt?

Wurde die Heterogenität zwischen den Schätzern beachtet? – Die Schätzer der Einzelstudien können sich deutlich voneinander unterscheiden. Diese statistische Heterogenität (*Glossar*) zwischen den Studien kann durch Unterschiede wie verschiedenes Studiendesign, unterschiedliche Studienpopulationen (Alter, Geschlecht, Ethnie), unterschiedliche Methoden der Rekrutierung, der Diagnose und unterschiedliche Messmethoden bedingt sein (17, 18). Eine deskriptive Darstellung der methodischen Heterogenität zwischen den Studien kann mit einer Übersichtstabelle erfolgen, in der die wichtigsten Charakteristika der Einzelstudi-

TABELLE 3

Ergebnisse einer gepoolten Reanalyse zum Zusammenhang zwischen oralen Kontrazeptiva und dem Zervixkarzinom – ausgewertet und dargestellt ähnlich einer Einzelstudie (24)

Dauer der Einnahme	Zeit seit letzter Einnahme	Fälle/Kontrollen	Durchschnittliche Dauer der Einnahmen in Jahren (Fälle)	RR	95%-KI*	P-Wert*
niemals		7 356/21 682	–	1,00	–	–
5+ Jahre	zurzeit Einnahme	880/1 466	11,1	1,90	1,69–2,13	s.
	2–9 Jahre	747/1 510	9,3	1,28	k. A.	s.
	10+ Jahre	412/1 654	8,1	0,94	k. A.	n. s.

Trendtest: $\chi^2 = 66,2; p < 0,0001$

RR, relatives Risiko, adjustiert nach Alter, Studie oder Studienzentrum, Alter bei erstem Geschlechtsverkehr, Anzahl der Sexualpartner, Anzahl der ausgetragenen Schwangerschaften, Raucher- und Screeningstatus;

* Informationen aus der Publikation; KI, Konfidenzintervall; k. A., keine Angabe; s., signifikant zum Niveau $\alpha = 5\%$; n. s., nicht signifikant zum Niveau $\alpha = 5\%$

(gekürzter und modifizierter Nachdruck aus: International Collaboration of Epidemiological Studies of Cervical Cancer:

Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. Lancet 2007; 370: 1609–21. Mit freundlicher Genehmigung des Elsevier Verlags.)

en dargestellt werden (Tabelle 2). Die Heterogenität kann formal mithilfe statistischer Tests untersucht werden. Bei Vorliegen von signifikanter statistischer Heterogenität zwischen den Studien soll bei der Berechnung des gepoolten Schätzers statt des „fixed effect“-Modells das „random effect“-Modell verwendet werden (7, 15, 16). Es ist allerdings nicht eindeutig definiert, ab welcher Größenordnung die statistische Heterogenität zwischen den Studien zu groß ist, sodass man von der Berechnung eines gepoolten Effektschätzers absehen sollte (1, 19). Des Weiteren ist die Heterogenität zwischen den Studien mittels Subgruppenanalysen (Glossar) zu untersuchen. Hier werden zum Beispiel nur Studien mit gleichen Merkmalen der Studienpopulation – wie etwa homogene Altersgruppen, gleiche Ethnien oder gleiche histologische Befunde – gemeinsam ausgewertet. Außerdem können Studien mit gleichen Merkmalen – beispielsweise bezüglich der Studienqualität oder der Studiengröße – in Subgruppenanalysen separat berücksichtigt werden. Dies kann Hinweise darauf ergeben, ob sich der Einfluss des betreffenden Risikofaktors (Glossar) in einzelnen Subgruppen unterscheidet.

Wurden Sensitivitätsanalysen durchgeführt? – Sensitivitätsanalysen (Glossar) dienen, ebenso wie Subgruppenanalysen, dazu, die Stabilität des gepoolten Schätzers zu prüfen. Es besteht zum Beispiel die Möglichkeit, dass der gepoolte Effektschätzer hauptsächlich von einer großen Studie bestimmt wird. Wird diese Studie aus der Analyse ausgeschlossen, kann sich der gepoolte Effektschätzer ändern. Dies muss bei der Diskussion und Interpretation der Ergebnisse beachtet werden.

Wurde ein möglicher Publikationsbias berücksichtigt? – Ein Publikationsbias (Glossar) kann mittels eines „funnel plot“ (Glossar) dargestellt werden (7, 20–22). Grafik 2 zeigt ein Beispiel mit simulierten Daten. Im oberen „funnel plot“ (Grafik 2a) zeigt sich annähernd eine trichterförmige Verteilung der Effektschätzer

der Einzelstudien um den gepoolten Effektschätzer (mittlere gestrichelte Linie). Hier liegt kein Publikationsbias vor. Im unteren „funnel plot“ (Grafik 2b) fehlen die kleinen Studien, die in diesem Beispiel keine erhöhten Risiken zeigen. Daher besteht hier wahrscheinlich ein Publikationsbias, weil diese Studien nicht veröffentlicht worden waren.

8. Wie wurden die Ergebnisse interpretiert?

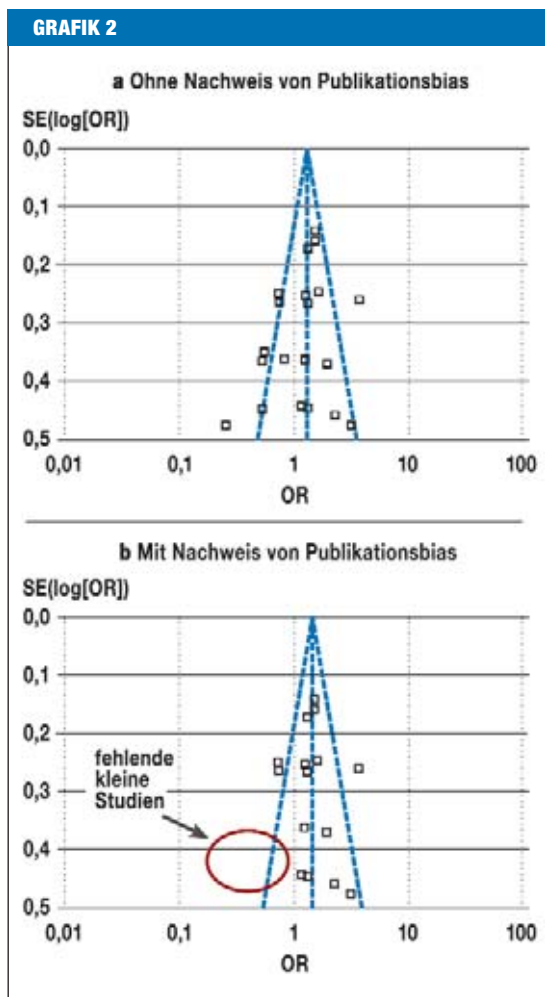
Bei der Interpretation der Ergebnisse sollen mögliche Limitationen diskutiert und berücksichtigt werden. Die Aussagekraft der Ergebnisse von Zusammenfassungen kann beispielsweise durch unzureichende Qualität der Einzelstudien – wie zum Beispiel eine starke Selektion der Studienpopulation oder durch aggregierte Daten (Glossar) – limitiert sein.

Ergebnisse

Aus den im Methodenteil genannten Einzelschritten wurden die relevanten Punkte extrahiert, die bei der Zusammenfassung wissenschaftlicher Artikel beachtet werden müssen (Kasten). Diese Checkliste kann auch herangezogen werden, um die Qualität einer systematischen Übersichtsarbeit oder Metaanalyse zu beurteilen.

Als Beispiele für die Durchführung einer systematischen Literaturübersicht (B), einer Metaanalyse mit publizierten Daten (C) und einer gepoolten Reanalyse (D) dienen Veröffentlichungen zur Assoziation zwischen der Einnahme von oralen Kontrazeptiva und der Entstehung des Zervixkarzinoms. Dieser Zusammenhang wird seit langer Zeit wissenschaftlich untersucht.

Im Jahr 1996 wurde eine systematische Übersichtsarbeit (B) von La Vecchia und Kollegen mit sechs Studien zu diesem Thema veröffentlicht (23). Es wurden verschiedene Informationen zu den Einzelstudien in einer Übersichtstabelle dargestellt. Ein gepoolter Effektschätzer wurde nicht berechnet.



Visualisierung von Publikationsbias mittels „funnel plots“ aus simulierten Daten a) kein Publikationsbias; b) Publikationsbias. SE = Standardfehler; OR = Odds Ratio

Im Jahr 2003 erschien eine Metaanalyse publizierter Daten (C) von Smith und Kollegen mit 28 Studien zum selben Thema (14). Die eingeschlossenen Studien wurden zunächst ebenfalls in einer deskriptiven Übersicht dargestellt, wie sie auch in systematischen Übersichtsarbeiten üblich ist (Tabelle 2). In der Übersichtstabelle zu dieser Zusammenfassung wird die methodische Heterogenität (Glossar) zwischen den Studien deutlich, beispielsweise bei den Unterschieden bezüglich des HPV-Nachweises (Tabelle 2). Die Heterogenität wurde außerdem formal mittels statistischer Tests untersucht, und es erfolgten verschiedene Subgruppenanalysen. Zusätzlich wurden im Unterschied zur systematischen Übersichtsarbeit (B) von La Vecchia und Kollegen auch gepoolte Effektschätzer mit den publizierten Daten berechnet (Grafik 1). Die Effektschätzer für die Einzelstudien und die gepoolten Effektschätzer mit ihren Konfidenzintervallen (Glossar) wurden mithilfe eines „forest plot“ dargestellt (Grafik 1).

Im Jahr 2007 wurde eine gepoolte Reanalyse (D) mit 24 Studien, von denen die Originaldaten vorlagen, zum

selben Thema veröffentlicht (24). Die gepoolten Effektschätzer wurden im Gegensatz zur Metaanalyse publizierter Daten aus den Originaldaten berechnet und ausschließlich gemeinsam dargestellt (Tabelle 3). Dies ist in detaillierter und einheitlicher Form nur in einer solchen gepoolten Reanalyse möglich, weil hier die Originaldaten mit den genauen Angaben zu allen Parametern für jede einzelne Studienteilnehmerin vorliegen. Allerdings muss auch hier die Herkunft der Individualdaten (Glossar) aus unterschiedlichen Studien berücksichtigt werden.

Diskussion

Systematische Übersichtsarbeiten (B) können einen umfassenden Überblick über den gegenwärtigen Forschungsstand geben (1). Notwendig sind sie auch für die Entwicklung von S2- und S3-Leitlinien im Rahmen der formalen Evidenzrecherche (25). Metaanalysen publizierter Daten (C) werden durchgeführt, um zusätzlich gepoolte Effektschätzer aus den Einzelstudien zu berechnen (1). Sie sind, wie auch systematische Übersichtsarbeiten, unabhängig von der Kooperationsbereitschaft der Autoren der Originalstudien möglich.

Die berechneten gepoolten Effektschätzer können aus verschiedenen Gründen eine begrenzte Aussagekraft haben. Erstens ist nicht klar definiert, bis zu welcher Größenordnung die Heterogenität zwischen den Studien vernachlässigbar und dann die Berechnung eines gepoolten Effektschätzers sinnvoll ist (1, 19). Sind die Einzelstudien zu heterogen, sollte von der Berechnung eines gepoolten Effektschätzers abgesehen werden. Zweitens wird der gepoolte Effektschätzer meist aus aggregierten Daten berechnet. Subgruppenanalysen und die Berücksichtigung potenzieller Störgrößen (Confounder; Glossar) können oft nur bedingt oder gar nicht erfolgen (1, 19). Drittens besteht auch bei der Metaanalyse publizierter Daten das Problem des Publikationsbias.

In einer gepoolten Reanalyse (D) können potenzielle Confounder und Risikofaktoren, zu denen die Angaben in den Publikationen meist nur aggregiert vorliegen, besser berücksichtigt werden (7). Mithilfe der Individualdaten können die in die Analyse eingegangenen Zielgrößen, Einflussgrößen und Confounder einheitlich kategorisiert und bei der Auswertung entsprechend beachtet werden. Es können, wenn entsprechend a priori im Studienprotokoll festgelegt, einzelne Datensätze herausgenommen werden, ohne eine ganze Studie ausschließen zu müssen. Nachteile der gepoolten Reanalyse sind zum einen die hohe Zeit- und Kostenintensität, zum anderen ist diese Art der Zusammenfassung auf die Kooperationsbereitschaft der Autoren der Einzelstudien angewiesen. Wenn nicht alle Autoren ihre Individualdaten schicken, kann eine Verzerrung der Ergebnisse die Folge sein.

Der Evidenzgrad der Art der Zusammenfassung nimmt von der systematischen Übersichtsarbeit über die Metaanalyse publizierter Daten bis zur gepoolten Reanalyse zu. Bei allen drei Formen der Zusammenfassung ist wichtig, dass sie mit hoher Qualität durchgeführt werden.

Klinische Kernaussagen

- Die verschiedenen Formen der Zusammenfassung lassen sich einteilen in: systematische Übersichtsarbeiten, Metaanalysen publizierter Daten und gepoolte Reanalysen.
- Systematische Übersichtsarbeiten können einen schnellen Überblick über den Stand der Forschung zu einem bestimmten Thema geben.
- Metaanalysen publizierter Daten und gepoolte Reanalysen ermöglichen zusätzlich die Berechnung von gepoolten Effektschätzern.
- Gepoolte Reanalysen erlauben eine detaillierte Auswertung auf Basis von Individualdaten.
- Alle genannten Formen der Zusammenfassung benötigen – wie jede Originalstudie – ein a priori angefertigtes Studienprotokoll, in dem die Fragestellung, die Hypothese, die Literaturrecherche, die Ein- und Ausschlusskriterien und die Auswertungsstrategien detailliert festgelegt werden.

Glossar

● Aggregierte Daten

Unter aggregierten Daten versteht man eine Zusammenfassung von Individualdaten.

● Bias

Verzerrung der Studienergebnisse durch systematische Fehler

● Effektschätzer

Ein Effektschätzer, z. B. Odds Ratio oder Relatives Risiko, schätzt das Ausmaß der Änderung in der Erkrankungshäufigkeit, welches von einer bestimmten Exposition verursacht wird.

● Exposition

Kontakt mit einem bestimmten Risikofaktor

● Forest Plot

Ein „forest plot“ ist eine grafische Darstellung der Effektschätzer der Einzelstudien und zusätzlich des gepoolten Schätzers. In der Regel ist der Effektschätzer jeder Einzelstudie mit Konfidenzintervall auf einer horizontalen oder vertikalen Achse repräsentiert. Je größer die Fläche der Effektschätzer der Einzelstudie ist, desto größer ist das Gewicht der Studie, u. a. bedingt durch die Studiengröße. Die gepoolten Effektschätzer werden meist in Form eines Diamanten dargestellt.

● Funnel Plot

In einem „funnel plot“ (Trichtergrafik) wird die Studiengröße gegen die Effektschätzer der Einzelstudien aufgetragen. Anstelle der Studiengröße werden meist die Varianzen oder die Standardfehler der Effektschätzer der Einzelstudien dargestellt. Je größer die Varianz bzw. der Standardfehler des Effektschätzers der Einzelstudie ist, desto kleiner ist die Studie. Die Effektschätzer von großen Studien streuen weniger um den gepoolten Effektschätzer als die Effektschätzer von kleinen Studien. Daher ergibt sich das Bild eines Trichters („funnel“). Ein Publikationsbias wird mithilfe von „funnel plots“ visualisiert.

● Heterogenität

Die statistische Heterogenität beschreibt die Unterschiede zwischen den Studien bezüglich der Effektschätzer. Diese kann durch methodische Heterogenität zwischen den Studien, also z. B. Unterschiede in der Studienpopulation, der Studiengröße oder bei den verwendeten Messmethoden, bedingt sein.

● Individualdaten

Bei Individualdaten liegen alle Angaben (z. B. Alter, Geschlecht, Diagnose) auf der Ebene der Einzelpersonen vor.

● Konfidenzintervall

Das Konfidenzintervall (Vertrauensbereich) ist ein Bereich, in dem mit einer bestimmten Wahrscheinlichkeit, in der Regel 95 %, der wahre Wert liegt.

● Odds Ratio

Odds (Chancenverhältnis) bedeutet in der Medizin und in der Epidemiologie ein Quotient aus der Wahrscheinlichkeit, exponiert zu sein, und der Wahrscheinlichkeit, nicht exponiert zu sein. Der Quotient aus den Odds der Fälle und den Odds der Kontrollen ergibt das Odds Ratio. Bei seltenen Erkrankungen ist das Odds Ratio eine Näherung für das Relative Risiko.

● Originaldaten

s. Individualdaten

● Publikationsbias

Publikationsbias bedeutet, dass Studien, die keinen Einfluss der Exposition auf die Zielerkrankung nachweisen konnten („negative“ Studien), seltener veröffentlicht werden als Studien, die eine positive oder statistisch signifikante Assoziation nachweisen konnten. Ein Publikationsbias kann mithilfe von „funnel plots“ visualisiert werden.

● Risikofaktor

Ein Risikofaktor modifiziert die Wahrscheinlichkeit für die Entstehung einer bestimmten Erkrankung. Dies kann z. B. ein äußerer Umwelteinfluss oder eine Prädisposition der Person sein.

● Relatives Risiko

Zur Berechnung des Relativen Risikos wird die Wahrscheinlichkeit für Exponierte, zu erkranken, durch die Wahrscheinlichkeit für Nichtexponierte, zu erkranken, geteilt. Die Berechnung des Relativen Risikos basiert auf Neuerkrankungen.

● Sensitivitätsanalysen

Mithilfe von Sensitivitätsanalysen wird untersucht, ob der Ausschluss einzelner Studien aus der Analyse einen Einfluss auf das Ergebnis des gepoolten Effektschätzers hat. Dadurch wird die Stabilität des gepoolten Effektschätzers geprüft.

● Störgröße („confounder“)

Ein „confounder“ ist ein Faktor, der sowohl mit der untersuchten Erkrankung als auch mit der untersuchten Exposition in Zusammenhang steht. Er kann daher den wahren Zusammenhang zwischen der Erkrankung und der Zielgröße verstärken oder abschwächen.

● Subgruppenanalysen

Hier werden verschiedene Untergruppen der Studienpopulation, z. B. eine homogene ethnische Gruppe, in Analysen separat berücksichtigt.

Interessenkonflikt

Die Autorinnen erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors besteht.

Manuskriptdaten

eingereicht: 9. 6. 2008, revidierte Fassung angenommen: 10. 3. 2009

LITERATUR

- Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C: Vergleich von traditionellen Reviews, Metaanalysen und gepoolten Analysen zur Bewertung von Risikofaktoren. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 1997; 28: 148–66.
- Higgins JPT, Green S (Hrsg.): *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.1 [aktualisiert September 2008]. The Cochrane Collaboration 2008. www.cochrane-handbook.org.
- Cook DJ, Sackett DL, Spitzer WO: Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995; 48: 167–71.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999; 354: 1896–900.
- Stroup DF, Berlin JA, Morton SC et al.: Meta-analysis of observational studies in epidemiology. *JAMA* 2000; 263: 2008–12.
- Sutton AJ, Higgins JP: Recent developments in meta-analysis. *Stat Med* 2008; 27: 625–50.
- Blettner M, Schlattmann P: Meta-analysis in epidemiology. In: Ahrens W, Pigeot I (Hrsg.): *Handbook of epidemiology*. Berlin, Heidelberg, New York: Springer 2005; 829–59.
- Egger M, Smith GD, Phillips AN: Meta-analysis: principles and procedures. *BMJ* 1997; 315: 1533–7.
- Petitti DB: Meta-analysis, decision analysis, and cost-effectiveness analysis. *Methods for quantitative synthesis in medicine*. 2. Auflage New York: Oxford University Press 2000.
- Friedenreich CM: Methods for pooled analyses of epidemiologic studies. *Epidemiology* 1993; 4: 295–302.
- Montori VM, Swiontkowski MF, Cook DJ: Methodologic issues in systematic reviews and meta-analyses. *Clin Orthop Relat Res* 2003; 413: 43–54.
- Meerpohl JJ, Timmer A, Antes G: Systematische Übersichten als Entscheidungshilfe für die Praxis. *Wie Sie die Informationsflut sinnvoll nutzen*. *MMW Fortschr Med* 2008; 150: 41–3.
- Khan KS, Kunz R, Kleijnen J, Antes G: Five steps to conducting a systematic review. *J R Soc Med* 2003; 96: 118–21.
- Smith J, Green J, Berrington de Gonzalez A et al.: Cervical cancer and use of hormonal contraceptives: a systematic review. *Lancet* 2003; 361: 1159–67.
- DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–88.
- Normand SL: Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999; 18: 321–59.
- Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539–58.
- Deeks JJ, Altman DG, Bradburn MJ: Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG (Hrsg.): *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group 2001; 285–312.
- Blettner M, Schlehofer B, Sauerbrei W: Grenzen von Metaanalysen aus publizierten Daten bei epidemiologischen Fragestellungen. *Soz Präventivmed* 1997; 42: 95–104.
- Dickersin K: The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1385–9.
- Ioannidis JP: Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; 279: 281–6.
- Dwan K, Altman DG, Arnaiz JA et al.: Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008; 3: e3081.
- La Vecchia C, Tavani A, Franceschi S, Parazzini F: Oral contraceptives and cancer. A review of the evidence. *Drug Saf* 1996; 14: 260–72.
- International Collaboration of Epidemiological Studies of Cervical Cancer: Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. *Lancet* 2007; 370: 1609–21.
- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Erarbeitung von Leitlinien für Diagnostik und Therapie: *Methodische Empfehlungen („Leitlinie für Leitlinien“*, Stand Dez. 2004). www.uni-duesseldorf.de/awmf/II/II_metho.htm.

Anschrift für die Verfasserinnen

PD Dr. rer. nat. et med. habil. Stefanie Klug, MPH
 Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)
 Universitätsmedizin der Johannes Gutenberg-Universität Mainz
 Obere Zahlbacher Straße 69
 55101 Mainz
 E-Mail: klug@imbei.uni-mainz.de

SUMMARY**Systematic Literature Reviews and Meta-Analyses—Part 6 of a Series on Evaluation of Scientific Publications**

Background: Because of the rising number of scientific publications, it is important to have a means of jointly summarizing and assessing different studies on a single topic. Systematic literature reviews, meta-analyses of published data, and meta-analyses of individual data (pooled reanalyses) are now being published with increasing frequency. We here describe the essential features of these methods and discuss their strengths and weaknesses.

Methods: This article is based on a selective literature search. The different types of review and meta-analysis are described, the methods used in each are outlined so that they can be evaluated, and a checklist is given for the assessment of reviews and meta-analyses of scientific articles.

Results: Systematic literature reviews provide an overview of the state of research on a given topic and enable an assessment of the quality of individual studies. They also allow the results of different studies to be evaluated together when these are inconsistent. Meta-analyses additionally allow calculation of pooled estimates of an effect. The different types of review and meta-analysis are discussed with examples from the literature on one particular topic.

Conclusions: Systematic literature reviews and meta-analyses enable the research findings and treatment effects obtained in different individual studies to be summed up and evaluated.

Dtsch Arztebl Int 2009; 106(27): 456–63
 DOI: 10.3238/arztebl.2009.0456

Key words: literature search, systematic review, meta-analysis, clinical research, epidemiology



Mit „e“ gekennzeichnete Literatur:
www.aerzteblatt.de/lit2709

The English version of this article is available online:
www.aerzteblatt-international.de

ÜBERSICHTSARBEIT

Systematische Übersichtsarbeiten und Metaanalysen

Teil 6 der Serie zur Bewertung wissenschaftlicher Publikationen

Meike Rassing, Maria Blettner, Stefanie J. Klug

eLITERATUR

- e1. Brinton LA, Huggins GR, Lehmann HF et al.: Long-term use of oral contraceptives and risk of invasive cervical cancer. *Int J Cancer* 1986; 38: 339–44.
- e2. Jones CJ, Brinton LA, Hammann RF et al.: Risk factors for in situ cervical cancer: results from a case-control-study. *Cancer Res* 1990; 50: 3657–62.
- e3. Peters RK, Thomas D, Hagan DG, Mack TM, Henderson BE: Risk factors for invasive cervical cancer among Latinas and non-Latinas in Los Angeles County. *J Natl Cancer Inst* 1986; 77: 1063–77.
- e4. Ebeling K, Nischan P, Schindler C: Use of oral contraceptives and risk of invasive cervical cancer in previously screened women. *Int J Cancer* 1987; 39: 427–30.
- e5. Brinton LA, Reeves WC, Brenes MM et al.: Oral contraceptive use and risk of invasive cervical cancer. *Int J Epidemiol* 1990; 19: 4–11.
- e6. Herrero R, Brinton LA, Reeves WC et al.: Injectable contraceptives and risk of invasive cervical cancer: evidence of an association. *Int J Cancer* 1990; 46: 5–7.
- e7. WHO: Invasive squamous-cell cervical carcinoma and combined oral contraceptives: results from a multinational study. WHO collaborative study of neoplasia and steroid contraceptives. *Int J Cancer* 1993; 55: 228–36.
- e8. WHO: Depot-medroxyprogesterone acetate (DMPA) and risk of invasive squamous cell cervical cancer. The WHO collaborative study of neoplasia and steroid contraceptives. *Contraception* 1992; 45: 299–312.
- e9. Thomas DB, Ray RM: Depot-medroxyprogesterone acetate (DMPA) and risk of invasive adenocarcinomas and adenosquamous carcinomas of the uterine cervix. WHO collaborative study of neoplasia and steroid contraceptives. *Contraception* 1995; 52: 307–12.
- e10. Thomas DB, Ye Z, Ray RM: Cervical carcinoma in situ and the use of depotmedroxyprogesterone acetate (DMPA). WHO collaborative study of neoplasia and steroid contraceptives. *Contraception* 1995; 51: 25–31.
- e11. Ye Z, Thomas DB, Ray RM: Combined oral contraceptives and risk of cervical carcinoma in situ. WHO collaborative study of neoplasia and steroid contraceptives. *Int J Epidemiol* 1995; 24: 19–26.
- e12. Thomas DB, Ray RM: Oral contraceptives and invasive adenocarcinomas and adenosquamous carcinomas of the uterine cervix. The WHO collaborative study of neoplasia and steroid contraceptives. *Am J Epidemiol* 1996; 144: 281–89.
- e13. Ursin G, Peters RK, Henderson BE, d'Ablaing G, Monroe KR, Pike MC: Oral contraceptive use and adenocarcinoma of the cervix. *Lancet* 1994; 344: 1390–4.
- e14. Cuzick J, Sasieni P, Singer A: Risk factors for invasive cervical cervix cancer in young women. *Eur J Cancer* 1996; 32A: 836–41.
- e15. Madeleine MM, Daling JR, Schwartz SM et al.: Human papillomavirus and long-term oral contraceptive use increase the risk of adenocarcinoma in situ of the cervix. *Cancer Epidemiol Biomarkers Prev* 2001; 10: 171–77.
- e16. Berrington A, Jha PK, Peto J, Green J, Hermon C: UK national case-control study of cervical cancer: Oral contraceptives and cervical cancer. *Lancet* 2002; 360: 410.
- e17. Moreno V, Bosch FX, Munoz N et al.: Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case control study. *Lancet* 2002; 359: 1085–92.