Oral Contraceptive Use and Cardiovascular Disease: Is the Relationship Real or Due to Study Bias?

David A. Katerndahl, MD, MA; Janet P. Realini, MD; and Peter A. Cohen, PhD
San Antonio, Texas, and Augusta, Georgia

Background. Epidemiologic studies link oral contraceptive use with several cardiovascular events, but the literature is difficult to summarize, and potential biases remain poorly addressed. This study uses meta-analysis to summarize study results and to analyze the influence of study characteristics, including susceptibility to bias, on study outcome.

Methods. Forty-seven case-control and cohort studies of oral contraceptives and four cardiovascular events were coded for relative risk (RR) and study characteristics, including adherence to 14 bias-control standards. Key RRs were pooled to summarize findings for each disease type. Univariate determinants of the magnitude of the relative risks were identified, and partial correlation analysis was performed for each disease type.

Results. Relative risks were significantly greater than 1.0 for venous thromboembolism (RR = 2.8, CI = 2.4 to 3.2), stroke (RR = 1.8, CI = 1.6 to 2.0), and myocardial infarction (RR = 1.6, CI = 1.4 to 1.8), but not for death due to any cardiovascular cause (RR = 1.0, CI = 0.8 to 1.3). Study characteristics were diverse, and potential biases were frequently uncontrolled. For three of ten study characteristics identified as independently influencing relative risk, methodologically stronger studies of venous thromboembolism tended to have higher RRs. The RRs for stroke and myocardial infarction were lower in studies that were methodologically stronger with regard to variables identified as important. In studies of cardiovascular death, bias-control standards identified as important were generally well addressed by the studies.

Conclusions. Oral contraceptive use does not appear to increase overall cardiovascular mortality. The associations noted with stroke and myocardial infarction may be due to methodologic flaws in the studies, while the association with venous thromboembolism is more likely to be valid.

Key words. Meta-analysis; contraceptives, oral; cardiovascular diseases; research design. *J Fam Pract 1992; 35:147-157.*

Numerous epidemiologic studies of the relationship between the use of oral contraceptives and cardiovascular disease have been published. Although several reviews of this vast literature have been published,[1–4] it is difficult to summarize and compare study findings. Different studies examine different disease types (eg, venous thromboembolism, stroke, myocardial infarction, overall cardiovascular mortality), disease subtypes (eg, subarachnoid hemorrhage and thrombotic stroke), patient populations, and controls. In addition, studies differ in their study type (case-control vs cohort), in the exclusions used to define the groups of women compared, and in their susceptibility to potential sources of bias. It is thus difficult for a critical reader to digest all of these studies and come to a meaningful understanding of what is known about oral contraceptives and cardiovascular diseases.

It has been generally accepted in the medical community that oral contraceptives are associated with some cardiovascular diseases. One review and evaluation of the literature, however, noted that important potential biases remain unaddressed by epidemiologic studies.[5] Specifically, bias in the detection of cardiovascular events may have influenced nearly all case-control and cohort studies of oral contraceptives and cardiovascular diseases. Seldom do the epidemiologic studies ensure equal susceptibility to cardiovascular disease among patients in the control and study groups. In addition, bias in the method of obtaining the patient's history of oral contraceptive use is an unresolved issue for most of the case-control studies.[5]

Using meta-analysis adds a quantitative dimension

Submitted, revised, May 19, 1992.

From the Department of Family Practice, University of Texas Health Science Center, San Antonio (Drs Katerndahl and Realini); and Educational Development, Medical College of Georgia School of Dentistry, Augusta (Dr Cohen). Requests for reprints should be addressed to David A. Katerndahl, MD, MA, Department of Family Practice, 7703 Floyd Curl Drive, San Antonio, TX 78284-7795.

 ISSN 0094-3509

to the assessment of a body of literature.[6–12] Meta-analysis is ordinarily used to synthesize results from several studies. In addition, through systematic evaluation of the characteristics of studies, meta-analysis can identify determinants of the magnitude of the relative risk of a set of studies. In this way, meta-analysis can help explain conflicting results in the literature. Some researchers believe meta-analysis will revolutionize how medical researchers assess several sets of research data.[13]

The purpose of the meta-analysis undertaken for this paper was to summarize the epidemiologic literature on oral contraceptive use and cardiovascular disease, and to investigate whether characteristics of the methods used in the study might affect the relative risk found. Because bias is such a common problem in even the largest and best-designed studies, we evaluated each study's adherence to standards for avoiding potential bias as described by Feinstein and Horwitz.[14–18] We were specifically interested in whether the susceptibility of a study to potential bias might affect the degree of association found between oral contraceptive use and cardiovascular events. We examined case-control and cohort studies of oral contraceptive use as it related to four types of cardiovascular events: venous thromboembolism, stroke, myocardial infarction, and death due to any cardiovascular cause. The meta-analysis reported herein was the first to evaluate this body of literature and the first to attempt to quantify the impact of bias on outcome.

Methods

Data Acquisition

Potential English-language case-control and cohort studies for inclusion were identified through a MEDLINE computer search (1960 through 1989), a *Dissertations Abstracts* computer search (1960 through 1987), a University Microfilms search, references from articles, verbal and written communications with investigators, and written communication with pharmaceutical companies that manufacture oral contraceptives. The term *oral contraceptive*, the various hormonal classes and specific agents, and *cardiovascular disease and mortality*, *myocardial infarction*, *cerebrovascular accident*, *subarachnoid hemorrhage*, *pulmonary embolism*, *thrombosis*, *phlebitis*, and *thrombophlebitis* were used in the computer searches. The published literature included 280 potential articles representing 201 studies. In addition, after review of abstracts from 1207 citations from *Dissertations Abstracts* and University Microfilms, copies of four dissertations were obtained as potentially relevant studies.

To be considered for inclusion, the study had to use women on oral contraceptives and deal with deep venous thrombosis, stroke, myocardial infarction, or death from all cardiovascular causes. Studies were excluded if there was no documentation of the presence of a control group or of direct oral contraceptive exposure. The findings of a study, ie, the relative risk, were not considered in the decision about inclusion. None of the four dissertations were deemed appropriate for inclusion. Of the 201 potential published studies, 47 appropriate studies were identified and selected for inclusion in the meta-analysis (see Appendix).*

Data Extraction

Studies were assessed by two independent reviewers (D.A.K. and J.P.R.) using a standardized coding sheet. The major outcome variable was relative risk (RR), directly measured by cohort studies and estimated by case-control studies. The variables investigated are listed in Table 1. The reviewers independently rated each study's compliance with standards designed to avoid bias in epidemiologic studies (Table 2).[20,21] Control of bias was rated "0" if the standard for evaluation was not met, and "2" if it was met. A "1" was assigned if it was unclear whether the standard had been met or if little or no bias was likely.[5] When ratings differed, a consensus rating was negotiated.

Because this assessment was subjective, interrater agreement was measured using percent agreement and kappa[22] statistics. Interrater agreement for coding was generally good. Of the 14 standards coded, five had kappas greater than 0.75 (excellent agreement), two had kappas between 0.60 and 0.74 (good agreement), and three had kappas between 0.40 and 0.59 (fair agreement).[23] Two of the remaining four standards, predetermined selection and equal clinical susceptibility, had kappas that were statistically significant ($\kappa = 0.384$, $P = .0006$, and $\kappa = 0.359$, $P = .0409$, respectively). The other two standards over which there was poor agreement, defined exposure and representativeness, were not met in any of the studies. This low rate could lead to falsely low kappa levels. Correction using Y statistics[24] showed that kappas increased for both standards, reaching 0.4 in the case of representativeness. Although kappa increased to 0.3 with calculation of the Y statistic for defined exposure, this still suggests poor agreement. However, in 96.7% of the studies assessed, the raters were within one category of each other, which we defined as "near agreement."

**A reference list of excluded studies is available from the corresponding author upon request.*

Table 1. Variables Investigated

- Relative risk
  - Relative risk
  - Standard error
  - Adjustment of relative risk
- Predisposing characteristics
  - Smoking status
  - Operative status
  - Other predisposition to the disease under study
- Cardiovascular disease
  - Type (eg, venous thromboembolism, stroke, myocardial infarction, cardiovascular death)
  - Subgroups (eg, subarachnoid hemorrhage, cerebral embolism)
  - Method of diagnosis (eg, clinical, autopsy)
  - Fatal vs nonfatal
- Demographics of sample
  - Minimum and maximum age
  - Education level (median years)
  - Race (percent white)
  - Socioeconomic status
  - Geographic location
  - Practice site (eg, hospital, family planning clinic)
- Study characteristics
  - Type (eg, case-control, cohort)
  - Matching of controls
  - Origin of controls (eg, inpatient, community)
  - Sample size
  - Statistical power
  - Duration of follow-up
  - Dropout or exclusion rate
  - Control of bias (*See* Table 2)
- Pill characteristics
  - Type of use (eg, past use, current use)
  - Progestin content
  - Estrogen content
- Publication characteristics
  - Journal type (eg, general medical journal, cardiology journal)
  - Country of journal origin
  - Year of publication
  - Journal impact[19]
  - Journal immediacy[19]

To avoid the effect of nonindependence of multiple RRs per study, one key RR per disease type per study was selected as most representative for summary purposes. Relative risks preferred for selection as key RRs were those that represented all of the age groups studied, current use of oral contraceptives, the diagnoses closest to the disease category of interest, nonpredisposed patients, the most comparable control group available, and statistical adjustment. As with all of the variables assessed, when the reviewers did not agree completely, a consensus was negotiated.

When standard errors for RRs were not reported, Woolf's method was used to calculate standard error.[25] Relative risks having at least one cell with a frequency of 0 were calculated by adding 0.5 to each cell as recommended by Feinstein.[26] Studies with any cell containing less than five subjects were defined as "small."[27]

Statistical power was calculated using the formulas of Schlesselman.[28] In this calculation, the incidences of disease in the nonexposed population were based on the Royal College of General Practitioners studies,[29–30] and the prevalence of oral contraceptive exposure was based on the 1982 Ortho Birth Control Study and the 1982 National Survey of Family Growth.[31] For the purpose of calculating power, the minimum RR considered important was 2.0.

Data Analysis

Key RRs were pooled to summarize findings for each disease type using the natural logarithms of the relative risk (LnRR). Weighted means were also calculated using methods described by Greenland.[27]

In an attempt to identify which study characteristics were significant in determining the magnitude of the RR, univariate statistics were performed. For each disease type, the natural logarithm of each key and nonkey relative risk (LnRR) was compared with each study characteristic using t tests, analyses of variance (ANOVAs), and Pearson (for interval or ratio variables) or Spearman (for ordinal variables) correlations, seeking a $P \leq .05$. Once univariate analysis had identified potential key characteristics, the effects of interactions between characteristics needed to be removed. Although regression analysis would be preferable for development of such predictive models, the small numbers of RRs per disease type, as well as the large numbers of missing values expected when using both case-control and cohort studies, made regression analysis inappropriate. Instead, partial correlation analysis was performed for each disease type using three data sets: case-control studies alone, cohort studies alone, and both case-control and cohort studies. At each step, the variable with the strongest correlation was included in the model and its effect apportioned out of the remaining variables. This procedure was continued until all variables were either eliminated because of lack of significant correlation ($P > .05$) or were added to the model.

Results

Table 3 summarizes important characteristics for each disease type. In general, study characteristics were diverse within all four disease categories. Studies varied widely in sample size and duration of follow-up, but few studies were small, by our definition. Information about race, education, socioeconomic status, smoking, and operative

Table 2. Control of Biases: Definitions and Interrater Agreement

Case-Control Studies	Cohort Studies	No. of Studies	Interrater Agreement (%)		κ	Z	P Value
			Complete	Near			
Predetermined selection of cases and controls The method for including subjects should be clearly stated and determined before the study is performed.	——	30	56.7	96.7	0.384	3.24	.0006
Definition of oral contraceptive exposure Exposure should be defined precisely before the study is performed.	——	30	60	96.7	0.216	1.21	.1131
Unbiased data collection Interviewers who collect the data should be blinded both to the study hypothesis and to the patients' classification as cases or controls.	——	30	80	96.7	0.654	4.19	0.000013
Equivalent patient recall Because exposure data based on patient interviews are subject to bias from patients' memories, physician's records or other confirmation of exposure should be used.	——	30	90	100	0.781	3.92	.000048
Exclusions unlikely to create bias Exclusions should be equally applied to cases and controls and should not constrain the sample to unrepresentative subjects.	——	30	63.3	96.7	0.412	2.89	.0019
Equal diagnostic examination In order to exclude "silent" cardiovascular events, controls should undergo the same diagnostic procedures and examination as the cases.	Equal diagnostic examination Because the cardiovascular events under study may be "silent" or misdiagnosed, exposed and unexposed women should undergo the same diagnostic procedures and examinations.	47	93.6	93.6	0.793	3.63	.00015
Equal medical surveillance Exposed and unexposed women in the population should undergo similar medical observation so that they are equally likely to be detected as cases.	Equal medical surveillance Exposed and unexposed women in the cohorts should undergo similar medical observation so that they are equally likely to be detected as having the cardiovascular event in question.	47	97.9	97.9	0.850	2.36	.0091
Equal demographic susceptibility Cases and controls should be comparable with regard to age, parity, marital status, and socioeconomic status; adjustments should be made for noted differences.	Equal demographic susceptibility Exposed and unexposed cohorts should be comparable with regard to age, parity, marital status, and socioeconomic status; adjustments should be made for noted differences.	47	74.5	97.9	0.542	4.16	.000013

Table 2. (continued)

Case-Control Studies	Cohort Studies	No. of Studies	Interrater Agreement (%) Complete	Near	κ	Z	P Value
Equal clinical susceptibility Cases and controls should be comparable with regard to clinical factors known to affect the risk of developing the cardiovascular disease in question; appropriate adjustments should be made for noted differences.	Equal clinical susceptibility Exposed and unexposed groups should be comparable with regard to clinical factors known to affect the risk of developing the cardiovascular disease in question; appropriate adjustments should be made for noted differences.	47	78.7	89.4	0.359	1.74	.0409
Avoidance of Berkson's (hospitalization) bias Both cases and controls should be derived from a community source rather than from hospitalized women.	———	30	96.7	96.7	0.925	4.64	.000002
———	Adherence monitored Systematic follow-up of the cohorts should monitor any changes in oral contraceptive use.	17	76.5	88.2	0.577	2.67	.0038
———	Dropouts analyzed Dropout rates of exposed and unexposed women should be low; if loss to follow-up is high, exposed and unexposed dropouts should have similar risks for the target disease.	17	88.2	94.1	0.744	2.83	.0023
———	Population representative The population from which the subjects are derived should be comparable to the population of US women with regard to susceptibility to cardiovascular disease.	17	82.4	100	0.301	0.72	.2358
———	Inception cohort Subjects should be included in the study from the beginning of their exposure to oral contraceptives.	17	100	100	1.0	1.46	.0721

status was often not specified. The disease subgroups included in the venous thromboembolism and stroke studies were varied; those of myocardial infarction and cardiovascular death were more uniform. The method of diagnosis was most commonly mixed, ie, various combinations of clinical and laboratory criteria, or unspecified. Seldom was confirmation of the diagnosis by radiologic or nuclear imaging required. Study subjects often included women with known predispositions to cardiovascular disease, and some studies gave no information about predisposing factors. Little information was available about the estrogen and progestin contents of the oral contraceptives taken by the study participants.

As Table 3 demonstrates, the results of the assessment of control of biases are disappointing. The criteria designed to prevent bias in the detection of cardiovascular events (equal diagnostic examination and equal medical surveillance) were seldom met. No case-control study adequately defined exposure, nor did any cohort study assure representativeness of the subjects.

Figure 1 summarizes the overall RRs for the outcome data, both uncorrected and corrected for differences in study standard errors. The weighted RR for venous thromboembolism is 2.8 (CI = 2.4 to 3.2), for cerebrovascular accident, 1.8 (CI = 1.6 to 2.0), for myocardial infarction, 1.6 (CI = 1.4 to 1.8), and for

Table 3. Summary of Study Characteristics for Each Disease Type

Study Characteristics	Venous Thromboembolism (n = 24)	Stroke (n = 20)	Myocardial Infarction (n = 24)	Cardiovascular Mortality (n = 8)
Total number of relative risks	80	95	91	29
Study type (% cohort)	58	40	38	88
Study size (% small study)	8	23	19	0
Adjustment of relative risk (% matched)	38	40	58	75
Fatality (% fatal cases only)	4	20	21	100
Matching of controls (% matched)	39	40	46	25
Socioeconomic status				
Lower class (%)	8	10	8	25
Middle class (%)	17	20	17	38
Geographic location (% based in the United States)	42	35	51	51
Practice site (% inpatient)	50	40	33	0
Dropout/exclusion rate ($\bar{x}$ %)	49	49	46	18
Oral contraceptive use (% current users)	88	75	83	50
Predisposition to CV disease (% nonpredisposed)	33	30	21	0
Minimum age, y ($\bar{x}$ + SD)	17.3 + 3.6	18.5 + 6.0	21.4 + 7.1	19 + 5.4
Maximum age, y ($\bar{x}$ + SD)	46.3 + 5.2	47.1 + 6.5	49.3 + 7.3	52.4 + 9.2
Statistical power, % ($\bar{x}$ + SD)	74.6 + 25.3	59.3 + 34.5	70.7 + 27.8	99.7 + 0.8
Method of diagnosis (% of studies)				
Clinical	4	5	4	13
Radiologic	13	—	4	—
Mixed	38	65	54	63
Unspecified	46	25	33	25
Other	—	5	4	—
Journal type (% general medical journal)	42	50	46	50
Publication site (% American journals)	39	35	33	25
Control of bias (% standard met)				
Predetermined selection	10	33	20	0
Defined exposure	0	0	0	0
Unbiased collection	10	0	13	0
Equivalent recall	10	25	33	0
Exclusions unlikely	70	58	40	0
Equal examination	8	15	13	0
Equal surveillance	13	0	0	0
Equal demographics	33	25	17	63
Equal clinical susceptibility	13	5	17	13
Berkson's bias avoided	20	42	40	100
Adherence monitored	29	25	22	14
Dropouts analyzed	21	0	0	0
Representativeness	0	0	0	0
Inception cohort	7	13	11	14

cardiovascular mortality, 1.0 (CI = 0.8 to 1.3). Except for cardiovascular mortality, the overall risks were significantly greater than 1.0 for each disease type. Summarizing P values for the key RRs using Stouffer's Z technique,[33] we found significant overall P values for venous thromboembolism ($Z = -12.33$, $P < .0001$), stroke ($Z = -7.93$, $P < .0001$), and myocardial infarction ($Z = -7.44$, $P < .0001$), but not for cardiovascular mortality ($Z = 0.72$, $P = .47$).

Table 4 presents the partial correlation analyses for each disease type for variables in case-control studies only, cohort studies only, and all studies. These analyses demonstrate the variables that independently contribute to the RR. In general, there were more significant differences and correlations with venous thromboembolism than with the other disease types. Variables showing a significant relationship in one disease type often failed to demonstrate a relationship in the other disease types. Failure to meet standards for control of bias was significantly related to the RR in some cases. Of note is that, although almost no study adequately defined exposure or analyzed the effect of dropouts, failure to control these potential biases showed significant relationships with RR for certain disease types.

For venous thromboembolism, 10 study characteristics were important. Cohort studies had lower RRs than did case-control studies. Statistical power and percentage of white subjects were inversely related to RR for all studies, while studying unspecified or all socioeconomic classes predicted higher RR. Current use of oral contraceptives, lower maximum age of subjects, and being a small study were related to the magnitude of the

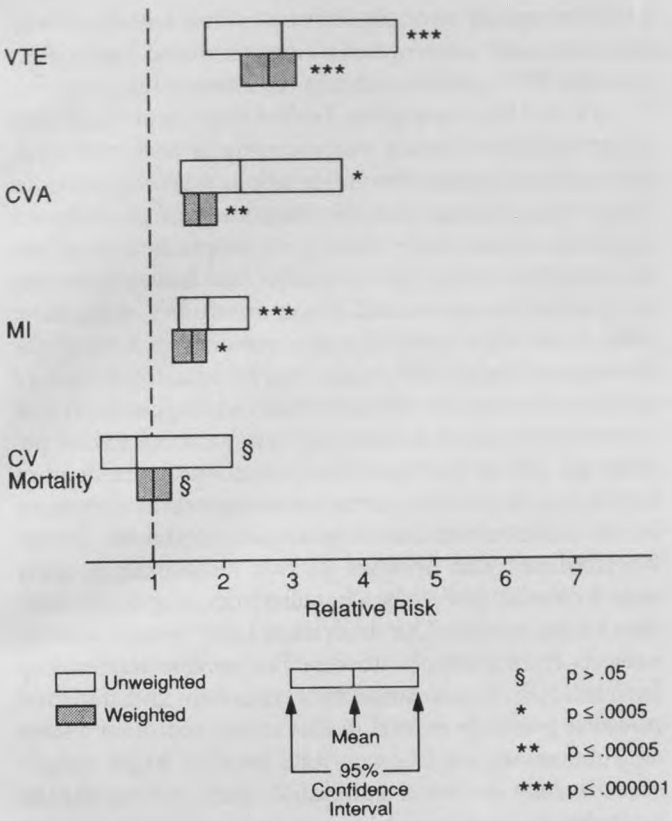

Figure 1. Summary of relative risks for venous thromboembolism (VTE), cerebrovascular accident (CVA), myocardial infarction (MI), and cardiovascular (CV) mortality.

Table 4. Results of Partial Correlational Analysis*

Variable	All Studies	Case-Control Studies	Cohort Studies
Venous thromboembolism			
Cohort study type	−0.39†	—	—
Small study		0.51‡	
Socioeconomic status	0.55†		
Geography (China-Others-Puerto Rico-Australia)			0.34§
Race (% white)	−0.43‡		−0.43‡
OC use (past, ever, current)		0.53†	
Maximum age		−0.42‡	
Statistical power	−0.63†		
Duration of follow-up of nonexposed			0.70†
Adherence monitored			0.54†
Stroke			
Defined exposure		−0.49†	—
Berkson's bias avoided		−0.31§	—
Myocardial infarction			
Dropout/exclusion rate	0.31‖	0.34‡	
Berkson's bias avoided		−0.28§	—
Cardiovascular mortality			
Dropout/exclusion rate	0.59‖		
Equal demographics			0.47§

**Initial Pearson correlation coefficient presented.*
†P ≤ .001; ‡P ≤ .01; §P≤ .05; ‖P ≤ .005.

RR in case-control studies of venous thromboembolism. Among the cohort studies of venous thromboembolism, RR was inversely related to the percentage of white subjects, and directly related to the duration of follow-up of nonexposed women, geographic location, and adequate monitoring of adherence.

Few variables independently contributed to the RR for disease types other than venous thromboembolism. The RR of stroke was inversely related to the definition of exposure and avoidance of Berkson's (hospitalization) bias. In myocardial infarction studies, RR depended on the dropout or exclusion rate. In addition, avoidance of Berkson's bias tended to lower the RR. In studies of cardiovascular mortality, RR was related to dropout or exclusion rate in all studies and ensured equal demographic susceptibility in cohort subjects.

Concern arises in the assessment of RR in venous thromboembolism because monitoring of adherence is important, yet adequate monitoring was done in only 29% of studies. Similarly, in stroke studies, definition of exposure appears to be important, but that standard was not met in any study.

Because few studies specified the estrogen and progestin content of the oral contraceptives used, this analysis was unable to adequately examine the relationships between the type of oral contraceptives taken and the incidence of cardiovascular disease. Analysis of the data available on the relationship between progestin content and RR failed to identify any significant pattern. Only for venous thromboembolism was estrogen content associated with RR; however, there was no definitive dose-related pattern to this relationship.

Discussion

The meta-analysis described in this paper documents agreement within the literature that significant relationships exist between oral contraceptive use and venous thromboembolism, stroke, and myocardial infarction. This meta-analysis also shows that, according to the literature, no relationship has been found between oral contraceptive use and overall cardiovascular mortality, despite good statistical power.

The problem with accepting the validity of the relationships observed between contraceptive use and venous thromboembolism, stroke, and myocardial infarction is that serious methodologic flaws existed within most of the studies. The method of diagnosis was rarely definitive. Although the diagnosis of myocardial infarc-

tion is adequately made using a combination of clinical, electrocardiographic, and laboratory (enzyme) data, the accurate diagnosis of venous thromboembolism, stroke, and death due to a cardiovascular cause often require radiologic or nuclear imaging. For venous thromboembolism, in particular, the sensitivity and specificity of accurate diagnosis by clinical examination are exceedingly poor.

Important standards for preventing bias were rarely met. Specifically, studies were almost always subject to bias in the detection of cardiovascular disease, ie, they failed to ensure equal diagnostic examination and equal medical surveillance. None of the case-control studies adequately defined exposure, and none of the cohort studies adequately demonstrated the representativeness of the population used. Most of the other standards for prevention of bias were seldom met as well.

If the associations found in the literature between oral contraceptive use and the various cardiovascular diseases were due to bias, then one would expect control of potentially important biases to be negatively correlated with the magnitude of the RR. We found, however, only a few significant relationships between susceptibility to bias and RR. We also anticipated that more definitive methods of diagnosis would correlate with lower levels of RR if bias was the reason for the associations; however, the method of diagnosis was not significantly related to RR for any disease type. The lack of significant relationships in our meta-analysis between control of important biases and level of RR may thus be interpreted as supporting the validity of the epidemiologic literature. On the other hand, the lack of a relationship may reflect the rarity with which the bias-control standards were met. So seldom was equal medical surveillance assured, for example, that detection bias remained a threat to nearly all studies. A meta-analysis could not be expected to demonstrate significant relationships if bias occurred in all, or nearly all, of the studies.

If a true relationship between oral contraceptive use and venous thromboembolism exists, then good control of biases should produce higher effect sizes. In fact, cohort studies that were methodologically stronger with regard to duration of follow-up and monitoring of adherence had higher RRs. In addition, because oral contraceptive use is suspected of increasing venous thromboembolic risk only during current or very recent use, our finding that studies of current users produced higher RRs supports the validity of the literature.

Two of the findings of the partial correlational analysis could be interpreted as refuting the relationship between oral contraceptive use and venous thromboembolism. That the cohort study type is inversely related to RR raises questions about the validity of the observed association because, in general, cohort studies are considered methodologically stronger than case-control studies. Similarly, that small case-control studies produced higher RRs could be interpreted as refuting the relationship.

Five of the findings in Table 4 were neutral, neither supporting nor refuting a relationship of oral contraceptive use with venous thromboembolism. We consider as neutral the findings that the magnitude of the RR was higher in studies with subjects of unspecified or all socioeconomic classes and in studies conducted in certain geographic locations, and lower in studies with more subjects of white race. We also consider as neutral the finding of higher RR with higher maximum age of subjects because the risk of thromboembolism with oral contraceptive use has generally not been related to patient age. These findings do not mean that individuals of higher age, white race, certain socioeconomic classes, or certain nationalities have higher risk of venous thromboembolism. The findings of our meta-analysis apply only to *studies* with these characteristics, not to the subjects of the studies. Our analysis did not include pooling subjects from multiple studies. The inverse relationship between RR of venous thromboembolism and statistical power is probably neutral in that it may represent a form of publication bias. Lower RRs require larger sample sizes to reach statistical significance that, in turn, may be a condition for publication.

Thus, this meta-analysis has mixed findings concerning the validity of the epidemiologic studies of oral contraceptives and venous thromboembolism. Three findings support the validity of the association, two tend to refute it, and five are neutral. While the observed association may still be due to systematic bias, eg, in the detection of venous thromboembolic events, our meta-analysis supports, at least in part, the validity of the association with oral contraceptive use.

The validity of the association between oral contraceptive use and stroke is challenged by our findings. Partial correlation analysis suggests that the relationship observed in the literature may be secondary to poorly controlled bias. Methodologic strength in definition of exposure and avoidance of Berkson's bias are both inversely related to RR in the case-control studies. Since none of the stroke studies adequately defined exposure, and only 42% controlled for Berkson's bias, the conclusion that oral contraceptive use is associated with stroke is questionable.

The overall RR for myocardial infarction is less than that for stroke. Again, partial correlation analysis raises doubt concerning the validity of this relationship. The relative risk of myocardial infarction is negatively correlated with control of Berkson's bias and positively correlated with the study's dropout or exclusion rate. With only 40% of the studies having been controlled for

Berkson's bias and a mean dropout or exclusion rate of 46%, the validity of association between oral contraceptive use and myocardial infarction can be challenged.

Finally, the lack of a significant overall relationship between oral contraceptive use and cardiovascular mortality is probably valid. Not only is statistical power excellent, but the two factors identified through partial correlation analysis as important, dropout or exclusion rate and equal demographic susceptibility, were generally addressed in the literature. The mean dropout or exclusion rate for cardiovascular mortality studies was 18%, and 63% of studies adequately ensured equal demographic susceptibility. Hence, the lack of an association between oral contraceptive use and cardiovascular mortality is probably valid. Because our study did not focus on any particular age or risk groups, our findings apply to overall risk of cardiovascular death. The question of oral contraceptive safety in older women, raised by the 1977 mortality report of the Royal College of General Practitioners,[34] and later narrowed to older women who smoke,[35] is thus not addressed by our study.

Many of the studies included in our meta-analysis were from eras when oral contraceptives contained higher doses of both estrogen and progestin than are currently prescribed. While we found no relationship between the year of study publication and magnitude of RR, we had little direct information about dosage. Thus, it is unclear whether associations noted in older studies in which patients had been taking oral contraceptives containing larger amounts of hormones will remain as prescribing habits shift to lower-dose oral contraceptives.

This meta-analysis is unusual in several respects. It is unique in its attempt to quantify the impact of susceptibility to biases that may affect epidemiologic studies. In addition, our meta-analysis attempts to address issues often ignored in previous meta-analyses. Sacks et al[9] reported that only 7% of the meta-analyses reviewed provided clear evidence of a predetermined protocol with inclusion criteria. Only 5% assessed interobserver agreement, and only 2% addressed publication bias. Of the 23 features Sacks et al used to assess the quality of a meta-analysis, none of the studies that they reviewed had adequately addressed more than 14. We believe our meta-analysis addressed 18 of those features.

The limitations of this study are the same as those that generally apply to all meta-analyses. Information concerning the studies and their sample characteristics was often incomplete. Our inability to identify any unpublished studies despite an apparently adequate search may be a cause for concern, although fail-safe Ns were computed, the results from which were reassuring. As with any meta-analysis, the paucity of studies for each disease type places this study at risk for type II errors. The small number of raters and the lack of adequate interrater agreement for a few bias-control standards, though they were not related to RR, may raise questions about the coding in general. Finally, the application of meta-analysis to epidemiologic issues is still somewhat controversial.

In conclusion, this meta-analysis suggests that the relationship between oral contraceptive use and venous thromboembolism may be valid. Overall cardiovascular mortality is not increased by oral contraceptive use. Finally, although the literature has found a significant relationship between oral contraceptive use and both myocardial infarction and stroke, these relationships may be due to methodologic flaws.

References

1. Stadel BV. Oral contraceptives and cardiovascular disease. N Engl J Med 1981; 305:612–8, 672–7.
2. Vessey MP. Female hormones and vascular disease—an epidemiological overview. Br J Fam Plann 1980; 6 (suppl):1–12.
3. Sartwell PE, Stolley PD. Oral contraceptives and vascular disease. Epidemiol Rev 1982; 4:95–109.
4. Meade TW. Risks and mechanisms of cardiovascular events in users of oral contraceptives. Am J Obstet Gynecol 1988; 158: 1646–52.
5. Realini JP, Goldzieher JW. Oral contraceptives and cardiovascular disease. Am J Obstet Gynecol 1985; 152:729–98.
6. Katerndahl DA, Cohen PA. Quantitatively reviewing the literature. Fam Pract Res J 1987; 6:123–9.
7. Strube MJ, Hartmann DP. Meta-analysis. J Consult Clin Psychol 1983; 51:14–27.
8. Rosenthal R, Rubin DB. Meta-analytic procedures for combining studies with multiple effect sizes. Psychol Bull 1986; 99:400–6.
9. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analysis of randomized controlled trials. N Engl J Med 1987; 316:450–5.
10. Kulik JA, Kulik CLC. Meta-analysis in education. Int J Educ Res 1989; 13:221–40.
11. Hunter JE, Schmidt FL. Methods of meta-analysis: correcting error and bias in research findings. Newbury Park, Calif: Sage, 1990.
12. Abrami PC, Cohen PA, d'Apollonia S. Implementation problems in meta-analysis. Rev Educ Res 1988; 58:151–79.
13. Mann C. Meta-analysis in the breech. Science 1990; 249:476–80.
14. Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. Am J Med 1979; 66:556–64.
15. Feinstein AR, Horwitz RI. Double standards, scientific methods, and epidemiologic research. N Engl J Med 1982; 307:1611–7.
16. Feinstein AR. Clinical biostatistics X. Sources of "transition bias" in cohort statistics. Clin Pharmacol Ther 1971; 12:704–21.
17. Feinstein AR. Clinical biostatistics XI. Sources of "chronology bias" in cohort statistics. Clin Pharmacol Ther 1971; 12:864–79.
18. Feinstein AR. Clinical biostatistics XLVIII. Efficacy of different research structures in preventing bias in the analysis of causation. Clin Pharmacol Ther 1979; 26:129–41.
19. Garfield E. Which medical journals have the greatest impact? Ann Intern Med 1986; 105:313–20.
20. Realini JP. Evaluation of case-control studies. In: Selected proceedings of the 1986 and 1987 Primary Care Research and Statistics Conferences. San Antonio: North American Primary Care Research Group, 1989: 1–8.
21. Realini JP. Evaluation of cohort studies. In: Selected proceedings