American Journal of Epidemiology

Vol. 185, No. 10
DOI: 10.1093/aje/kww193
Advance Access publication:
April 21, 2017

Practice of Epidemiology

Exposure Opportunity: The Advantages of Including Men in Analyses of Female-Related Risk Factors

Kirsten Rozemeijer, Saskia le Cessie, Astrid van Hylckama Vlieg, Frits R. Rosendaal, Jan P. Vandenbroucke, Charles Poole, and Suzanne C. Cannegieter*

* Correspondence to Dr. Suzanne C. Cannegieter, Department of Clinical Epidemiology, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands (e-mail: s.c.cannegieter@lumc.nl).

Initially submitted September 16, 2015; accepted for publication April 6, 2016.

Intuitively, researchers do not include subjects who do not have the opportunity to be exposed, such as men in studies on oral contraceptives (OCs). We aimed to explore in which situations it is nevertheless beneficial to do so. We considered the effect of including men in case-control analyses of 8 different hypothetical data sets on the effect of OC use and venous thrombosis. In all scenarios, OC use was the exposure of interest, sex the factor that determined exposure opportunity, and air travel another risk factor. In some of these scenarios, sex and air travel were included as confounders or effect modifiers. Logistic regression was used to estimate odds ratios. Standard errors of the estimated log odds ratios, including and excluding men, were compared. We also studied the effect of including men using data from 1999–2004 from a case-control study on risk factors for venous thrombosis, conducted in the Netherlands. In all hypothetical examples, and in the real-data study, addition of men to the analysis yielded the same odds ratios when correctly adjusting for confounding. Moreover, use of additional subjects often led to more precise estimates. We suggest that subjects who do not have the opportunity to be exposed should not routinely be excluded from epidemiologic studies.

case-control studies; epidemiologic methods; exclusion; exposure opportunity; inclusion

Abbreviations: FVL, factor V Leiden; MEGA, Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis; OC, oral contraceptive; VT, venous thrombosis.

In a case-control study designed to estimate an incidence rate ratio, the incident cases are included along with a sample of the source population selected to represent the distribution of the exposure in the source population's person-time at risk. The exposed-to-unexposed ratio of the odds of being a case in the study (odds ratio) then estimates the exposed-to-unexposed ratio of the incidence rates. Over the years there have been discussions about whether individuals who could not have the exposure of interest should be included in case control studies (1–4). About 30 years ago, it was postulated that including subjects who do not have the opportunity to be exposed could introduce bias and dilute the odds ratio towards unity (1). Indeed, researchers sometimes do not include subjects in a study who do not have the opportunity to be exposed, such as men in a study on oral contraceptive (OC) use and an outcome that both men and women can experience or women in a study on prostate cancer and its effects on survival. However, Poole (2) demonstrated in 1986 that the inclusion of infertile women in case-control studies on OC use and myocardial infarction did not result in a dilution of the effect estimate, as long as infertility did not negatively confound the association. Moreover, the precision of the effect measure estimate improved by including infertile women. He concluded that subjects who have reasons for nonexposure do not necessarily need to be excluded from a study or analysis and that, in many situations, including them will have advantages (2, 4). However, his conclusions were restricted to reasons for nonexposure that were unrelated to the outcome in the study (i.e., that were not confounding factors).

Our interest in the benefit of including men in a study on female-related risk factors arose when we came across a problem in an analysis of the Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis (MEGA) study (5). In this case-control study on risk factors

for venous thrombosis (VT), we planned to study the effect of air travel and OC use on the incidence of VT. The controls had been sought among partners of the patients, so controls were matched on a 1-to-1 basis to the cases and were generally of the opposite sex. We assumed that in this design a matched analysis on the combined effect of air travel and OC use was not possible. We therefore performed a case-only analysis including only female cases, which showed a highly elevated risk for the combination of air travel and exposure to OCs. However, in a commentary that accompanied our paper, Kenneth Rothman pointed out that a case-control analysis with men included could in principle have been performed (6). Moreover, a matched analysis including men would have been superior, because air travel and OC use are not independent in the population, which is a prerequisite for a case-only analysis. In our subsequent response, we showed that a case-control analysis of the MEGA data led to more realistic results, in line with data from other studies (7, 8). Hence in the MEGA study, including men seemed to be the only method to obtain correct estimates for the combination of air travel and OC use on the risk of thrombosis.

The aim of the current paper was to describe several situations addressing whether and under what assumptions individuals who do not have the opportunity to be exposed can be included in an analysis. First, we examined the effect of OC use on the risk of VT in a case-control study, both when including and excluding men from the analysis, using hypothetical examples based on different scenarios with different assumptions. These scenarios include some in which the reason for nonexposure (i.e., sex) is a confounder, some in which air travel is added as risk factor and confounder, and some in which statistical interaction is present between different risk factors. In addition to these scenarios, we illustrate the effect of including and excluding men using real data from the MEGA case-control study.

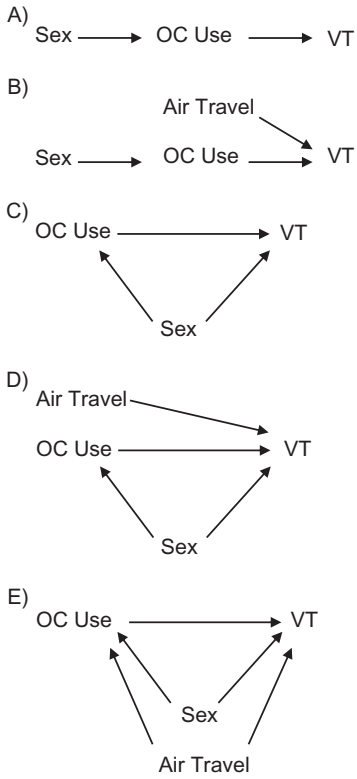

Figure 1. Directed acyclic graphics (DAGs) for the different scenarios considered. A) corresponds with scenario 1; B) corresponds with scenarios 2 and 3; C) corresponds with scenario 4; D) corresponds with scenarios 5 and 6; E) corresponds with scenarios 7 and 8. Oral contraceptive (OC) use was in all scenarios as the exposure of interest, venous thrombosis (VT) was the outcome, and sex was the reason for nonexposure. Air travel was an additional risk factor in (B) and (D). Sex was, in addition to the reason for nonexposure, also a confounder in (C), (D), and (E). Air travel was a confounder in (E).

METHODS AND RESULTS OF HYPOTHETICAL EXAMPLES

We considered 8 different hypothetical scenarios on the effect of OC use and the risk of VT in an unmatched case-control study. As an additional risk factor for VT, air travel is introduced. Directed acyclic graphs for each scenario are shown in Figure 1. In all situations, female nonusers of OCs who did not travel by air are used as reference group. OC use was the exposure of interest and sex the factor that determined exposure opportunity (i.e., no men use OCs, while women can). This is depicted in Figure 1A, which shows the relationship between sex, OC use, and VT in a hypothetical situation without additional risk factors or confounders. In scenario 2, air travel was a second risk factor for VT (Figure 1B). Scenario 3 was similar to scenario 2, except that there was interaction between OC use and air travel on a multiplicative scale (Figure 1B). In these 3 scenarios the causal assumption holds as there is no confounding and it is assumed that unexposed women and men have the same probability of developing VT. Scenarios 4–6 were identical to scenarios 1–3 except that we let sex confound the relationship between OC use and VT (Figure 1C and 1D). Thus, we assumed that men have a different risk of thrombosis than women do, while sex was also associated with the exposure of interest. In scenarios 7 and 8, both sex and air travel confounded the relationship between OC use and VT (Figure 1E). Supramultiplicative interaction between sex and air travel is present in scenario 8.

We constructed 8 hypothetical data sets corresponding to the 8 scenarios (Table 1). For each of the data examples, the odds ratio of VT for OC use was set to 2 and the odds ratio for air travel to 3. In the scenarios in which sex was a confounder of the relationship between OC use and VT, the odds ratio of VT for a male versus female was set to 2.5. When we wished interaction between OC use and air travel to be present, the ratio of odds ratios corresponding to supramultiplicative interaction was set to 4. This implies that, because the odds ratios for OC use and air travel were 2 and 3, respectively, the odds ratio for VT when both OC use and air travel are present would be 24 ($2 \times 3 \times 4$). In scenario 8, the product of the separate odds ratios for sex and air travel was multiplied by a factor of 4.5.

For each scenario, a multivariate unconditional logistic regression analysis was performed to estimate the odds ratios

Table 1. Number of Subjects Included in 8 Hypothetical Data Examples of Case-Control Studies on the Association of Oral Contraceptive Use With Venous Thrombosis

Scenario Description					Men		Women	
Risk Factor	Interaction	Confounder	OC Use	Air Travel	No. of Cases	No. of Controls	No. of Cases	No. of Controls
		Scenario 1[a,b]						
OC use			Yes	No	0	0	80	360
			No	No	80	720	40	360
		Scenario 2[a,c]						
OC use, air travel			Yes	Yes	0	0	160	240
			Yes	No	0	0	80	360
			No	Yes	160	480	80	240
			No	No	80	720	40	360
		Scenario 3[a,c]						
OC use, air travel	OC use × air travel		Yes	Yes	0	0	640	240
			Yes	No	0	0	80	360
			No	Yes	160	480	80	240
			No	No	80	720	40	360
		Scenario 4[a,d]						
OC use		Sex	Yes	No	0	0	80	360
			No	No	200	720	40	360
		Scenario 5[a,e]						
OC use, air travel		Sex	Yes	Yes	0	0	160	240
			Yes	No	0	0	80	360
			No	Yes	400	480	80	240
			No	No	200	720	40	360
		Scenario 6[a,e]						
OC use, air travel	OC use × air travel	Sex	Yes	Yes	0	0	640	240
			Yes	No	0	0	80	360
			No	Yes	400	480	80	240
			No	No	200	720	40	360
		Scenario 7[a,f]						
OC use		Air travel, sex	Yes	Yes	0	0	220	330
			Yes	No	0	0	60	270
			No	Yes	400	480	80	240
			No	No	200	720	40	360
		Scenario 8[a,f]						
OC use	Air travel × sex	Air travel, sex	Yes	Yes	0	0	220	330
			Yes	No	0	0	60	270
			No	Yes	1,800	480	80	240
			No	No	200	720	40	360

Abbreviations: DAG, directed acyclic graph; OC, oral contraceptive; VT, venous thrombosis.

[a] OC use was the exposure of interest, VT the outcome, and sex the reason for nonexposure. If present, effect modification was supramultiplicative.

[b] Scenario 1 corresponds with the DAG in Figure 1A.

[c] Scenarios 2 and 3 correspond with the DAG in Figure 1B.

[d] Scenario 4 corresponds with the DAG in Figure 1C.

[e] Scenarios 5 and 6 correspond with the DAG in Figure 1D.

[f] Scenarios 7 and 8 correspond with the DAG in Figure 1E.

for the risk factors. This was performed twice, once including men in the analysis and once excluding them. We used the standard error of the log odds ratio as a measure for precision and the relative inefficiency as a measure for the loss in precision when men were excluded from the analysis rather than being included. The relative inefficiency was defined as the ratio of the standard error of the log odds ratio when men were excluded from the analysis and when men were included. For example, a relative inefficiency of 1.5 represents a 1.5-fold increase in the standard error of the log odds ratio when men are excluded from the analysis.

RESULTS

The results of the analyses of the 8 scenarios are given in Table 2, with and without the inclusion of men, who make up half of the source population and, therefore, half of the controls. In scenario 1, sex (i.e., the determinant of nonexposure) is not associated with VT (i.e., is not a confounder). The odds ratio for OC use is estimated correctly, both with and without inclusion of men. However, the estimate becomes more precise when men are included. The relative inefficiency of not including men is 1.32.

In scenario 2, air travel is an additional risk factor for VT. In the logistic regression analysis with OC use and air travel as covariates, the addition of men to the analysis yields odds ratios identical to those in the women-only analysis. Again, the estimate is more precise when men are included in the analysis. The relative inefficiencies of excluding men for OC use and air travel are 1.31 and 1.32, respectively.

In scenario 3, apart from OC use and air travel being risk factors for VT, statistical interaction is also present between air travel and OC use. The odds ratios for OC use and air travel and the ratio of odds ratios for the interaction between OC use and air travel are, again, unaffected by the inclusion of men in the analysis. The relative inefficiencies of excluding men for OC use, air travel and for their statistical interaction are 1.32, 1.73 and 1.35, respectively.

In scenarios 4–8, sex is a confounder in the association between OC use and VT. Scenarios 4–6 are identical to scenarios 1–3 apart from the confounding by sex. In scenario 4, where only OC use is a risk factor for VT, it is shown that ignoring sex when men are included in the analysis yields a biased estimate of the odds ratio for OC use. Therefore, in scenarios 4–6, sex was added as a covariate in the logistic regression. Adjustment for sex results in obtaining the correct odds ratio, but no efficiency is gained by including the men.

When air travel is present as second risk factor for VT (scenario 5), the addition of men to the analysis does not improve the precision of the odds ratio for OC use. However, the precision for air travel increases, and both analyses yield the correct odds ratios.

In scenario 6, statistical interaction is present between air travel and OC use. Here all estimates are correct, and they are more precise when men are included in the analysis.

Air travel and sex are both confounders in the association between OC use and the risk of VT in scenario 7. Including men yields the correct estimates of the odds ratios provided that sex is included in the model. As in scenario 5, there is no gain in precision of the odds ratio for OC use, but the precision of the odds ratio for air travel is increased.

In scenario 8, statistical interaction is present between traveling by air and sex, which are also both confounders. A term for interaction between air travel and sex should be added to the model in this scenario, if men are included, in order to obtain correct parameter estimates. No gain in precision is obtained.

The presence of submultiplicative instead of supramultiplicative interaction (scenarios 3, 6, and 8) and/or downward instead of upward confounding (scenarios 4–8) results in similar findings as described above (Web Tables 1 and 2). For example, if sex is a downward confounder and air travel a second risk factor (scenario 5), the inclusion of men in the analysis does not affect the odds ratios for OC use and air travel. In addition, the precision of the odds ratio for air travel improves while the precision of the odds ratio for OC use is not affected.

FACTORS INFLUENCING THE RELATIVE INEFFICIENCIES

The relative inefficiencies depend on the male-to-female ratio and on the fraction of women using OCs. This can intuitively be understood: Including men will increase the size of the group of nonusers of OCs. When there are more men, more precision can be gained by including them in the analysis. Furthermore, because all men are nonusers of OCs by definition, more efficiency will be gained if there are relatively few nonusers of OCs among the women. To illustrate the magnitude of the effect of changing the male-to-female ratio, we considered scenario 2 and varied the male-to-female ratio while keeping the rest of the design fixed (i.e., no change in the total number of subjects, the percentage of subjects who travelled by air, and the percentage who used OC) using numbers in Table 1. We calculated the relative inefficiencies for different ratios using the asymptotic standard errors (see Web Appendix 1 for details). Results are shown in Figure 2. A change in the magnitude of the odds ratios can also influence the relative efficiency. However, no overall conclusion can be given here. Whether the inefficiency increases or decreases with increasing odds ratio depends on factors such as the ratio of cases and controls, the fraction of women who are using OCs, and the male-to-female ratio, as is illustrated in Web Appendix 1.

MEGA STUDY, METHODS, AND RESULTS

To further illustrate the effect of including and excluding men, we performed analyses on data from the MEGA study, a case-control study on risk factors for VT in the Netherlands (5) similar to those discussed in the previous sections. Although the MEGA study was designed as a matched study (partners of the controls were selected as cases), we performed an unmatched analysis, because we wanted to compare the same analyses, once with men included and once with men excluded (which is not possible in a matched analysis). Because the controls must be representative for the source population (in this case, the general population at risk for VT)

Table 2. Results of Multivariate Logistic Regression Analysis of the 8 Data Examples of Oral Contraceptive Use and Venous Thrombosis With and Without the Inclusion of Men in the Analysis

Variable	Including Men in the Analysis[a]		Excluding Men From the Analysis[a]		Relative Inefficiency[a]
	OR	SE of Log OR	OR	SE of Log OR	
Scenario 1[b]					
OC use	2.00	0.157	2.00	0.207	1.32
Scenario 2[b,c]					
OC use	2.00	0.098	2.00	0.129	1.31
Air travel	3.00	0.097	3.00	0.128	1.32
Scenario 3[b–d]					
OC use	2.00	0.157	2.00	0.207	1.32
Air travel	3.00	0.122	3.00	0.211	1.73
Air travel × OC use	4.00	0.189	4.00	0.256	1.35
Scenario 4[b,e,f]					
OC use	1.00	0.143	2.00	0.207	1.45
Scenario 4[b,e,g]					
OC use	2.00	0.207	2.00	0.207	1.00
Sex	2.50	0.185			
Scenario 5[b,c,e]					
OC use	2.00	0.129	2.00	0.129	1.00
Air travel	3.00	0.081	3.00	0.128	1.58
Sex	2.50	0.114			
Scenario 6[b–e]					
OC use	2.00	0.171	2.00	0.207	1.22
Air travel	3.00	0.094	3.00	0.211	2.25
Sex	2.50	0.114			
Air travel × OC use	4.00	0.173	4.00	0.256	1.48
Scenario 7[b,e,h]					
OC use	2.00	0.127	2.00	0.127	1.00
Air travel	3.00	0.082	3.00	0.131	1.60
Sex	2.50	0.114			
Scenario 8[b,e,h,i]					
OC use	2.00	0.127	2.00	0.127	1.00
Air travel	3.00	0.131	3.00	0.131	1.00
Sex	2.50	0.153			
Air travel × sex	4.50	0.162			

Abbreviations: OC, oral contraceptive; OR, odds ratio; SE, standard error; VT, venous thrombosis.

[a] ORs, SE of the log ORs, and relative inefficiencies for different risk factors and confounders are given when men are included and excluded from the analysis.

[b] Female OC nonusers who did not travel by air form the reference group. OC use was the exposure of interest, VT the outcome, and sex was the reason for nonexposure.

[c] Air travel was an additional risk factor.

[d] Interaction between air travel and OC use was present, indicated by ×.

[e] Sex was, besides the reason for nonexposure, also a confounder.

[f] Sex was not added in the model of scenario 4.

[g] Sex was added in the model of scenario 4.

[h] Air travel was a confounder.

[i] Interaction between air travel and sex was present, indicated by ×.

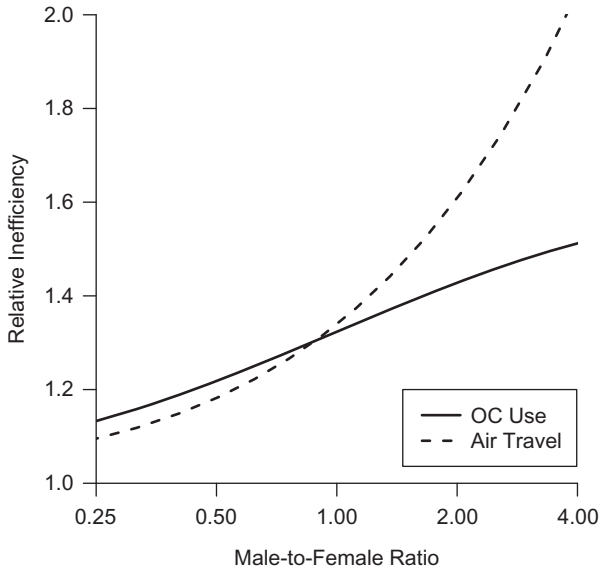

Figure 2. The relationship between the male-to-female ratio and the relative inefficiency in the scenario of a case-control study with 2 risk factors: oral contraceptive (OC) use and air travel (scenario 2). The male-to-female ratio varies while the total number of subjects, the percentage of women using OCs, and the percentage of subjects who travelled by airplane was held fixed, using the values of scenario 2. A relative inefficiency above 1.0 corresponds with a gain in precision when men are included in the analysis, as compared with a women-only analysis.

with respect to exposure and covariates, we chose to study factor V Leiden (FVL) rather than air travel, because we expected the frequency of FVL among the original matched controls to be more similar to that of the general population than the frequency of air travel. FVL is a genetic defect in clotting factor V and therefore its prevalence in the controls is unlikely to be associated with its prevalence in their partners. FVL is associated with a 3- to 5-fold increased risk of VT, and it has been established from previous studies that FVL and OC use display supra-additive interaction (9, 10). We assessed the effect of the exclusion of men from the analysis when studying the relationship between FVL, OC use, and their statistical interaction on the risk of VT, using ordinary logistic regression models.

In this study with 1,804 cases and 1,667 controls, 407 female cases used OCs (22.6% of all cases; 44.8% of female cases) and 122 female controls used OCs (7.3% of all controls; 14.3% of female controls). FVL was present in 276 cases (15.3%) and in 84 controls (5.0%); among women, it was present in 130 cases (14.3%) and 43 controls (5.1%). We considered several scenarios with and without including men. The results are given in Table 3. In scenario 1, the odds ratio for OC use and VT was estimated, and we adjusted for sex. Here, using all subjects yielded the same estimated odds ratio and 95% confidence interval for OC use as in the analysis where the men were excluded. In scenario 2, in which FVL was added to the model, the effect of the second risk factor, FVL, was more precisely estimated with men included in the analysis. The odds ratios for FVL in these analyses were comparable when men were included (odds ratio = 3.30, 95% confidence interval: 2.55, 4.28) and when men were excluded (odds ratio = 2.92, 95% confidence interval: 2.01, 4.24), while the relative inefficiency was 1.45.

A term for interaction between FVL and OC use was added in scenario 3. Again, including men yielded more precise estimates of the effect of FVL and the interaction between FVL and OC use. The change in effect of FVL when including men was larger than the change in effect of OC use. This is because we considered sex as a possible confounder, for which we adjusted in the analysis. This implies that the OC effect is estimated as a contrast (between the women who do and who do not use OC) both when men are included and when they are excluded. In the analyses including men, the effect of FVL is estimated using both men and women (under the assumption in scenario 2 that the effect of FVL is the same for either men or women on the odds ratio scale and under the assumption in scenario 3 that the effect of FVL is the same for men and unexposed women), which yields somewhat different estimates than do analyses using data from female participants only.

Terms for interaction between FVL and OC use and between FVL and sex were added in scenario 4. In this situation, the effect of FVL on thrombosis is estimated separately for OC users and nonusers and for men and women. Therefore, in this analysis the odds ratios of OC use and FVL for the women were exactly equal to the odds ratios obtained when excluding the men. In general the estimates became less precise when the model became more complicated.

DISCUSSION

We have considered several scenarios in a case-control study design in which including subjects who are at risk for the outcome but do not have the opportunity to be exposed yields correct estimates of the odds ratios for exposure. These scenarios include presence of confounding (as long as the correct model is fitted) and statistical interaction between the exposure and other risk factors. We demonstrated that the use of additional, unexposed subjects can lead to a more precise estimate of the exposure effect if the reason for nonexposure does not confound the relationship between exposure and outcome. These results correspond with findings of Poole (2), who demonstrated that including infertile women in case-control studies of the relationship between OC use and myocardial infarction did not result in a dilution of the effect.

Adding men does not increase the precision of the estimate of OC use in a situation in which sex is a confounder, so it is unnecessary to include men if OC use is the only parameter of interest. However, when other factors are also considered, it can be useful to include men, because the effect of other risk factors and confounders can be more efficiently estimated under the model assumption that the effect of these variables is the same for men and women. Furthermore, effects of interaction between exposure and other risk factors can be estimated more precisely, again under the assumption that the effect of the other risk factors is the same for men and women

Table 3. Logistic Regression Analyses Including and Excluding Men From the Analyses, Using Data from the Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis Study, The Netherlands, 1999–2004

Variable	Including Men in the Analysis[a]		Excluding Men From the Analysis[a]		Relative Inefficiency[a]
	OR	SE of Log OR	OR	SE of Log OR	
			Scenario 1[b]		
OC use	4.84	0.118	4.84	0.118	1.00
Sex	1.59	0.076			
			Scenario 2[b,c]		
OC use	4.75	0.120	4.75	0.119	1.00
FVL	3.30	0.132	2.92	0.191	1.45
Sex	1.56	0.077			
			Scenario 3[b–d]		
OC use	4.65	0.123	4.57	0.123	1.01
FVL	3.20	0.138	2.61	0.214	1.55
OC use × FVL	1.44	0.496	1.77	0.522	1.05
Sex	1.56	0.077			
			Scenario 4[b–e]		
OC use	4.57	0.123	4.57	0.123	1.00
FVL	2.61	0.214	2.61	0.214	1.00
OC use × FVL	1.77	0.522	1.77	0.522	1.00
FVL × sex	1.41	0.282			
Sex	1.52	0.080			

Abbreviations: FVL, factor V Leiden; OR, odds ratio; OC, oral contraceptive; SE, standard error; VT, venous thrombosis.

[a] ORs, SE of the log ORs, and relative inefficiencies for different risk factors and confounders are given when men are included and excluded from the analysis.

[b] Female OC nonusers without FLV form the reference group. OC use was the exposure of interest, VT the outcome, and sex was the reason for nonexposure as well as a confounder.

[c] FVL was an additional risk factor.

[d] Interaction between FVL and OC use took place, indicated by ×.

[e] Additional interaction between FVL and sex took place, indicated by ×.

(no interaction between the other risk factors and the reason for nonexposure). Especially when the male-to-female ratio among controls is large, including men in the analysis can improve the relative inefficiency. These findings were also observed in the real-data example of the MEGA study, where including men yielded a more precise estimate of the effect of factor V Leiden on thrombosis (5).

Some issues regarding study design and assumptions related to causal inference need particular attention, first being the positivity assumption. This assumption requires that there be both exposed and unexposed subjects at every combination of confounders. This assumption is valid in the first scenarios, in which sex was not a confounder and men and unexposed women have the same risk of the outcome. However the positivity assumption does not hold if the reason for nonexposure—in our example, sex—is a confounder, because there are no exposed men. We demonstrated that in this situation adjustment for sex was needed. Adding both OC use and sex as independent variables in a logistic model yields a model equivalent to adding a categorical covariate with 3 categories: men, women with OC use, and women without OC use. Hence, even if, technically, the positivity assumption is invalid, we can still estimate the effect for exposed women, unexposed women, and unexposed men. However, no claims about exposed men should be made. The second assumption for causal inference concerns exchangeability between exposed and unexposed subjects (i.e., that the unexposed men and women should represent the experience the exposed women would have had if they had not been exposed). Exposure opportunity can be strongly linked to other factors that confound the relationship that is actually being studied, so the extent to which exchangeability is present will depend on the actual research question. For the research question of our study, in which sex determines exposure opportunity, exchangeability will be unlikely, because men and women differ in many respects. For other research questions, this might be different: Consider for example the relationship between UV radiation and melanoma. Here there may also be reasons for nonexposure or minimal exposure opportunity, such as living in northern Scandinavia or adherence to a religion such as Islam in

which women are completely covered when outside (by burka or niqab). Another example could be exposure to pathogen-bearing mosquitoes (malaria), in which exposure opportunity is present only in those parts of a country where the mosquito is prevalent. In these situations people who have and do not have exposure opportunity are probably more exchangeable than in our study. Nevertheless, even if the reason for nonexposure is related to outcome (as sex usually is), adjustment for sex solves this because adding both OC use and sex as variables to the model is equivalent to defining one categorical covariate (men, women with OC use, and women without OC use). As shown in our examples, the effect of OC use is estimated with the same precision whether or not men are included.

With respect to study design, in this paper we focused on case-control studies because we were motivated to conduct this study by problems we ran into in our own case-control study. However, including subjects without exposure opportunity could also be beneficial in other types of designs. If incidence density sampling has been done correctly, the same results should be expected in a cohort study, as has been demonstrated by Poole (2) for the first scenario. For the current analysis, in which we used data from the MEGA study to determine the difference between including and excluding men in "real-life" data, it was necessary to perform an unmatched analysis because we wanted to compare the same analyses with men included and excluded. The expectation was that an unmatched analysis of the relationship between FVL, OC use, and their interaction on the risk of VT would yield results similar to those of a matched analysis. Indeed, a matched analysis including men yielded results very similar to the unmatched analysis with men included (data not shown). Another design issue is the determination of the source population. For the example we chose for the current study most researchers would probably feel that women form the source population, rather than men and women combined. However, if we consider a less extreme example, such as those discussed above, we believe that researchers would define the source population as including people who have no or minimal opportunity to be exposed. So, in a study of UV radiation and melanoma, women who wear a burka would be part of the source population. Likewise, in a study of malaria, people living in areas where malaria is not present would not automatically be excluded. While these examples are in fact not fundamentally different from our example using OC use, they are perceived to be different, probably because we assume implicitly that men will never be exposed to OCs, while in the other 2 examples there is still a chance that those without exposure opportunity may one day be exposed. That the probability of ever being exposed is either zero or very small should not, in principle, make a difference to the decision of including these people in a study. (Indeed, the probability of taking OCs in men is not zero because male-to-female transgender persons take these hormones as well.) So, we propose that everyone in a population, independent of their exposure opportunity, should form the source population but that actual inclusion in a study depends on the practical circumstances present for a particular study.

With respect to these practical circumstances, it should be noted that there is a difference between excluding men in the design of a study and excluding them from the analysis. When the study has been set up as a study in women only, additional collection of data from men is not recommended because this will result in a considerable effort that does not justify the (relatively small) gain in precision. If, in this situation, one considered adding more people to the sample, it would in general be more efficient to add more women (who can be either unexposed or exposed) rather than unexposed men. However, in studies where data from both sexes are available it is advisable, for the scenarios we considered, to include both sexes in all analyses, even when the exposure is not present in one of them.

In conclusion, there are situations in which it is beneficial to include subjects who do not have the opportunity to be exposed, because the use of additional subjects can lead to more precise estimates if the correct model is fitted, especially when the reason for nonexposure is not a confounder for the exposure of interest or for the second risk factor. We therefore suggest that subjects who are at risk for the outcome but do not have the opportunity to be exposed should not routinely be excluded from epidemiologic studies.

ACKNOWLEDGMENTS

Author affiliations: Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands (Kirsten Rozemeijer, Saskia le Cessie, Astrid van Hylckama Vlieg, Frits R. Rosendaal, Jan P. Vandenbroucke, Suzanne C. Cannegieter); Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands (Kirsten Rozemeijer); Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands (Saskia le Cessie); Department of Epidemiology, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC (Charles Poole); and Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, The Netherlands (Suzanne C. Cannegieter).

Conflict of interest: none declared.

REFERENCES

1. Schlesselman JJ. *Case-Control Studies: Design, Conduct, Analysis*. New York, NY: Oxford University Press; 1982: 71–78.
2. Poole C. Exposure opportunity in case-control studies. *Am J Epidemiol*. 1986;123(2):352–358.
3. Schlesselman JJ. Exposure opportunity in epidemiologic studies. *Am J Epidemiol*. 1987;125(2):174–178.
4. Poole C. Critical appraisal of the exposure-potential restriction rule. *Am J Epidemiol*. 1987;125(2):179–183.
5. Cannegieter SC, Doggen CJ, van Houwelingen HC, et al. Travel-related venous thrombosis: results from a large population-based case-control study (MEGA) study. *PLoS Med*. 2006;3(8):e307.
6. Rothman KJ. Thrombosis after travel. *PLoS Med*. 2006;3(8):e300.
7. Cannegieter SC, van Hylckama Vlieg A, le Cessie S, et al. Thrombosis after travel: reply to a Perspective. [Article

comment]. http://www.plosmedicine.org/article/comments/info%3Adoi%2F10.1371%2Fjournal.pmed.0030307. Published March 31, 2009. Accessed March 31, 2009.
8. Kuipers S, Cannegieter SC, Middeldorp S, et al. The absolute risk of venous thrombosis after air travel: a cohort study of 8,755 employees of international organisations. *PLoS Med*. 2007;4(9):e290.
9. Rosendaal FR, Koster T, Vandenbroucke JP, et al. High risk of thrombosis in patients homozygous for factor V Leiden (activated protein C resistance). *Blood*. 1995;85(6):1504–1508.
10. Vandenbroucke JP, Koster T, Briët E, et al. Increased risk of venous thrombosis in oral contraceptive users who are carriers of factor V Leiden mutation. *Lancet*. 1994;344(8935):1453–1457.