Lifestyle and venous thrombosis
Pomp, E.R.

Citation
Pomp, E. R. (2008, December 3). *Lifestyle and venous thrombosis*. Retrieved from https://hdl.handle.net/1887/13308

Version:	Corrected Publisher's Version
License:	[Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](https://hdl.handle.net/1887/13308)
Downloaded from:	https://hdl.handle.net/1887/13308

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 7

The experience of multiple control groups in a large case-control study on gene-environment interaction

Pomp ER, Le Cessie S, Van Stralen KJ, Vandenbroucke JP, Rosendaal FR, Doggen CJM.

ABSTRACT

In a large case-control study on risk factors for venous thrombosis (MEGA study) we enrolled two different control groups; partners of patients and a random digit dialing group (RDD). This presented unexpected challenges in the analysis of three different types of research questions. For the evaluation of body mass index, a general life style factor, partners had to be analyzed with a matched analysis, RDD controls with an unmatched analysis. We developed a statistical approach which enabled us to pool the results of both analyses. For the analysis of pregnancy as risk factor for venous thrombosis only in women, simple pooling of both control groups was possible. However, lower pregnancy rates than expected were encountered in the partner group and higher rates in the RDD group. After combining both control groups, pregnancy frequencies were comparable with data from the general Dutch population. Frequencies of the factor V Leiden mutation, an example of a genetic risk factor, were identical in both control groups and in line with published data, indicating that for the analyses of this genetic risk factor both control groups were equally suitable. Our experience with the inclusion of two different control groups might be useful to others for choosing the most optimal research design and statistical approach.

INTRODUCTION

When designing a case-control study a very important decision is the choice of the appropriate control group. The purpose of a control group in a case-control study is to indicate the expected frequency of an exposure in patients under the null-hypothesis that there is no relation between exposure and disease. Therefore a uniform requirement for control selection is that the control group should be selected from the same source population as the cases independently of their exposure status[1;2].

This general aim nevertheless leads to several options in practice. Control subjects can be selected from the general population, such as random population control subjects, partners, friends or neighbors. Another potential source of control subjects is the hospital in which cases are hospitalized. Usually, there will be advantages for one group that are missing in the other, and vice versa. For example, random population controls may be more difficult to locate and less motivated to take part in the study than partners, friends or neighbors[3]. Situations arise in which the investigator may face a choice between two or more possible control groups to use. When different types of research questions are addressed and adjustment for different variables is required, multiple control groups can be useful. However, it has been suggested that the value of multiple control groups is limited[1], since it can lead to inconsistent results and proper analysis may become complex.

In the Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA study), a very large population-based case-control study, we believed to have good reasons to include two different control groups, a partner control group and a random population control group. We included partners because the main focus of the study was on genetic risk factors for venous thrombosis and their interaction with environmental and lifestyle factors. It seemed unlikely that partners would select each other based on genetic differences in coagulation parameters. We also wanted to study environmental factors that are closely linked to lifestyle and for which partner controls might control for unmeasured confounding. In addition we assumed that asking partners would make it easier to recruit control subjects with malignancies or chronic diseases, which was a requisite if we wanted to study these diseases in relation to the risk of venous thrombosis. For the analysis with partner controls we had envisaged either a matched analysis of partners or an unmatched analysis with the opposite sex partners of cases becoming controls for same-sex cases. However, the proportion of men and women in the patient and control group was different in specific age categories; in particular there were very few young men with venous thrombosis (cases), resulting in few young female partners (controls), while there were many young female cases of

venous thrombosis. A population control group was added, that would be useful for certain analyses (such as pregnancy in young women), and might increase the overall numbers for the genetic analyses – as we did not expect differences in genetic make-up between partner controls and population controls. In the analysis phase, however, we learned that we had to distinguish quite carefully which analyses would be done with what control groups, as there were some unexpected differences, predominantly in environmental and lifestyle factors. In the process, we also had to devise a method for statistically combining the control groups if the analysis with one control group had to be matched and the other not. This process, as well as our solutions, might be useful to others who embark on large-scale gene environment interaction studies.

MEGA STUDY

Patients and partners

Between March 1999 and September 2004, we included consecutive patients with a first diagnosis of venous thrombosis. Patients were selected from the files of six large anticoagulation clinics in the Netherlands, which monitor anticoagulation treatment in all patients in a geographically well-defined area. Patients between the age of 18 and 70 with deep venous thrombosis of the leg, pulmonary embolism or a combination of these diagnoses were included. Patients with severe psychiatric problems or those unable to speak Dutch were considered as ineligible for practical reasons.

During the inclusion period partners of patients were asked to participate as control subjects. Only partner control subjects between the age of 18 and 70 with no history of deep venous thrombosis were included and the same exclusion criteria were applied as for patients.

Random digit dialing control subjects

From January 2002 until September 2004, another control group was recruited by using the random digit dialing (RDD) method according to Waksberg[4]. Only RDD control subjects between the age of 18 and 70 with no recent history of deep venous thrombosis were included and the same exclusion criteria were applied as for patients. The RDD method has proved to be a constructive method to collect a nearly random sample of all individuals in the population[5]. This method employs a two stage design which increases the likelihood of contacting households. Within

the geographical inclusion area, area codes and prefix numbers (first three digits of personal telephone number) combinations were obtained. For efficiency reasons, the prefixes were not generated completely at random in our study but were generated from the prefix numbers of the patients. To these prefixes, different random combinations of the next two digits were added. These eight digits formed the first stage of the sampling unit, i.e. the Primary Sampling Units (PSUs). To each PSU again two digits were added which were randomly generated by the computer. This number was dialed to determine whether or not it reached a household. If it did not reach a household because the telephone number was not in use or was used by a business or institution, the PSU was dropped from further consideration. If it did reach a household, 19 new numbers with the same PSU were randomly generated by the computer. Per household a maximum of seven attempts were made at different time points of the day and at different week days, with once at least three weeks between two attempts.

This procedure of control sampling was expensive and time-consuming; on average only three persons per hour were included. The response rate is hereby dependent on demographic characteristics of the target population and telephone skill of the interviewers[5]. In addition the RDD method is only useful if the vast majority of individuals live in households with a fixed telephone. In December 2005 fixed telephone coverage in the Netherlands was still very high (96%)[6], indicating that telephone coverage was more than enough for our RDD method. However in the nearby future, increasing use of mobile phones will decrease the ability for the RDD method to target specific areas within a country and achieve complete coverage.

An important consideration in random digit dialing surveys is bias introduced by non-responders. Non-response bias can be a problem if responders differ from non-responders for the measured variables[7;8]. Most studies have found that reluctant respondents are older and less educated than respondents who readily agree. Differences with respect to income, occupation, race and marital status have been inconsistent[8].

For efficiency reasons, we frequency matched the random control subjects to the patients who provided a blood sample according to age and sex. With each telephone call we asked a specific person within a household to participate (e.g. youngest woman between 20 and 50) and therefore avoided that the first person who picked up the phone, who maybe more mobile and healthier, was constantly included as control subject.

Data collection

Within a few weeks after diagnosis and registration at the anticoagulation clinics patients received a letter with information about the study and were subsequently contacted by phone. Partners of patients were also invited to participate. If patients or partners refused to participate the reason for refusal was asked for. Patients, partners and random digit dialing control subjects received a questionnaire shortly after inclusion by phone. The questionnaires included items on potential risk factors for venous thrombosis e.g. body weight, body height and pregnancies. Most questions referred to a period of 12 months prior to the index date, i.e. the date of diagnosis of thrombosis of the patient or the date of filling in the questionnaire for the random control subjects. For partners the date of diagnosis of thrombosis of the patient was used as index date in the body mass index analyses and in the pregnancy analyses the date of filling in the questionnaire was used.

From March 1999 till June 2002, patients and their partners were asked to visit the anticoagulation clinic where after an overnight fast a blood sample was drawn at least three months after withdrawal of anticoagulation. Only in case of continuous use for more than one year a blood sample was taken during anticoagulation therapy. From December 1999 onwards, self-administered buccal swabs were obtained by mail when participants were unable or unwilling to provide a blood sample. From June 2002 onwards, blood draws were no longer performed in patients and their partners, and the study was restricted to DNA collection by buccal swabs sent by mail. The RDD controls were invited for a blood draw within a few weeks after the questionnaire was sent. Within this group buccal swabs were sent when someone refused the blood draw. In the blood samples and buccal swabs prothrombotic mutations including the Factor V Leiden (G1691A) mutation were determined. A detailed description of blood collection and DNA analysis for factor V Leiden in the MEGA study has been published previously[9].

RESPONSE RATES AND GENERAL CHARACTERISTICS

During the inclusion period, 5961 eligible patients, 3586 eligible partners and 4346 eligible RDD control subjects were approached to participate. In the patient group, 4957 patients (83%) were willing to participate, partners had a similar response rate (n=2917, 81%), and 3000 (69%) RDD control subjects participated (figure 1).

Of the participating patients 92% returned a questionnaire compared to 95% and 93% in partners and RDD control subjects. During the first part of the MEGA study (March 1999-June 2002) a blood sample was provided by 73% of
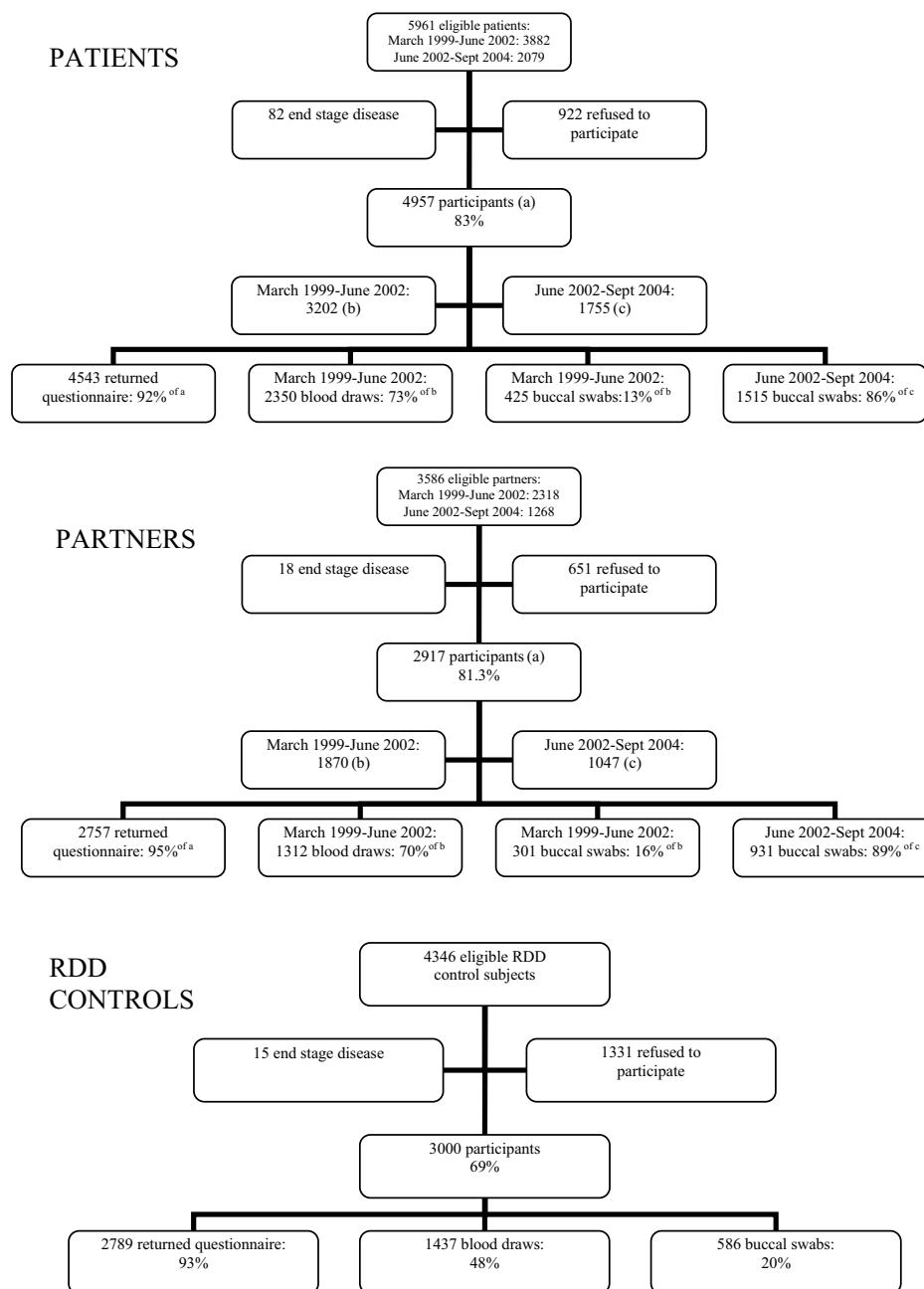
PATIENTS

5961 eligible patients:
March 1999-June 2002: 3882
June 2002-Sept 2004: 2079

82 end stage disease

922 refused to participate

4957 participants (a)
83%

March 1999-June 2002: 3202 (b)

June 2002-Sept 2004: 1755 (c)

4543 returned questionnaire: 92% [of a]

March 1999-June 2002: 2350 blood draws: 73% [of b]

March 1999-June 2002: 425 buccal swabs: 13% [of b]

June 2002-Sept 2004: 1515 buccal swabs: 86% [of c]

PARTNERS

3586 eligible partners:
March 1999-June 2002: 2318
June 2002-Sept 2004: 1268

18 end stage disease

651 refused to participate

2917 participants (a)
81.3%

March 1999-June 2002: 1870 (b)

June 2002-Sept 2004: 1047 (c)

2757 returned questionnaire: 95% [of a]

March 1999-June 2002: 1312 blood draws: 70% [of b]

March 1999-June 2002: 301 buccal swabs: 16% [of b]

June 2002-Sept 2004: 931 buccal swabs: 89% [of c]

RDD CONTROLS

4346 eligible RDD control subjects

15 end stage disease

1331 refused to participate

3000 participants
69%

2789 returned questionnaire: 93%

1437 blood draws: 48%

586 buccal swabs: 20%

Figure 1. Response rates of patients, partners and RDD control subjects

participating patients and 70% of partners. Forty-eight percent of eligible RDD control subjects provided a blood sample. During the second part of the study (June 2002-September 2004), a buccal swab was obtained from 86% of patients

Table 1. Reasons for non-response in patients, partners and RDD control subjects

| | Patients | | Control subjects | | | |
| | | | Partners of participating patients | | RDD controls | |
	N	%	N	%	N	%
Refused to participate	922	100	651	100	1331	
No willingness	514	55.7	628	96.5	1243	93.4
Too many hospitals	93	10.1	–		–	
Not mobile	17	1.8	1	0.2	–	
Untraceable	271	29.3	14	2.2	88	6.6
Filled in questionnaire about recurrent VT	5	0.5	–		–	
Reason unknown	22	2.4	8	1.2	–	

– = not specified

and 89% of partners. Reasons why persons refused to participate are presented in more detail in table 1.

Mean age of 4957 patients was 48.6 (5th-95th percentiles, 25.7-67.9), the 2917 partners were on average 48.3 years (5th-95th percentiles, 28.0-66.1) and the 3000 RDD control subjects had a mean age of 45.3 (5th-95th percentiles, 23.5-66.9). Fifty four percent (n=2680) of patients, 50% (n=1463) of partners and 57% (n=1719) of RDD control subjects were women.

DIFFERENT RESEARCH QUESTIONS, DIFFERENT USE OF CONTROLS

In the MEGA study we investigated genetic or acquired factors and their interaction as possible risk factors for venous thrombosis. As genetic risk factors several prothrombotic mutations, such as the factor V Leiden mutation, were measured in blood samples or buccal swabs collected from the participants. Included were a wide range of acquired risk factors like malignancies, surgery, injuries and various lifestyle related risk factors as pregnancy, oral contraceptive use, overweight, smoking, physical activity, alcohol use and (air) travel. When analyzing lifestyle factors as possible risk factors for venous thrombosis different considerations concerning the choice of a control group have to be made compared to the analysis of genetic risk factors. It is challenging to use both control groups in such a way that statistical power is maintained and bias is reduced to a minimum.

For set forth the analytic considerations of two different control groups we will describe the association of a general lifestyle risk factor (body mass index), a lifestyle risk factor in women (pregnancy) and an example of a genetic risk factor (factor V Leiden mutation) with the risk of venous thrombosis. We will present

a statistical method that allowed us to use both control groups in the analyses of body mass index as risk factor for venous thrombosis.

Body mass index- General lifestyle risk factor

For the analyses of body mass index (BMI) as risk factor for venous thrombosis[10] the most obvious control group seems to be the RDD control group because one instinctively would say that partners are too much alike. When we investigated the BMI distribution in patients, partners and the RDD controls, frequencies of overweight (BMI: 25-29 kg/m^2) and obesity (BMI: ≥ 30 kg/m^2) were indeed more similar in patients and their partners than in patients and the RDD controls, resulting in lower risk estimates when using the partner control group compared to the RDD control group (table 2). These results were obtained with an unconditional logistical regression analysis, which is not correct because it uses partners of cases as control subjects for other cases. Partners are matched with patients and this matching has to be considered in the statistical analysis since ignoring matching generally introduces bias, even if the matched variable is not a confounder[1]. Performing an unmatched analysis with matched data will result in an underestimation of the true effect. Matching was accounted for with a conditional logistic regression analysis[11], i.e. matched analysis, which adjusts for similar lifestyle factors between patients and their partners by including only discordant pairs. In table 3 the results of the matched analysis with patient-partner pairs is presented. Risk estimates appeared to be still somewhat lower compared to the analysis with the RDD control subjects (overweight$_{partners}$ OR 1.45, CI95 1.26-1.67; overweight$_{RDD}$ OR 1.83, CI95 1.63-2.05; obesity$_{partners}$ OR 1.81, CI95 1.49-2.20; obesity$_{RDD}$ OR 2.87, CI95 2.45-3.35). A possible explanation for this difference is that adjustment for similar lifestyle factors in the matched analysis may include some unknown, unmeasured confounders, which will lead to risk estimates closer to the real estimates compared to the risk estimates obtained from the analysis with the RDD controls. It is important to

Table 2. Body mass index (BMI) distribution in patients, partners and RDD control subjects – Unmatched analyses

BMI (kg/m^2)	Patients		Partners		RDD		OR$_{partner}$* (CI95)	OR$_{RDD}$* (CI95)
	N	%	N	%	N	%		
<25	1369	36.5	1306	44.8	1409	55.7	1	1
25-29	1593	42.4	1172	40.2	848	33.5	1.33 (1.20-1.49)	1.83 (1.63-2.05)
≥30	794	21.1	438	15.0	274	10.8	1.75 (1.52-2.01)	2.87 (2.45-3.35)
Total	3756		2916		2531			

*adjusted for age and sex

Odds ratios (ORs) calculated with unconditional logistic regression

Note: BMI analyses were performed in non-pregnant individuals without malignancies

Table 3. BMI as risk factor for venous thrombosis - Matched analyses

BMI (kg/m^2)	Patients	Partners	OR$_{matched}$* (CI95)
<25	739	925	1
25-29	949	860	1.45 (1.26-1.67)
≥30	415	318	1.81 (1.49-2.20)
Total	2103	2103	

*adjusted for age and sex

Table 4. BMI as risk factor for venous thrombosis - Combined analyses with patients, partners and RDD control subjects

BMI (kg/m^2)	Patients	Partners	RDD	OR$_{combined}$* (CI95)
<25	1369	925	1409	1
25-30	1593	860	848	1.71 (1.54-1.89)
≥30	794	318	274	2.45 (2.14-2.80)
Total	3756	2103	2531	

*adjusted for age and sex

Table 2-4: Adapted from British Journal of Haemaology 2007;139:289-269

realize that in the matched analysis only patient-partner pairs can be included, resulting in less power than the analysis with the RDD control subjects. Besides this, both the patient and the partner of a pair must have valid data for the required variable, otherwise the whole pair cannot be included in the analysis. Finally, the matched analysis itself only uses pairs who are discordant for the variable of interest, resulting in further reduced power.

Using the RDD control subjects in the analyses of BMI as risk factor for venous thrombosis may result in a slightly overestimation of the true risk estimates because there were somewhat fewer RDD controls with overweight compared to the general Dutch population. According to data of the Central Bureau of Statistics in the Netherlands the prevalence of overweight and obesity was respectively 36% and 11% during the study period[12], compared to a 33% and 11% found in the RDD group.

Both partner and RDD analyses showed consistent results in terms of clearly increased risks. In a combined analysis the most powerful estimate was obtained. We used a simple approach in which the estimates of the odds ratios of the two analyses were pooled[13]. In this combined analysis we accounted for the correlation between the estimated odds ratios since most patients were included both in the matched and the unmatched analysis. Table 4 presents the odds ratios of the combined analysis (OR$_{overweight}$ 1.71, CI95 1.54-1.89, OR$_{obesity}$ 2.45, CI95 2.14-2.80), which were of course in between partner and RDD odds ratios.

When analyzing the risks in men and women separately it was not possible to perform a matched analysis with the partner controls, as control individuals were nearly always of the opposite sex to the cases.

Pregnancy- Lifestyle risk factor in women

In the MEGA questionnaire, participants were asked if they had been pregnant in the year before the index date or if they were still pregnant, and what the (expected) date of delivery was.

In the analysis of pregnancy as risk factor for venous thrombosis, only women were included. In addition, only participants with a partner were included in the analysis, since being in a relationship affects the probability of getting pregnant. During the invitation by phone, patients were asked if they had a partner, partner controls had a partner per definition, and civil status was asked for in the questionnaire, also allowing the inclusion of only RDD controls with a partner. However, we encountered a much higher frequency of pregnancies in the RDD control subjects with a partner than in the partner control subjects (table 5). The percentage of pregnant or postpartum women was 12.3% in the RDD control group and 3.9% in the partner control group compared to 8.8% in the general population. These

Table 5. Pregnancy and postpartum in patients, partners and RDD control subjects

Pregnancy status	Patients		Partners		RDD		$OR_{partner}^{*}$	OR_{RDD}^{*}	OR_{total}^{*}
	N	%	N	%	N	%			
Neither	163	61.3	394	96.1	371	87.7	1	1	1
Pregnant	35	13.2	14	3.4	44	10.4	9.28 (4.37-19.70)	3.60 (2.01-6.44)	4.67 (2.72-8.00)
Postpartum[†]	68	25.5	2	0.5	8	1.9	198.07 (38.00-1032.29)	42.22 (17.38-102.60)	61.21 (27.06-138.48)
Total	266		410		423				

*adjusted for age, [†]three months after delivery
ORs calculated with unconditional logistic regression
Note: Pregnancy analyses were performed using women who were between 18 and 50 years of age, had a partner, did not use oral contraceptives or hormone replacement therapy and had no malignancies or a partner* with malignancies (*for patients and partner controls).

Table 6. Different stages of pregnancy and postpartum in patients, partners and RDD control subjects

Pregnancy status	Patients		Partners		RDD	
	N	%	N	%	N	%
Neither	161	60.5	378	92.2	347	82.0
1st and 2nd trimester	8	3.0	6	1.5	30	7.1
3rd trimester	27	10.2	8	2.0	14	3.3
Puerperium (1-6 weeks)	65	24.4	1	0.2	5	1.2
7 weeks to 3rd month postpartum	3	1.1	1	0.2	3	0.7
4th month to 1 year postpartum	2	0.8	16	3.9	24	5.7

Table 5-6: Adapted from Journal of Thrombosis and Haemostasis 2008;6:632-637

Chapter 7

frequencies in the control groups were unexpected; before the start of our study we assumed that including partners would make it easier to recruit pregnant individuals because pregnant women in general would be less motivated to participate in a study. However, the opposite appeared to be true. The high frequency in the RDD group may be due to more awareness of health issues in pregnant women and therefore more willingness to participate than non-pregnant women. The low frequency in the partner control group remains difficult to explain.

It was possible to combine the two separate control groups into one large group. The prevalence of pregnant or postpartum control women (8.1%) then became similar to that of the general population (8.8%). Not only for the overall analysis but also for the stratified analysis of different stages of pregnancy and the postpartum period it was important that the proportion of control subjects in each time frame during and after pregnancy was a good reflection of the general population (table 6). To verify this, we calculated the expected number of controls in each period, using data from the general population[14]. During pregnancy the number of controls in the overall group was similar to what we would expect to find (6.9% compared to an expected 6.6%). In the first three months postpartum we observed a lower number of controls (1.2% compared to an expected 2.2%), possibly due to a reduced motivation to participate in our study after child delivery. In the period from four months up to one year postpartum the number of controls was still somewhat reduced (4.7% compared to an expected 6.6%). These lower proportions might have resulted in a slight overestimation of relative risks in the postpartum period.

These analyses illustrate that the inclusion of multiple control groups appeared to be very useful. A priori assumptions about control group characteristics were not in line with the collected data. If only a partner control group or only the RDD control group was collected, pregnancy associated risks were either over- or underestimated.

Factor V Leiden- Genetic risk factor

For genetic risk factors it is unlikely that their frequency is different in partners compared to RDD control subjects. However, the prevalence of factor V Leiden is related to ethnicity[15] so you could speculate that if partners chose their partner according to ethnicity the factor V Leiden distribution in partners would be dissimilar compared to RDD control subjects. In the MEGA study most participants were of Dutch origin, so differences in the distribution of factor V Leiden due to intra-racial partnerships were unlikely. For the RDD controls you could hypothesize that RDD controls with a positive family history of venous thrombosis will

Table 7. Factor V Leiden mutation (FVL) in patients, partners and RDD control subjects

FVL	Patients		Partners		RDD		OR$_{partner}$*	OR$_{RDD}$*	OR$_{total}$*
	N	%	N	%	N	%			
-	3612	84.3	2403	94.7	1914	94.6	1	1	1
+	675	15.7	134	5.3	109	5.4	3.38 (2.78-4.09)	3.36 (2.72-4.15)	3.36 (2.88-3.92)
Total	4287		2537		2023				

*adjusted for age and sex
ORs calculated with unconditional logistic regression

be more willing to give blood than RDD controls without a positive family history, leading to an overestimation of the prevalence of factor V Leiden in this group. However, we found the same percentage of individuals with factor V Leiden in the partner and the RDD group (partner controls, 5.3%; RDD controls, 5.4%) (table 7). Obviously, both percentages could be an overestimation of the true prevalence, but the percentages were equal to a previously recorded prevalence of factor V Leiden in Caucasians[16].

Since both control groups had the same percentage of factor V Leiden carriers and this percentage was supported by literature both control groups were combined as if they were one in an unconditional logistic regression analysis (table 7).

DISCUSSION

In the MEGA study, a large population-based case-control study, we evaluated the use of two different control groups, a partner control group and a RDD control group, in the analyses of three different types of research questions. We learned that we had to distinguish quite carefully which analyses would be done with what control groups, as there were some unexpected differences. For the evaluation of body mass index, we had to devise a method for statistically combining the control groups in the analysis. Using the partner control group asked for a matched analysis and for the RDD group an unmatched analysis was required. For pregnancy, simple pooling of both female control groups was possible. However, lower pregnancy rates than expected were encountered in the partner group and higher rates in the RDD group. After combining both control groups, pregnancy frequencies were comparable with data from the general Dutch population. Frequencies of the factor V Leiden mutation were identical in both control groups and in line with published data, indicating that for the analyses of genetic risk factor both control groups were equally suitable.

Chapter 7

There are only a few studies reporting their experience with multiple control groups. In 1983, Savraky and Clarke wrote a paper that summarizes their experience in using hospital and neighborhood control subjects[17]. When testing the hypothesis if oxidative hair dyes were carcinogenic, they found to their surprise lower rates of hair dye use among 314 hospitals (40.5%) than among 470 neighborhood control subjects (52.8%). Several other striking differences were observed. Compared with hospital controls, neighborhood controls were older, ethnically more heterogeneous, less likely to be oral contraceptive users and more likely to be smokers. The investigators believed that most of these differences arose from different lifestyles in the relatively rural region from which the hospital controls were derived and in the urban region that provided the neighborhood group. These geographical differences demonstrate the importance of selection of patients and control subjects from the same source population[18]. A study investigating the association between machining fluid and laryngeal cancer risk used control subjects with oral cancer and a stratified random sample of all deaths in a distinct geographical area as control subjects[19]. When cases (n=888) were compared to oral cancer controls (n=752) high exposure to machining fluids resulted in a 1.5-fold increased of laryngeal cancer. However, when cases were compared with population controls (n=3594) no increased risk of exposure was found. A possible explanation, besides a chance finding, may be that exposure data quality for the cases and oral cancer controls may have differed from that of the population controls. These studies illustrate the problem of multiple results; at least one of the results is biased. Only further external information could help to evaluate the likely extent of bias in the estimates from different controls.

Not only characteristics may differ substantially between control groups, but also response rates may vary. In the MEGA study partner controls were more willing to participate than RDD controls (83% versus 69%). Especially for blood draws the difference was considerable; 70% percent of participating partners and 48% of participating RDD controls provided a blood sample. A possible explanation for this difference may be that partners motivated each other to participate and were able to join each other to the location of the blood draw. Another consideration which may explain differences between RDD and partner response rates is the fact that partners of non-participating patients were not included in the non-response; if a patient refused to participate, we did not ask the patients partner to participate. Thus beforehand a selection of more willing couples, with participating patients, was made which could have positively influenced the partner response.

In the analyses of BMI as risk factor for venous thrombosis partner controls were included in a conditional logistic regression analysis (matched analysis) since ignoring matching introduces bias. Aside from the complication of matching, the

fact that partners have a relationship may be associated with certain characteristics which make partners somewhat different from the source population[1].

It is important to realize that a priori assumptions about control group characteristics may not be confirmed by the data. We had the wrong assumption that including partners would make it easier to recruit pregnant women or individuals with severe diseases. Besides the low frequency of pregnancies in partners, frequencies of malignancies were also different from what we expected; both control groups had about the same percentage of malignancies (data not shown). These findings could indicate that health issues for RDD controls are an extra motivation to participate. The low frequency of pregnancies in partners is however difficult to understand. It may be due to the fact that partners were approached via the patient. It is possible that because of the pregnancy or disease of the partner, the patients decided on their own that their partner was not willing to participate. This illustrates the importance of asking in detail reasons for non-response.

In conclusion, when different types of research questions are addressed in a case-control study, it is important to think thoroughly about control group choice and the way controls are to be used in the statistical analyses. We hope the discussion of our experience in using multiple control groups can help others to create the most optimal study design and statistical approach for answering their research questions.

ACKNOWLEDGEMENT

Chapter 7

REFERENCES

1. Rothman K, Greenland S. Modern Epidemiology. Philadelphia: Lippincott-Raven publishers; 1998.

2. Miettinen OS. The "case-control" study: valid selection of subjects. J.Chronic.Dis. 1985; 38:543-548.

3. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. Am.J.Epidemiol. 1992;135:1029-1041.

4. Waksberg J. Sampling methods for Random Digit Dialing. Journal of the American Statistical Association 1978;73:40-46.

5. Hartge P, Brinton LA, Rosenthal JF et al. Random digit dialing in selecting a population-based control group. Am.J.Epidemiol. 1984;120:825-833.

6. European Commission. E-communications Household Survey (accessed 3 December 2007, online). http://ec.europa.eu/information_society/policy/ecomm/doc/library/ext_studies/ecomm_household_study/eb_jul06_main_report_en.pdf

7. Funkhouser E, Macaluso M, Wang X. Alternative strategies for selecting population controls: comparison of random digit dialing and targeted telephone calls. Ann.Epidemiol. 2000;10:59-67.

8. Voigt LF, Koepsell TD, Daling JR. Characteristics of telephone survey respondents according to willingness to participate. Am.J.Epidemiol. 2003;157:66-73.

9. Blom JW, Doggen CJM, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. JAMA 2005;293:715-722.

10. Pomp ER, le Cessie S, Rosendaal FR, Doggen CJM. Risk of venous thrombosis: obesity and its joint effect with oral contraceptive use and prothrombotic mutations. Br.J.Haematol. 2007;139:289-296.

11. Cannegieter SC, Doggen CJM, van Houwelingen HC, Rosendaal FR. Travel-related venous thrombosis: results from a large population-based case control study (MEGA study). PLoS.Med. 2006;3:e307.

12. Central Bureau of Statistics The Netherlands. Reported health and lifestyle (Accessed 3 December 2007, online). http://statline.cbs.nl/StatWeb/table.asp?PA=03799&D1=267-270&D2=0-17&D3=0&D4=2-4&DM=SLNL&LA=nl&STB=G1&HDR=G2,T,G3

13. Le Cessie S, Nagelkerke N, Rosendaal FR, Van Stralen KJ, Pomp ER, Van Houwelingen HC. Combining matched and unmatched control groups in case-control studies. Am.J.Epidemiol. (accepted for publication).

14. Central Bureau of Statistics The Netherlands. CBS history population 2002 (accessed 2 May 2007, online). http://statline.cbs.nl/

15. Dowling NF, Austin H, Dilley A et al. The epidemiology of venous thromboembolism in Caucasians and African-Americans: the GATE Study. J.Thromb.Haemost. 2003;1:80-87.

16. Rosendaal FR. Venous thrombosis: a multicausal disease. Lancet 1999;353:1167-1173.

17. Stavraky KM, Clarke EA. Hospital or population controls? An unanswered question. J. Chron. Dis. 1983;36:301-307.

18. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. Lancet 2005;365:1429-1433.

19. Russi M, Dubrow R, Flannery JT, Cullen MR, Mayne ST. Occupational exposure to machining fluids and laryngeal cancer risk: contrasting results using two separate control groups. Am.J.Ind.Med. 1997;31:166-171.