January 2014

Is Smoking Protective Against The Development Of Endoleak?

Ziyi Li
Yale University, lizy0820@gmail.com

Follow this and additional works at: http://elischolar.library.yale.edu/ysphtdl

Is smoking protective against the development of endoleak?

Results of analyzing Endoleak data from a VA trial

Li, Ziyi

Master Thesis

Advisor:

Shuangge S. Ma PhD

Tassos C. Kyriakedes PhD

Yale School of Public Health, 2014

ABSTRACT

Background: Endoleak has been reported as a major complication among patients who have undergone endovascular repair of their abdominal aortic aneurysm (AAA). Several studies made efforts towards identifying the risk factors for the development of endoleak through various studies and data analyses. Among all the findings, one result--smoking significantly lowers a patient's risk for an endoleak--has been confirmed multiple times. But such a result appears to be contradictory to what researchers have anticipated; yet the contradictory has not been closely studied.

Methods: Data on endoleaks from the Department of Veterans' Affairs Open Versus Endovascular Repair (OVER) randomized controlled trial was used, and 419 male subjects with smoking histories were included in the analysis. A series of logistic regression models and propensity score models was constructed with the baseline and follow-up variables. Multiple imputation techniques were utilized to minimize the impact of missing data and to improve analytical robustness. In addition, a simulation study was also undertaken to better evaluate the models above.

Results: About half (5/12) of the logistic regression models supported the significant effect on endoleak of smoking in the model, and a smaller proportion (3/18) of propensity score models indicated that smoking was a significant factor for endoleak. Missing values had an important impact on the results. Although smoking's effects were not significant for little more than half of the models, odds ratio of developing endoleak for current smokers were always less than 1 compared with non-current smokers, which may be clinically meaningful. Results from simulation studies suggest that clinical trials with larger sample size might be necessary to reach a definitive conclusion on this topic.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

INTRODUCTION

Open heart surgery used to be the routine treatment for patients with Abdominal Aortic Aneurysm (AAA) more than twenty years ago[1]. Ever since the introduction of minimally invasive endovascular stent grafts in 1986[2], endovascular aneurysm repair (EVAR) had become a popular alternative treatment for AAA patients, partly because of the increasing recognition of EVAR's superior advantage of decreasing perioperative mortality[1,3-5]. In spite of the benefits, endovascular repair had drawbacks. Endoleak, a symptom of continuous blood flow between graft and aneurysm sac, had been discovered as a major complication among patients who had received endovascular repair[6]. Endoleak was also regarded as a signal of surgery failure in several studies[7,8]. Although the absence of an endoleak does not necessarily mean that the patient is free from risk (Gilling-Smith and his colleagues found that ruptures could happen to patients with no endoleak)[9], there is no doubt that the appearance of an endoleak implies unsuccessful exclusion of the aneurysm. It had been found that certain types of endoleak (for example Type I endoleak) were more dangerous than some other types, since these types of endoleak were closely related with a higher risk of aneurysm sac enlargement, and the aneurysm sac enlargement may subsequently cause aneurysm rupture[10,11]. Consequently, patients with endoleak may have to receive secondary interventions to fix the related problems, and the cost of treatment increases at the same time[12].

Experience from several studies had showed that 20% - 35% of the patients undergoing endovascular aneurysm reparation developed endoleaks after interventions[12-15]. Because of the high frequency and potential risks of endoleak, researchers had spent efforts to investigated the risk factors for the development of endoleak [11,15-18]. Knowledge about these risk factors can potentially assist health practitioners to pre-select AAA patients who may benefit from endovascular aneurysm repairation and to

lower the possibility of endoleak.   Studies about endoleak suggested that age, aneurysm size, smoking

status, etc. were related with the incidence of endoleak.   In analyzing data from these risk factors,

smokers had been identified to be at lower risk of endoleak, was is the most unexpected and least

understood[11, 15-18].

As part of the EUROSTAR study, Marrewijk et al., Mohan et al., and Buth et al. collected pre- and

post-operative data of AAA patients who had received endovascular repairation from 110 medical centers

all over Europe[11,15,17].   They reported that the non-endoleak group had lower proportion of smokers

compared to the endoleak group.   Mohan et al. had completed the most comprehensive study among the

three studies on the topic of smoking and endoleak[15], and through a series of logistic regression models,

Mohan et al. found that current smokers smoking more than 20 cigarettes per day had the lowest risk of

endoleak compared with current smokers smoking less than 20 cigarettes, and ex-smokers.   Such results

appeared to confirm   the hypothesis that smoking provided some protection against the development of

endoleak.   The results above demonstrating the relation between smoking and endoleak were not

coincidental.   Zarins et al. published similar results from their AneuRx multicenter clinical trial in 1999

and also found higher proportions of smokers in patients without endoleak than in those with endoleak[16].

Although the results about the relation between smoking and endoleak had accumulated through different

studies, a close scrutiny over the analysis demonstrated that results were generated through the application of

student t-tests, chi-square tests, and/or multivariate regression models[11,15-17].   Because smoking would

obviously raise some ethical concerns in the context of a randomized trial, data on such topics were mostly

observational.   Observational data could have certain disadvantages, such as unbalanced baseline

characteristics between two experiment arms[19]. However, it didn't mean that simple methodology was not good enough. On the contrary, when both simple and complex methods were feasible in solving problems, simple methods should always be the first consideration. We suggested that researchers should be more cautious when conducting analysis based on observational data. In this context, more comprehensive analysis needs to be done to better understand the relationship of quitting smoke and developing endoleak.

In this paper, I used both logistic regression model and propensity score model to investigate the effect of quitting smoke on the incidence of an endoleak in male veterans. The analysis was based on the data from Open versus Endovascular Repair (OVER) trial, which was conducted by Veterans Healthcare System from 2002 to 2011. More description about dataset can be found in the methods section. In addition to analyzing the real experiment data from OVER trial, I also used imputation methods and a simulation study to minimize the influence of missing data and maximize our understanding of parameter estimates in both models.

METHOD

Data source

OVER trial was a randomized, multicenter clinical trial conducted by Veterans Affairs Cooperative Study Group (VACSPCC) from October 15, 2002 to October 15, 2011. The main goal was to provide information on short-term and long-term comparison between open repair and endovascular repair of AAA. The OVER trial enrolled 881 veterans who were 49-year or order and who were eligible for both open and endovascular repair of their AAA. Of the 891 study participants, a total of 444 were randomized to endovascular group and 437 were randomized to open group. Not all subjects successfully received the treatment they were assigned to (31, 3.5% of the subjects failed to accept the

6

assigned treatment, including 18 subjects randomized to EVAR group and 13 subjects randomized to

Open group).

Baseline measurements were collected for all subjects including age, height, weight, left ABI, right ABI,

smoking status, and family history of abdominal aortic aneurysm.   After repairation, follow-up visits were

scheduled at 1 month, 3 months and 6 months in the first year, and once every year from the second year on, .

The follow-up visits for patients who had received endovascular repairs included computed tomography and

plain radiography of the abdomen.   Endoleak status was ascertained at the same time. More details about

OVER trial could be found in the published short-term and long-term reports[3,4].

Although the OVER trial collected data of the subjects in both open repair and endovascular repair groups,

I only used the data of patients who received endovascular repair, because only those who received

endovascular repair had the risk of developing endoleak.   The original endovascular population

included 3 female subjects (0.68%) and they were eliminated from the analysis to keep the homogeneity

of study group.   In addition, since our study topic was about quitting smoke, subjects who never smoked

were also excluded.   The final data set included 419 endovascular patients.   Of all subjects, a total of

252 were current non-smokers but had smoking history which considered as treatment group, a total of 167

were current smokers which considered as control group.   In this study, "treatment" stood for quitting

smoke.

Data analysis

Preliminary analysis – Population characteristics (at time of randomization) were tabulated by smoking

status and endoleak incidence respectively.   P values based on two sample t test (for continuous variables) or chi-square test (for binary variables) were also generated for each baseline variables.   For unbalanced variables, histograms of variables against smoking status or endoleak status were generated to better understand their distributions.

Propensity score methods – As discussed in introduction section, data from an observational study may have   unbalanced characteristics between treatment group and control group.   And these differences may lead to biased estimation for treatment effects.   Propensity score for each subject was usually calculated in this context to balance the covariates differences between treatment group and control group, in order to reduce the bias.   The methods described by Rosenbaum and Robin[19] were used to generate propensity score for the purposes of this study.

Suppose subject i (i=1, …, N) were randomized to treatment (Zi=1) and control (Zi=0).   Given a vector of observed covariates, propensity score stood for the conditional probability of assignment to treatment (Zi=1) for this subject, and it was denoted by

$$e(x_i) = pr(z_i = 1|x_i)$$

where it was assumed that

$$pr(z_1, z_2, …, z_n|x_1, x_2, …, x_n) = \prod_i^N e(x_i)^{z_i}\{1 - e(x_i)\}^{1-z_i}$$

Rosenbausm and Rubin concluded that propensity score was the coarsest balancing score of observed covariates x[19].   Balancing score b(x) was defined as a function of x such that given b(x) the conditional distributions of x for subject in treatment group (Zi=1) and the subject in control group (Zi=0) were the same; in other words, given treated and controlled subjects had a same propensity score, the treatment assignment was strongly ignorable and the treatment effect could be calculated by the differences between the responses

of treated and controlled groups.   Balancing score was a bridge for non-randomized experiments to imitate randomized experiments.   A critical difference between randomized experiments and non-randomized experiment   s was whether the two groups were comparable.   Specifically, randomized experiments had two similar treatment groups (in terms of baseline variables, and except for the treatment modality), thus they may be directly compared; while non-randomized experiments may have systematically different treatment groups and direct comparison may not be meaningful.   Through balancing the score, the subjects in two groups could be matched or stratified, so that t h e meaningful treatment effect estimation could be achieved.

Rosenbaum and Rubin also suggested that propensity score could be modeled using logistic regression or discriminant score [19].   Discriminant score utilized the assumption that observed covariates have multivariate normal distribution, while logistic regression models had no prior distribution assumption for covariates.   One requirement for both calculation methods was that the data should not contain any missing data.   Logistic regression was used to estimate the propensity scores.   A detailed primary analysis of the characteristics of estimated propensity score can be found in the Results Section(page 13).   After propensity scores were generated, three methods – matching, stratification, and covariate adjustment – were used in the following analysis.

Matching by propensity score: Although it was straight-forward that the subjects with similar propensity score in treatment group and control group were matched , several techniques such as greedy matching with caliper distance, optimal matching, exact matching, complete matching, and mahalanobis matching, were usually used and these methods could either use 1:1 matching or 1:n matching.   Austin had conducted

comprehensive research to compare most of these matching techniques (matching on the logit of the

propensity score using calipers of width either 0.2 or 0.6 of the standard deviation of the logit of the

propensity score; matching on the propensity score using calipers of 0.005, 0.01, 0.02, 0.03, and 0.1; and 5 to

1 digit matching on the propensity score)[20,21].   Austin concluded that matching using caliper of width of 0.2

of the standard deviation of the logit of the propensity score and using caliper of width of 0.02 and 0.03 were

likely to have better performance for estimating treatment effects compared with other techniques.

Matching using caliper of width of 0.2 of the standard deviation of the logit of the propensity score was also

used in the current analysis.   In the endoleak dataset, the sample size of the treatment group was similar to

the sample size of the control group, so a greedy 1:1 matching with (a) 0.1 of the standard deviation of the

logit of the propensity score; (b) 0.2 of the standard deviation of the logit of the propensity score; (c) 0.25 of

the standard deviation of the logit of the propensity score, and (d) mahalanobis matching.

Stratification, also known as subclassification, was the second major method of using propensity scores to

adjust unbalanced characteristics between the treated and the control groups in observational studies.

Stratification was proposed by Cochran in 1965 and the original method was only used to adjust for

confounding variable(s)[22].   The main problem of the original method was that with the number of

confounding variables increasing, the number of strata or subclasses increased exponentially.   Stratifying

by propensity score was a useful way to solve such problem.   Moreover, propensity score methods can

generate more meaningful strata by considering all unbalanced binary and continuous variables when

modeling.   A question that usually arose when stratifying by propensity score was the number of strata used.

Researchers found that stratification with creating five classes was sufficient to eliminate 90 percent of the

bias due to the unbalanced covariates in treatment and control groups[22], and such results was supported

theoretically by Rosenbaum and Rubin's article[19].   As a result, propensity score-based stratification with

five strata was used in the analysis in this paper.   Detailed results and diagnostics can be found in the Result

Section (page 17).

Covariate adjustment was a method using propensity score directly as a covariate in modeling treatment

effects as responses.   Other unbalanced variables after adjusting with propensity score would also be

included in the model.   The difference between using propensity score as covariate in the regression

model and using all variables in the regression model as covariates might be a question.   Rosenbaum

and Rubin commented that "the point estimate of the treatment effect obtained from an analysis of

covariance adjustment for multivariate x, in fact, equaled to the estimation obtained from univariate

covariate adjustment for the sample linear discriminant based on x, whenever the same sample

covariance matrix was used for both the covariance adjustment and the discriminant analysis"[19].

Therefore, these two methods should generate similar results.   However, as suggested in the article by

RB D'Agostino. Jr., propensity score had the "dim-decrease" advantage[23].   When generating

propensity score, all related variables, including interaction terms and high order terms, can be included,

since the goal of logistic model at that stage was for prediction and including many variables did not

harm the results, rather would improve the accuracy.   On the contrary, parsimonious models were

usually preferred when regression model with all variables was used for easier diagnostics and

interpretation.   The Endoleak dataset used for analysis had more than 20 baseline variables, and thus in

this case propensity score may provide more accurate estimation for treatment effects than regression

model with all covariates.

Multiple Imputation - Missing values had always been a challenge in data analysis.   Rubin's multiple

imputation procedure filled in each missing position with a series of plausible values which may probably

represent the true values[25].   This approach provided a helpful strategy to address the missing value issue.

In practice, several techniques had been proposed and could be easily implemented by software, including

parametric regression method, propensity score method, and Markov chain Monte Carlo (MCMC) method[24].

For different missing patterns, different imputation techniques could be used.   For instance, parametric

regression method and propensity score method fitted monotone missing data patterns, and MCMC method

fitted any arbitrary missing pattern.   In the endoleak dataset, missing data pattern was not monotone, thus

MCMC method had been used for the imputation process.

Simulation - To better understand the variability of the parameter estimation and to further compare the

results of logistic regression and propensity score, I simulated 100 data sets on the basis of MCMC-imputed

endoleak data.   The first MCMC-imputed endoleak dataset was generated after 200 burn-in

estimation-maximization algorithm lops.   At that time, the algorithm had already converged and the

Markov chain had reached the state of stable.   The rest of the imputed endoleak datasets were generated

with 100 iterations between one and the successive one.

Each observation in the simulated data set was generated on the basis of imputed endoleak data set to make

the simulated data similar to endoleak data.   The simulation processes of variables were different between

continuous variables and binary variables.   For a continuous variable X, denote that the standard deviation

of X by SD_x and the simulated data point of X by X_n. X_n was generated by

$$X_n = X + rN(0, (0.1 \times SD_X)^2)$$

rN(0, $(0.1 \times SD_X)^2$) was the random number from normal distribution with mean 0 and variance $(0.1 \times SD_X)^2$. For a binary variable Y, denote that the simulated data point of Y by Y_n. Y_n was generated by

$$Y_n = I(rU(0,1) < 0.9) \times Y + I(rU(0,1) \geq 0.9) \times (1 - Y)$$

$rU(0,1)$ was the random number of uniform(0,1) distribution.

The simulated data was analyzed by logistic regression and propensity score as above.

RESULTS

Preliminary analysis

After eliminating female subjects and male subjects who had never smoked, the endoleak dataset included

419 subjects in total. Table1 contains the characteristics of all patients at baseline by smoking status and

primary outcomes. From Table 1, the first and the most unbalanced variable detected was age. The

average age of current smokers was 67.1, while the average age of current non-smokers was 70.8. Current

non-smokers were significantly older than current smokers (P value from two sample t-test was less than

0.0001). Age was also significantly related with endoleak occurrence (P value from two sample t-test

equaled 0.032). The average age of the subjects who had endoleaks was 70.5, which was larger than the

average age of those did not have endoleaks (68.8). Figure 1 presents the histogram of age versus smoking

status and versus endoleak occurrence. The figure confirmed that smokers were, on average, younger than

non-smokers; and the subjects with endoleaks were, on average, older than subjects without endoleaks.

In addition to age, weight was significantly related with smoking status or endoleak occurrence (although

diabetes history and anticoagulants medicine taking history had borderline P value against endoleak, 0.052

13

for both variables) among all other baseline measurements.   The average weight of current non-smokers was 92.6kg which was greater than the average weight of current smokers, 88.0kg.   No significant relation was found between weight and endoleak.   Figure 2 shows the histogram of weight versus smoking status and versus endoleak occurrence.

Table 1: Characteristics of the patients at the time of randomization

Variables	N	Smoking Status		P*	Outcome		P*
		Current Smokers (N=167)	Current Non-smokers (N=252)		Endoleak (N=125)	No-endoleak (N=294)	
Continuous variables – mean (sd)							
Age-yr	419	67.1(6.8)	70.8(7.9)	<.0001	70.5(8.2)	68.8(7.4)	0.032
Height-cm	419	177.8(7.3)	177.5(7.7)	0.74	177.7(7.4)	177.6(7.6)	0.90
Weight-kg	419	88.0(17.8)	92.6(16.3)	0.0063	90.2(16.6)	91.0(17.3)	0.67
Left ABI	383	1.0(0.2)	1.0(0.6)	0.61	1.1(0.8)	1.0(0.2)	0.20
Right ABI	382	1.0(0.2)	1.0(0.6)	0.75	1.1(0.9)	1.0(0.2)	0.22
Binary variables – no. (%)							
Currently smokes	419	-	-	-	39(31.2)	128(43.54)	0.018
Family history of abdominal aortic aneurysm	419	29(17.4)	36(14.3)	0.39	23(18.4)	42(14.3)	0.29
Diabetes	419	32(19.2)	57(22.6)	0.40	34(27.2)	55(18.7)	0.052
High cholesterol	419	121(72.5)	183(72.6)	0.97	93(74.4)	211(71.8)	0.58
Thrombosis	419	11(6.6)	11(4.4)	0.32	5(4.0)	17(5.78)	0.45
Hypertension	419	128(76.7)	201(79.8)	0.45	101(80.8)	228(77.6)	0.46
Emphysema	419	47(28.1)	72(28.6)	0.92	32(25.6)	87(29.6)	0.41
Coagulopathy	419	2(1.2)	3(1.2)	0.99	2(1.6)	3(1.0)	0.62
Stroke	419	30(18.0)	32(12.7)	0.14	18(14.4)	44(15.0)	0.88
Cardiac disease	419	1(0.6)	1(0.4)	0.44	72(57.6)	142(48.3)	0.16
Medication[¶]: Beta-blockers	419	111(66.5)	163(64.7)	0.71	88(70.4)	186(63.3)	0.16
Medication[¶]: Aspirin	419	89(53.3)	142(56.4)	0.54	74(59.2)	157(53.4)	0.27
Medication[¶]: ACE inhibitors	419	76(45.5)	107(42.5)	0.54	56(44.8)	127(43.2)	0.76
Medication[¶]: Anticoagulants	419	15(9.0)	27(10.7)	0.56	18(14.4)	24(8.2)	0.052
Medication[¶]: Other platelet inhibitors	337	8(6.3)	16(7.7)	0.63	11(10.5)	13(5.6)	0.11

* P value of two sample t-test (for continuous variables) or chi-square test (for binary variables).

¶ Medications last 6 months prior to randomization.

Age by Smoking Status   Age by Endoleak
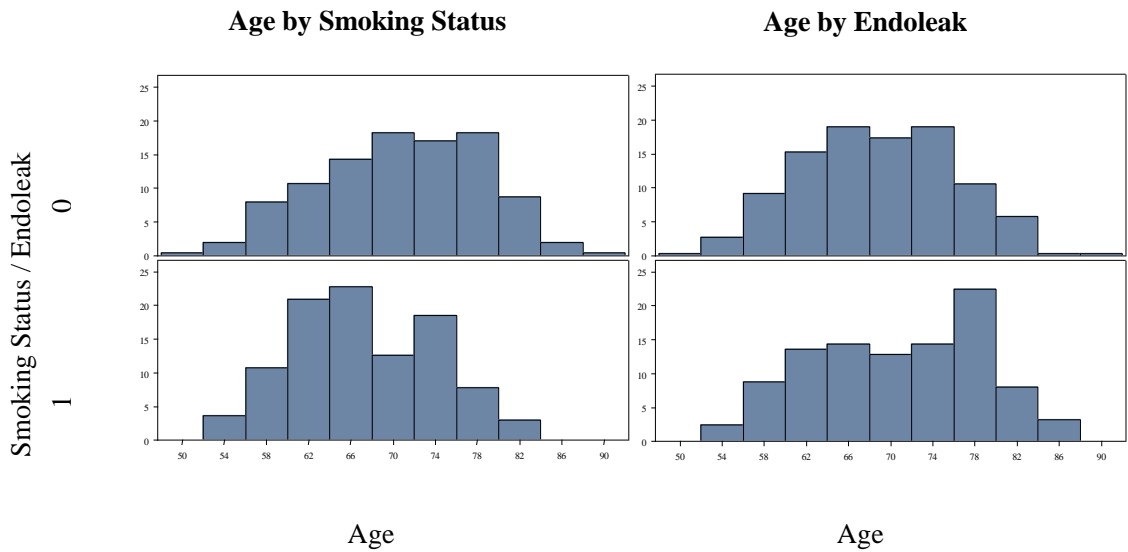
Figure 1: Histogram of age by smoking status (left) and by endoleak occurrence (right). X-axis stands for age, ranging from 50 to 90. Y-axis stands for incidence percent. The two plots on the upper side are age for current non-smokers (left) and for subjects with no endoleaks (right). The two plots on the lower side are age for current smokers (left) and for subjects with endoleaks (right). For age and smoking status, P value from two sample t-test is less than 0.0001. For age and endoleak, P value from two sample t-test equals 0.032.

Weight by Smoking Status   Weight by Endoleak
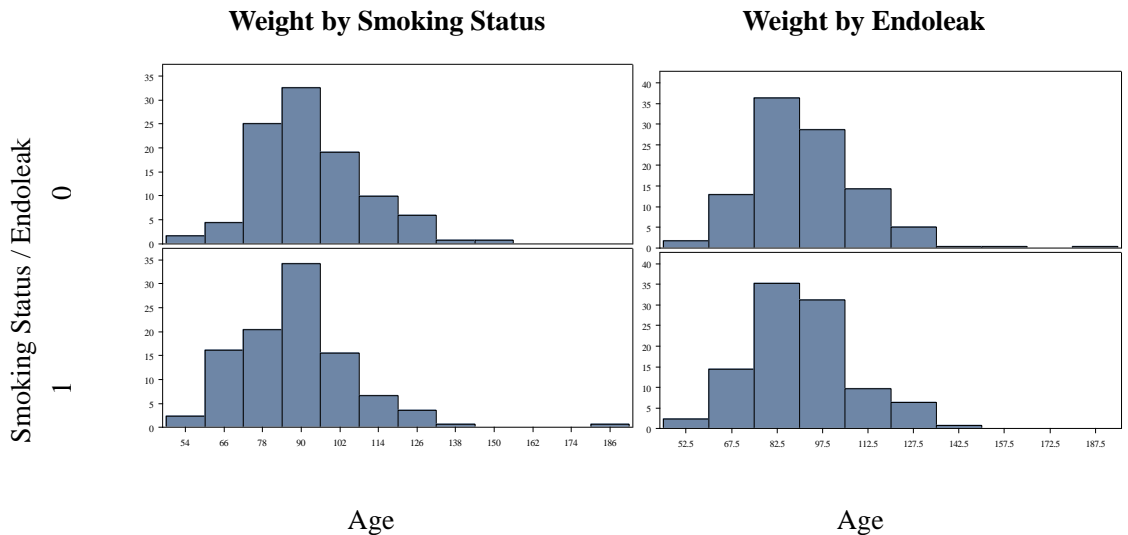
Figure 2: Histogram of weight by smoking status (left) and by endoleak occurrence (right). X-axis stands for age, ranging from 52(52.5) to 187(187.5). Y-axis stands for incidence percent. The two plots on the upper side are weight for current non-smokers (left) and for subjects with no endoleaks (right). The two plots on the lower side are weight for current smokers (left) and for subjects with endoleaks (right). For weight and smoking status, P value from two sample t-test is less than 0.0063. For age and endoleak, P value from two sample t-test equals 0.67.

Another issue in the table 1 was the missing values for three variables – left ABI 36(8.6%), right ABI 37(8.8%), and other platelet inhibitor use history 82(19.6%).   To better understand the missing pattern, baseline characteristics were compared between subjects who had missing values and subject who were observed.   The results were demonstrated in Table 2.   And all subjects who missed left ABI measurement also missed right ABI.   But only one subject with the right ABI value missing, had the variable of left ABI not missing.   Therefore, left ABI miss/not-miss in the analysis (left column) was used to represent results for left ABI and right ABI.   From Table 2, it was discovered that subjects who had ABI values missing were significantly taller (P= 0.0072) and significantly less likely to take aspirin medicines (P= 0.04).   In addition, subjects who had other platelet inhibitor medication history missing had significantly lower proportions of AAA family history and thrombosis history (P=0.0086).

Table 2: Check the balance of missing observations

Variables	Left/Right ABI*		P**	Medication: Other platelet inhibitors		P**
	Not missing (N=383/382)	Missing (N=36/37)		Not missing (N=337)	Missing (N=82)	
Continuous variables – mean (sd)						
Age	69.2(7.6)	70.9(8.0)	0.20	69.4(7.8)	68.9(7.0)	0.61
Height-cm	177.3(7.5)	180.9(7.7)	0.0072	177.6(7.3)	177.6(8.5)	0.98
Weight-kg	90.8(17.0)	90.6(18.4)	0.95	89.9(15.6)	94.0(22.0)	0.11
Left ABI	-	-	-	0.99(0.2)	1.12(1.0)	0.29
Right ABI	-	-	-	0.99(0.2)	1.11(1.1)	0.32
Binary variables – no. (%)						
Currently smokes	154(40.2)	13(36.1)	0.63	128(38.0)	39(47.6)	0.11
Endoleak	118(30.8)	7(19.4)	0.15	105(31.2)	20(24.4)	0.23
Family history of AAA	59(15.4)	6(16.7)	0.84	60(17.8)	6(6.1)	0.0086
Diabetes	83(21.7)	6(16.7)	0.48	74(22.0)	15(18.3)	0.47
High cholesterol	281(73.4)	23(63.9)	0.22	248(73.6)	56(68.3)	0.34
Thrombosis	20(5.2)	2(5.6)	0.93	22(6.5)	0(0)	0.018

Hypertension	302(78.9)	27(75.0)	0.59	266(78.9)	63(76.8)	0.68
Emphysema	106(27.7)	13(36.1)	0.28	92(27.3)	27(32.9)	0.31
Coagulopathy	5(1.3)	0(0)	0.49	5(1.5)	0(0)	0.27
Stroke	58(15.1)	4(11.1)	0.51	47(14.0)	15(18.2)	0.32
Cardiac disease	199(52.0)	15(7.0)	0.44	165(49.0)	49(60.0)	0.18
Medication[¶]: Beta-blockers	252(65.8)	22(61.1)	0.57	220(65.3)	54(65.9)	0.92
Medication[¶]: Aspirin	217(56.7)	14(39.9)	0.040	194(57.6)	37(45.1)	0.042
Medication[¶]: ACE inhibitors	171(44.7)	12(33.3)	0.19	151(44.8)	32(39.0)	0.34
Medication[¶]: Anticoagulants	40(10.4)	2(5.56)	0.35	31(9.2)	11(13.4)	0.25
Medication[¶]: Other platelet inhibitors	22(7.1)	2(6.9)	0.96	-	-	-

Left ABI had 36 missing observations and right ABI had 37 missing observations. The missing observations shared 36 subjects in common (36 subjects were missing both left and right ABI values). Calculation in this column was based on left ABI observations.

** P value of two sample t-test (for continuous variables) or chi-square test (for binary variables).

Consequently based on the evidence above it was hard to conclude that the three variables were missing randomly.   On the other hand, 419 subjects might not large enough to definitively make a precise conclusion.   Under this circumstance, four datasets were constructed and used for further analysis: (i) one with all variables but incomplete subjects; (ii) one with all subjects and variables with no missing value; (iii) one filled with logistic regression; and (iv) one filled with multiple imputation.

Logistic regression and diagnostics

Logistic regression models were built with different variable selection techniques (no selection, forward selection, backward selection, and stepwise selection).   Criteria (significance level) for variable entry in forward and stepwise selection was set at 0.05, significance level for variable elimination in backward and

stepwise selection was set at 0.10.   As suggested in the section above, analysis was conducted with three

datasets.   Part of the results including P value, OR estimate, and confidence interval of OR estimate of

current smokes generated by logistic regression models were listed in Table 3.   An expanded result table

containing the final model information can be found in the Appendix.   In addition, Figure 3 was generated

to better illustrate the odds ratio estimation and confidence intervals.   From Figure 3, one can deduce that

odds ratio for current smoking was around 0.6, although more than half (7/12) of the   confidence interval

contained the value of 1.   Of the three datasets, the one with all variables had two insignificant P values,

the one with all subjects but partial variable list had 3 out of 4 P values significant, and the one filled with

multiple imputation data had all P value significant.

Table 3: Logistic regression results.

Include all variables* (308 subjects)				
Variable selection method[**]	**Currently smokes**			
	P	**OR estimate**	**95% Wald Confidence Interval**	
No selection	0.084	0.61	0.34	1.07
Forward	0.022	0.55	0.33	0.92
Backward	0.12	0.65	0.38	1.11
Stepwise	0.022	0.55	0.33	0.92
Include only no-missing variables[¶] (419 subjects)				
Variable selection method[**]	**Currently smokes**			
	P	**OR estimate**	**95% Wald Confidence Interval**	
No selection	0.074	0.65	0.40	1.04
Forward	0.019	0.59	0.38	0.92
Backward	0.025	0.60	0.38	0.94
Stepwise	0.019	0.59	0.38	0.92
Data filled with multiple imputation (all variables, 419 subjects)				
Variable selection method[**]	**Currently smokes**			
	P	**OR estimate**	**95% Wald Confidence Interval**	
No selection	0.060	0.63	0.38	1.02
Forward	0.063	0.64	0.40	1.03
Backward	0.063	0.64	0.40	1.02

Stepwise	0.063	0.64	0.40	1.03

*All variables = Current smoking, age, height, weight, AAA family history, diabetes, high cholesterol, thrombosis, emphysema, coagulopathy, stroke, cardiac disease history, beta blocker, aspirin, ACE inhibitor, anticoagulants, platelet inhibitor, left ABI, and right ABI.
**Significance criteria: SLENTRY=0.05, SLSTAY=0.10.
¶All variables except left ABI, right ABI, and platelet inhibitor.
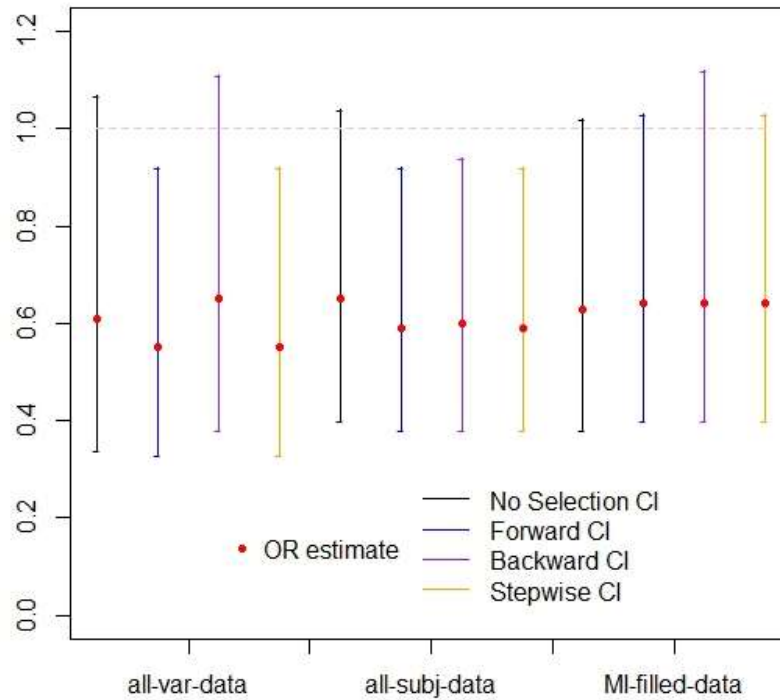
Figure 3: *OR estimate and confidence interval by logistic regressions. Left four lines stand for OR confidence intervals calculated with data including all variables. Four lines in the middle stand for OR confidence intervals calculated with data including all subjects. And right four lines stand for OR confidence intervals calculated with data filled by multiple imputation.*

For all the logistic regression models, the Hosmer and Lemeshow goodness-of-fit test was used to assess the overall fitness of the models. Predicted probability diagnostics plot, leverage diagnostics plot, and influence on the model fit and parameter estimates plot were used to identify the outliers and high-influenced plots. Results of goodness-of-fit test can be found in Table 4. All the logistic models listed in the table had P value greater than 0.05, indicating that none of the model fit was significantly bad. The predicated probability diagnostics plots, leverage diagnostics plots, and influence on the model fit and

parameter estimation plots were provided in the Appendix.

Table 4: Logistic regression model diagnostic (Hosmer and Lemeshow Goodness-of-fit test)

Model selection method	Model Detail	Hosmer and Lemeshow Goodness-of-fit test		
		Chi-square	DF	Pr>ChiSq
Include all variables* (308 subjects)				
No selection	Logit(endoleak) = Current Smokes + All variables	13.31	8	0.10
Forward	Logit(endoleak) = Current Smokes + Beta Blocker	0.076	2	0.96
Backward	Logit(endoleak) = Current Smokes + Age + Aspirin + Anticoag + PlateletInhib	12.07	8	0.14
Stepwise	Logit(endoleak) = Current Smokes + Beta Blocker	0.076	2	0.96
Include only no-missing variables¶ (419 subjects)				
No selection	Logit(endoleak) = Current Smokes + All variables	9.48	8	0.30
Forward	Logit(endoleak) = Current Smokes	-	-	-
Backward	Logit(endoleak) = Current Smokes + Diabetes + Anticoag	0.22	3	0.97
Stepwise	Logit(endoleak) = Current Smokes	-	-	-
Data filled with multiple imputation (all variables, 419 subjects)				
No selection	Logit(endoleak) = Current Smokes + All variables	12.14	8	0.15
Forward	Logit(endoleak) = Current Smokes + Age + Diabetes + Right_ABI + PlateletInhib	6.62	8	0.58
Backward	Logit(endoleak) = Current Smokes + Age + Family History + Diabetes + Anticoagulants + Right_ABI + PlateletInhib	9.90	8	0.27
Stepwise	Logit(endoleak) = Current Smokes + Age + Diabetes + Right_ABI + PlateletInhib	6.62	8	0.58

Propensity score analysis and diagnostics

The first step of implementing propensity score methodology was to generate propensity scores. In the

logistic regression models that calculated propensity score, all the available baseline effects were included

and variable lists varied by data set. Figure 4 showed the histograms of propensity scores by smoking

status in different datasets, one could see that propensity scores in all data sets had significantly different

distributions (P values of two sample t-test were less than .0001 in all cases).   Although the distributions of

propensity score were different, the ranges of distributions for smokers and non-smokers almost overlapped.

In this situation, matching using propensity score can allow for reasonable matched pairs and can also

improve the results by matching cases to similar controls.   In addition, covariate adjustment and

stratification with propensity score may also worked well because of the comparability of case and control

groups.   Therefore, it appeared that the propensity score method addressed the problem well.

Include all variables (308 subjects)　　　　**Include only no-missing variables (419 subjects)**
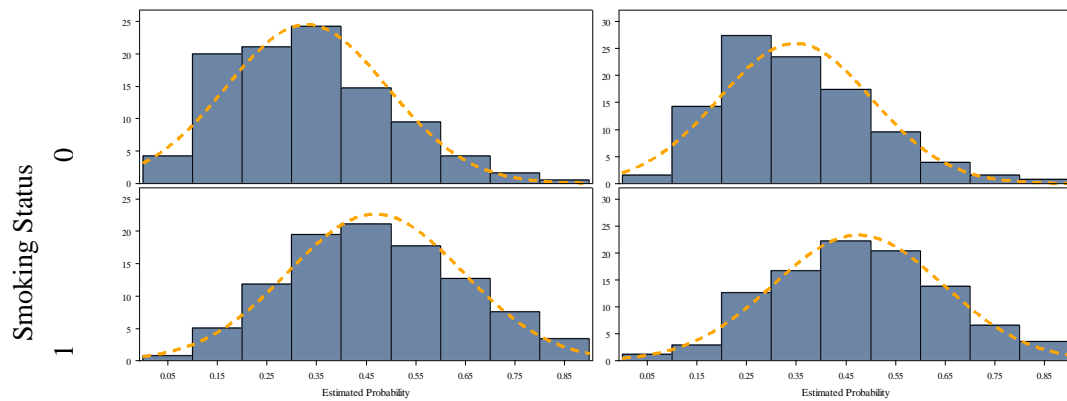
Data filled by multiple imputation (all variables, all subjects)
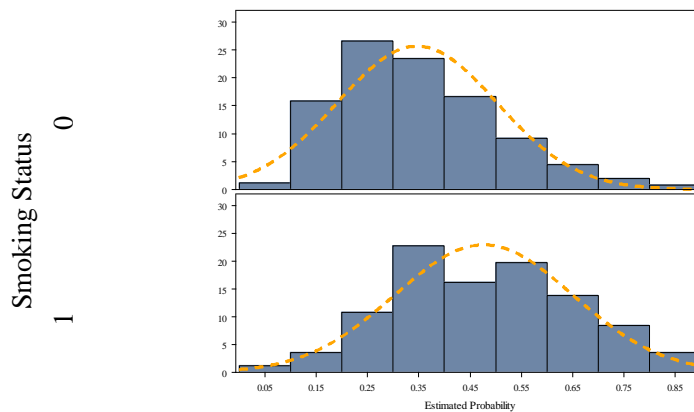
Figure 4: Histograms of propensity score by smoking status for data set including all variables (top left), data set including all subjects and only no-missing variables (top right), data set filled by multiple imputation (bottom).   X-axis was estimated probability (propensity score estimate).   Y-axis is frequency. Yellow dotted lines were normal approximation based on data.   P values of two sample t-test for all three situations are less than .0001.

The results by multiple propensity score methods were presented in Table 5.   First, it was noticed that

the numbers of matched pairs were pretty satisfactory, especially in the second and third data sets. The

first dataset included all variables but only 308 subjects (118 current smokers and 190 current

non-smokers), which was less than the second and third data sets, and thus less matched pairs.   For the

second dataset, no less than 130 pairs were matched, that was 68 percent of all cases or 80 percent of the

all the controls.   Second, from the perspective of results, all confidence intervals of odds ratios of current

smoking included 1 in the range except in three cases.   To understand the results better, the results of

odds ratio estimation and confidence intervals were presented in Figure 5.   Compared with Figure 3, the

odds ratio estimation demonstrated in Figure 5 had a larger range (0.47 to 0.77).   In addition, larger

proportions of confidence intervals included 1 in Figure 5 compared to in Figure 3.   To conclude, more

evidence pointed towards the suggestion that smoking was not significantly related with endoleak

occurrence after propensity score methods were used.

Table 5: Propensity score models results.

Include all variables* (308 subjects)					
	Currently smokes				
Methods	**Matched pairs**	**P**	**OR estimate**	**95% Wald Confidence Interval**	
Covariate adjustment	-	0.090	0.62	0.36	1.08
Stratification	-	0.092	0.62	0.36	1.08
Greedy 1 to 1 0.1 logit ps.sd matching	95	0.062	0.53	0.27	1.03
Greedy 1 to 1 0.2 logit ps.sd matching	97	0.040	0.49	0.25	0.97
Greedy 1 to 1 0.5 logit ps.sd matching	101	0.030	0.47	0.24	0.93
Mahalanobis matching	99	0.25	0.69	0.37	1.30
Include only no-missing variables¶ (419 subjects)					
	Currently smokes				
Methods	**Matched pairs**	**P**	**OR estimate**	**95% Wald Confidence Interval**	
Covariate adjustment	-	0.079	0.66	0.41	1.05
Stratification	-	0.058	0.63	0.39	1.02
Greedy 1 to 1 0.1 logit ps.sd matching	130	0.30	0.72	0.38	1.34

Greedy 1 to 1 0.2 logit ps.sd matching	133	0.50	0.82	0.46	1.45
Greedy 1 to 1 0.5 logit ps.sd matching	141	0.36	0.77	0.44	1.35
Mahalanobis matching	136	0.19	0.70	0.42	1.19

Data filled with multiple imputation (all variables, 419 subjects)

Methods	Currently smokes				
	Matched pairs	P	OR estimate	95% Wald Confidence Interval	
Covariate adjustment	-	0.083	0.66	0.41	1.06
Stratification	-	0.076	0.65	0.41	1.05
Greedy 1 to 1 0.1 logit ps.sd matching	129	0.091	0.59	0.32	1.09
Greedy 1 to 1 0.2 logit ps.sd matching	131	0.11	0.63	0.35	1.12
Greedy 1 to 1 0.5 logit ps.sd matching	141	0.04	0.55	0.31	0.98
Mahalanobis matching	99	0.26	0.70	0.37	1.31

*All variables = Currently smokes, age, height, weight, family history, diabetes, high cholesterol, thrombosis, emphysema, coagulopathy, stroke, cardiac disease history, beta blocker, aspirin, ACE inhibitor, anticoagulants, platelet inhibitor, left ABI, and right ABI.

¶All variables except left ABI, right ABI, and platelet inhibitor.
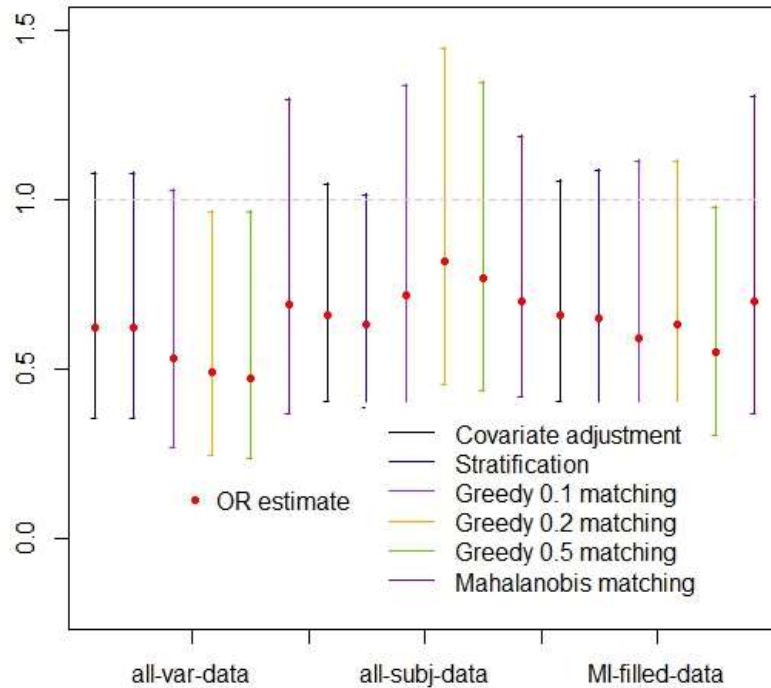

Figure 5: *OR estimate and confidence interval by propensity score methods. Left six lines stand for OR confidence intervals calculated with data including all variables. Six lines in the middle stand for OR confidence intervals calculated with data including all subjects. And right six lines stand for OR confidence intervals calculated with data filled by multiple imputation.*

To make sure that propensity score matching did make the case and control groups more balanced,

standardized differences of each variable before and after matching were checked in all models.

Table 6 listed the standardized differences of all variables before and after matching in greedy 1:1

matching with 0.1 logit propensity score standard deviation using data including all variables.

Standardized difference equal to 0.1 usually indicated a big unbalance between the two groups.

Before matching, age (sdd=0.51), weight (sdd=0.28), right ABI (sdd=0.11), stroke (sdd=0.12), and

medication-beta blocker (sdd=0.15) were all considered unbalanced variables. Age was the most

unbalanced variable and had standardized difference 0.51 which demonstrated a substantial

unbalance. After matching, almost all variables were balanced except for right ABI (sdd=0.15) and

cardiac disease (sdd=0.15), which were just mildly unbalanced. And in the following analysis, right

ABI and cardiac disease were added to the model to ensure the minimum bias caused by unbalanced

variables.

Table 6: Check balance before and after matching with propensity score: standardized difference.
(Use Greedy 1 to 1 0.1 logit ps.sd matching applying to data including all variables as an example)

Variables	Current smokers vs. Current non-smokers	
	By smoking status	By endoleak occurrence
Continuous variables		
Age-yr	0.51	0.076
Height-cm	0.031	0.023
Weight-kg	0.28	0.023
Left ABI	0.084	0.002
Right ABI	0.11	0.15
Binary variables		
Family history of abdominal aortic aneurysm	0.083	0.056
Diabetes	0.068	0.025
High cholesterol	0.008	0.048
Thrombosis	0.074	0.000

Hypertension	0.029	0.029
Emphysema	0.006	0.071
Coagulopathy	0.009	0.085
Stroke	0.12	0.056
Cardiac disease	0.069	0.15
Medication[¶]: Beta-blockers	0.15	0.045
Medication[¶]: Aspirin	0.059	0.000
Medication[¶]: ACE inhibitors	0.028	0.021
Medication[¶]: Anticoagulants	0.053	0.076
Medication[¶]: Other platelet inhibitors	0.023	0.039

Because of the differences between logistic regression results (Table 3) and propensity score results (Table 5), one would question whether the data matched by propensity score would still generate the same results if logistic regression models were applied.   Following this hypothesis, logistic regressions were used in the propensity-score-matching to generate datasets and the results.   The results presented in Table 7 indicated that more results were significant in some datasets, especially the data sets including all variables.   However, in datasets including all subjects and datasets filled with multiple imputations, all the results were not significant.

Table 7: Logistic regression with propensity-score-matching generated datasets

Include all variables*

			Currently smokes			
Using dataset generated by	**Method**	**Matched Pair**	**P**	**OR estimate**	**95% Wald Confidence Interval**	
Greedy 1 to 1 0.1 logit ps.sd matching	No selection	95	0.032	0.46	0.23	0.94
	Forward	95	0.049	0.52	0.27	1.00
	Backward	95	0.053	0.52	0.27	1.01
	Stepwise	95	0.049	0.52	0.27	1.00
Greedy 1 to 1 0.2 logit ps.sd matching	No selection	97	0.028	0.46	0.23	0.92
	Forward	97	0.048	0.51	0.26	1.00
	Backward	97	0.053	0.52	0.26	1.01

	Stepwise	97	0.048	0.51	0.26	1.00
Greedy 1 to 1 0.5 logit ps.sd matching	No selection	101	0.023	0.49	0.22	0.90
	Forward	101	0.041	0.50	0.26	0.97
	Backward	101	0.041	0.50	0.26	0.97
	Stepwise	101	0.041	0.50	0.26	0.97
Mahalanobis matching with 0.2 logit ps.sd	No selection	99	0.28	0.69	0.35	1.36
	Forward	99	0.34	0.73	0.39	1.39
	Backward	99	0.38	0.75	0.39	1.43
	Stepwise	99	0.34	0.73	0.39	1.39

Include only no-missing variables[¶]

Using dataset generated by	Method	Matched pairs	P	OR estimate	95% Wald Confidence Interval	
					Currently smokes	
Greedy 1 to 1 0.1 logit ps.sd matching	No selection	130	0.20	0.69	0.39	1.22
	Forward	130	0.17	0.68	0.39	1.19
	Backward	130	0.15	0.66	0.38	1.16
	Stepwise	130	0.17	0.68	0.39	1.19
Greedy 1 to 1 0.2 logit ps.sd matching	No selection	133	0.43	0.80	0.46	1.39
	Forward	133	0.38	0.79	0.46	1.34
	Backward	133	0.39	0.79	0.46	1.36
	Stepwise	133	0.38	0.79	0.46	1.34
Greedy 1 to 1 0.5 logit ps.sd matching	No selection	141	0.20	0.70	0.41	1.21
	Forward	141	0.16	0.69	0.40	1.16
	Backward	141	0.16	0.69	0.40	1.16
	Stepwise	141	0.16	0.69	0.40	1.16
Mahalanobis matching with 0.2 logit ps.sd	No selection	136	0.14	0.65	0.37	1.15
	Forward	136	0.18	0.68	0.39	1.19
	Backward	136	0.18	0.68	0.39	1.19
	Stepwise	136	0.18	0.68	0.39	1.19

Data filled with multiple imputation (all variables, all subjects)

Using dataset generated by	Method	Matched pairs	P	OR estimate	95% Wald Confidence Interval	
					Currently smokes	
Greedy 1 to 1 0.1 logit ps.sd matching	No selection	128	0.10	0.61	0.34	1.10
	Forward	128	0.15	0.66	0.38	1.16
	Backward	128	0.15	0.66	0.38	1.16
	Stepwise	128	0.15	0.66	0.38	1.16
Greedy 1 to 1 0.2 logit ps.sd matching	No selection	131	0.11	0.62	0.35	1.11
	Forward	131	0.12	0.64	0.37	1.12
	Backward	131	0.14	0.66	0.38	1.15
	Stepwise	131	0.12	0.64	0.37	1.12
Greedy 1 to 1 0.5 logit ps.sd matching	No selection	141	0.078	0.61	0.35	1.06

	Forward	141	0.098	0.64	0.37	1.09
	Backward	141	0.098	0.64	0.37	1.09
	Stepwise	141	0.098	0.64	0.37	1.09
Mahalanobis matching with 0.2 logit ps.sd	No selection	99	0.37	0.73	0.37	1.44
	Forward	99	0.44	0.77	0.41	1.48
	Backward	99	0.46	0.78	0.41	1.50
	Stepwise	99	0.44	0.77	0.41	1.48

Simulation

A total of one hundred datasets had been generated by multiple imputation and the following simulation

process.   Then, these data sets were analyzed by logistic regression and propensity score method as

described above.   The results were collected and summarized in Tables 8 and 9.   First, the p value

estimations and OR estimations in the two tables were stable over different methods, except for the p

value estimation of propensity score method with mahalanobis matching.   All the means of odds ratio

estimations of developing endoleaks in smoking subjects versus non-smoking subjects remained

between 0.75 and 0.80.

Further, neither the results of p value nor the results of odds ratio were significant.   All the confidence

intervals of P values contained 0.05 and all the confidence intervals of odds ratio included 1.   Therefore,

we failed to conclude that smoking was a significant factor.   Last but not the least, although 1 was

included in the confidence intervals of odds ratio estimates, all the odds ratio estimations were smaller

than 1, which provided evidence to support that smoking had a protective effect on endoleak incidence.

In summary, simulations indicated that smokers may have lower risk of developing endoleak compared

with non-smokers, although this effect was not significant in this analysis.   Larger sample size might be

necessary for further confirming this finding.

Table 8: Simulation: logistics regression

Variable selection method	Currently smokes				
	P		OR		
	Mean	95% Confidence Interval	Estimate mean	95% Confidence Interval	95% Prediction Interval
No selection	0.32	(0.004, 0.96)	0.77	(0.5, 1.16)	(0.32, 1.86)
Forward	0.28	(0.002, 0.85)	0.77	(0.51, 1.18)	(0.34, 1.80)
Backward	0.30	(0.003, 0.90)	0.78	(0.52, 1.18)	(0.33, 1.80)
Stepwise	0.29	(0.002, 0.85)	0.77	(0.51, 1.17)	(0.34, 1.80)

Table 9: Simulation: propensity score

Methods	Currently smokes				
	P		OR		
	Mean	95% Confidence Interval	Estimate mean	95% Confidence Interval	95% Prediction Interval
Covariate adjustment	0.32	(0.005, 0.91)	0.80	(0.52, 1.15)	(0.33, 1.81)
Stratification	0.33	(0.005, 0.90)	0.79	(0.53, 1.15)	(0.34, 1.80)
Greedy 1 to 1 0.1 logit ps.sd matching	0.32	(0.003, 0.88)	0.76	(0.44, 1.20)	(0.25, 2.0)
Greedy 1 to 1 0.2 logit ps.sd matching	0.31	(0.003, 0.91)	0.78	(0.46, 1.23)	(0.26, 2.04)
Greedy 1 to 1 0.5 logit ps.sd matching	0.30	(0.0036, 0.93)	0.77	(0.45, 1.20)	(0.26, 2.02)
Mahalanobis matching	0.09	(0, 0.64)	0.75	(0.46, 1.13)	(0.39, 1.42)

DISCUSSION

Several studies have investigated the risk factors for endoleak in recent years and had made significant

progress[11,15-18]. Although not every studies included smoking as a potential risk factor, all studies that

had included smoking in the analysis found that the proportions of previous or current smokers who had

developed an endoleak were smaller[11,15-17]. The methods used in these studies were either multivariate

(logistic) regression techniques or chi-square tests. In order to address the finding that smokers may

have lower incidence of endoleaks, a few explanations were given in the above studies: Buth et al.

suggested that changes in the coagulation profile of the blood caused by smoking may lead to a tendency

of quick occlusion of small vessels, which caused the spontaneous heal of endoleaks[15,17]. The analysis

described in this project confirmed the conclusion that the odds of developing endoleak for current

smokers were smaller than the odds of developing endoleak for current non-smokers even if the

confidence intervals of odds ratio were not statistically significant. Compared to the previous studies,

this study used more sophisticated statistical approaches/analytical techniques and utilized more

relavantcovariates in the analysis.

The propensity score method was chosen to analyze the data due to its ability to balance the dissimilar

characteristics between case and control groups. In reality, smokers and non-smokers usually had

systematically different characteristics, such as age and cardiac disease history. As in the case of the VA

trial's endoleak dataset, it had been discovered in the preliminary analysis that current non-smokers were

older and heavier than current smokers, which might cause bias to the following analysis. After

matching or stratifying by propensity score, the difference between characteristics of smokers and

non-smokers became small. The second positive aspect about propensity score was that all suspected

covariates can be contained in the model, which was in contrast with the parsimonious variable selection

in logistic regressions. Including more variables in the model was able to further minimize the bias and

improve fit of the model, and the VA endoleak data set included more than 20 variables. Lastly, but not

the least, propensity score method fitted the scenario of the question very well. Hernan and Taubman

argued that it was important to have a well-defined intervention when raising questions about the

investigation of causal relationships[26]. In the current study, the main interest was the presence or not of a

causal relationship between smoking and risk of endoleak development. A randomized trial with smoking

cessation as intervention to reduce the incidence of endoleaks in this patient population may be difficult,

although not impossible, to implement.   Therefore, all subjects with a smoking history were included and propensity score was used to itimate the endoleak study with an ideal randomized experiment trial.

However, propensity score method had its own limitations.   First, as suggested by Rubin and Rosenbaum, propensity score can only balance the observed covariates[19].   If the study was biased by unmeasured or hidden covariates, the propensity-score-generated results would not be corrected.   Second, not all observations were used in the analysis.   One might observe the differences between the total number of subjects in the data set and the number of subjects left after matching.   This partial-use of information may also cause bias.

The pros and cons of logistic regression and propensity score method had been discussed for a few years, and no consensus had been reached.   Two recent literature review studies concluded that treatment effects estimated by propensity score methods and regression techniques were similar to each other, while another simulation study found that propensity score method provided systematically better estimation for treatment effect compared with logistics regression[27,28].   In the current study, the results of propensity score and logistic regression were similar.   Although they both generated odds ratio estimations less than 1, one may observe that propensity score had less significant p value estimations and wider confidence interval for odds ratio.   Two reasons may provide possible explanations: (a)   The sample size was reduced after matching by propensity score and such reduction increased the estimated standard deviation of odds ratio estimations and widened the confidence intervals; (b) The propensity score indeed decreased the bias and made the estimations shift towards null effects.   Larger sample sizes were needed to decide which one(s) was the reason(s) for the differences.

In addition to the potential drawbacks of the propensity score methodology, this study had other limitations. For instance, all the subjects in the study were male. This may restrict the finding to only male patients. Another limitation was the missing data. Although multiple imputation techniques and simulation were used to decrease the impact of the missing data, the missing pattern was impossible to detect through the existing data and negative influence may still exist.

Smoking cessation had almost always been found as a beneficial intervention towards improving cardiac health, especially for patients after cardiac surgeries[31-34]. It has been reported that patients who had quitted smoking experienced a significant reduction in sudden cardiac death[27] and had a significant decrease in mortality after myocardial infarction[30]. However, the pros and cons of smoking for patients after endovascular repair needed further examination and discussion.

REFERENCE

1. Greenhalgh, Roger M., et al. "Endovascular repair of aortic aneurysm in patients physically ineligible for open repair." The New England journal of medicine362.20 (2010): 1872-1880.

2. Volodos, N. L., et al. "A self-fixing synthetic blood vessel endoprosthesis."Vestnik khirurgii imeni II Grekova 137.11 (1986): 123.

3. Lederle, Frank A., et al. "Outcomes following endovascular vs open repair of abdominal aortic aneurysm: a randomized trial." Jama 302.14 (2009): 1535-1542.

4. Lederle, Frank A., et al. "Long-term comparison of endovascular and open repair of abdominal aortic aneurysm." New England Journal of Medicine 367.21 (2012): 1988-1997.

5. Prinssen, Monique, et al. "A randomized trial comparing conventional and endovascular repair of abdominal aortic aneurysms." New England Journal of Medicine 351.16 (2004): 1607-1618.

6. White, Geoffrey H., et al. "Endoleak as a complication of endoluminal grafting of abdominal aortic aneurysms: classification, incidence, diagnosis, and management." (2009).

6. Brown, L. C., et al. "The UK Endovascular Aneurysm Repair (EVAR) trials: design, methodology and progress." European journal of vascular and endovascular surgery 27.4 (2004): 372-381.

7. Blum, Ulrich, et al. "Endoluminal stent–grafts for infrarenal abdominal aortic aneurysms." *New England Journal of Medicine* 336.1 (1997): 13-20.

8. Gilling-Smith, G. L., et al. "Freedom from endoleak after endovascular aneurysm repair does not equal treatment success." *European Journal of Vascular and Endovascular Surgery* 19.4 (2000): 421-425.

9. Chuter, Timothy AM, et al. "Endoleak after endovascular repair of abdominal aortic aneurysm." *Journal of vascular surgery* 34.1 (2001): 98-105.

10. van Marrewijk, Corine, et al. "Significance of endoleaks after endovascular repair of abdominal aortic aneurysms: the EUROSTAR experience." *Journal of vascular surgery* 35.3 (2002): 461-473.

11. Greenhalgh, R. M. "Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (EVAR trial 1), 30-day operative mortality results: randomised controlled trial." *The Lancet* 364.9437 (2004): 843-848.

12. Mialhe, Claude, Claude Amicabile, and Jean Pierre Becquemin. "Endovascular treatment of infrarenal abdominal aneurysms by the Stentor system: preliminary results of 79 cases." *Journal of vascular surgery* 26.2 (1997): 199-209.

13. Corriere, Matthew A., et al. "Endoleak following endovascular abdominal aortic aneurysm repair: implications for duration of screening." *Annals of surgery* 239.6 (2004): 800.

14. Mohan, I. V., R. J. F. Laheij, and P. L. Harris. "Risk factors for endoleak and the evidence for stent-graft oversizing in patients undergoing endovascular aneurysm repair." *European Journal of Vascular and Endovascular Surgery* 21.4 (2001): 344-349.

15. Zarins, Christopher K., et al. "Endoleak as a predictor of outcome after endovascular aneurysm repair: AneuRx multicenter clinical trial." *Journal of vascular surgery* 32.1 (2000): 90-107.

16. Buth, Jacob, et al. "The significance and management of different types of endoleaks." *Seminars in vascular surgery*. Vol. 16. No. 2. WB Saunders, 2003.

17. Green, N., et al. "Endoleak after endovascular aneurysm repair: Current status." *Reviews in Vascular Medicine* (2013).

18. Rosenbaum, Paul R., and Donald B. Rubin. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70.1 (1983): 41-55.

19. Austin, Peter C. "Some Methods of Propensity‐Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo simulations." *Biometrical Journal* 51.1 (2009): 171-184.

20. Austin, Peter C., Paul Grootendorst, and Geoffrey M. Anderson. "A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study." *Statistics in medicine* 26.4 (2007): 734-753.

21. COCHRAN, W. G. (1965). The planning of observational studies of human populations (with discussion). *J. R. Statist. Soc. A* 128, 234-55

22. d'Agostino, Ralph B. "Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group." *Stat Med* 17.19 (1998): 2265-2281.

23. Yuan, Yang C. "Multiple imputation for missing data: Concepts and new development (Version 9.0)." SAS Institute Inc, Rockville, MD (2010).

24. Rubin, Donald B. Multiple imputation for nonresponse in surveys. Vol. 307. John Wiley & Sons, 2009.

26. Hernán, M. A., and S. L. Taubman. "Does obesity shorten life? The importance of well-defined interventions to answer causal questions."International Journal of Obesity 32 (2008): S8-S14.

25. Shah, Baiju R., et al. "Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review."Journal of clinical epidemiology 58.6 (2005): 550-559.

26. Stürmer, Til, et al. "A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods." Journal of clinical epidemiology 59.5 (2006): 437-e1.

27. Goldenberg I, Jonas M, Tenenbaum A, et al. Current Smoking, Smoking Cessation, and the Risk of Sudden Cardiac Death in Patients With Coronary Artery Disease. Arch Intern Med. 2003;163(19):2301-2305. doi:10.1001/archinte.163.19.2301.

28. Wilson K, Gibson N, Willan A, Cook D. Effect of Smoking Cessation on Mortality After Myocardial Infarction: Meta-analysis of Cohort Studies. Arch Intern Med.2000;160(7):939-944. doi:10.1001/archinte.160.7.939.

29. Voors, Adriaan A., et al. "Smoking and Cardiac Events After Venous Coronary Bypass Surgery A 15-Year Follow-up Study." Circulation 93.1 (1996): 42-47.

30. Irving, David W., and T. Yamamoto. "Cigarette smoking and cardiac output."British heart journal 25.1 (1963): 126.

31. Kannel, William B., Jonathan F. Plehn, and L. Adrienne Cupples. "Cardiac failure and sudden death in the Framingham Study." American heart journal115.4 .(1988): 869-875

32. Hallstrom, Alfred P., Leonard A. Cobb, and Roberta Ray. "Smoking as a risk factor for recurrence of sudden cardiac arrest." New England Journal of Medicine 314.5 (1986): 271-275.

Appendix 1: Logistic regression (with whole model column).

Include all variables* (308 subjects)

Variable selection method[**]	Currently smokes				Whole model
	P	OR estimate	95% Wald Confidence Interval		
No selection	0.084	0.61	0.34	1.07	Logit(endoleak) = Current Smokes + All variables
Forward	0.022	0.55	0.33	0.92	Logit(endoleak) = Current Smokes + Beta Blocker
Backward	0.12	0.65	0.38	1.11	Logit(endoleak) = Current Smokes + Age + Aspirin + Anticoag + PlateletInhib
Stepwise	0.022	0.55	0.33	0.92	Logit(endoleak) = Current Smokes + Beta Blocker

Include only no-missing variables[¶] (419 subjects)

Variable selection method**	Currently smokes				Whole model
	P	OR estimate	95% Wald Confidence Interval		
No selection	0.074	0.65	0.40	1.04	Logit(endoleak) = Current Smokes + All variables
Forward	0.019	0.59	0.38	0.92	Logit(endoleak) = Current Smokes
Backward	0.025	0.60	0.38	0.94	Logit(endoleak) = Current Smokes + Diabetes + Anticoag
Stepwise	0.019	0.59	0.38	0.92	Logit(endoleak) = Current Smokes

Data filled with multiple imputation (all variables, 419 subjects)

Variable selection method**	Currently smokes				Whole model
	P	OR estimate	95% Wald Confidence Interval		
No selection	0.0601	0.63	0.38	1.02	Logit(endoleak) = Current Smokes + All variables
Forward	0.0632	0.64	0.40	1.03	Logit(endoleak) = Current Smokes + Age + Diabetes + Right_ABI + PlateletInhib
Backward	0.063	0.64	0.40	1.02	Logit(endoleak) = Current Smokes + Age + Family History + Diabetes + Anticoagulants + Right_ABI + PlateletInhib
Stepwise	0.063	0.64	0.40	1.03	Logit(endoleak) = Current Smokes + Age + Diabetes + Right_ABI + PlateletInhib

*All variables = Currently smokes, age, height, weight, family history, diabetes, high cholesterol, thrombosis, emphysema, coagulopathy, stroke, cardiac disease history, beta blocker, aspirin, ACE inhibitor, anticoagulants, platelet inhibitor, left ABI, and right ABI.

**Significance criteria: SLENTRY=0.05, SLSTAY=0.10.

[¶]All variables except left ABI, right ABI, and platelet inhibitor.