Identifying the Potential Risk Factors of Venous Thromboembolism Events in Clinical Trials

Kao-Tai Tsai

Celgene Corporation, NJ & JPHCOPH, Georgia Southern University, GA

Abstract

Analyzing safety data and trying to identify the potential risk factors of an adverse event is a both interesting and extremely challenging task. Due to the correlations and possible interactions of patient's medical history, medications received before and after the trial, as well as other factors, it can make this a complex multivariate data analysis problem and the simple summary can be very misleading. In addition, subjects with or without the safety event are usually not readily comparable in most of the randomized trials. In this article, as a case study, we show the exploratory data analysis performed in identifying the risk factors for Venous Thrombus Embolism using data analytical techniques such as matching, mediation, and regressions. Data from recent clinical trials are used to demonstrate the statistical procedures utilized.

Keywords: Drug Safety Data, Venous Thromboembolism, Mediation Effect, Matching.

1 Introduction

Analyzing safety data and trying to identify the potential risk factors of an adverse event is a both interesting and extremely challenging task. Due to the inter-correlations and possible interactions of patient's medical history, medications received before and after the trial, as well as other confounding factors, it can make this a complex multivariate data analysis problem and the summaries of commonly methods can be very misleading. In addition, subjects with or without the safety event are usually not readily comparable in most of the randomized trials.

In this article, as a case study, we show the exploratory data analysis performed in identifying the risk factors for Venous Thrombus Embolism (VTE) using data analytical techniques such as matching, recursive partitioning, and regressions.

Data from recent clinical trials are used to demonstrate the statistical procedures utilized.

We start with some background knowledge of VTE and the related risk factors identified by various registry studies. We then, using the recent clinical trial data, show the results from the straightforward regressions and discuss the pitfalls and the potential remedy using other statistical methods. The methods used in this data analysis are not new in the literature, however, it had not commonly been used in clinical trial data analysis. We conclude the article with some general discussions.

2 Economic burden of VTE

The Surgeon General estimates that 100,000 to 180,000 deaths occur annually because of VTE in the U.S. alone. VTE and its complications place a substantial burden on the U.S. health care system. In addition, the sequelae of post-thrombotic syndrome and chronic thromboembolic pulmonary hypertension are sources of morbidity, diminished quality of life, and loss in functional status. However, many of the VTE risk factors are likely modifiable by adopting a heart healthy lifestyle.

3 Studies of VTE and potential risk factors

Since VTE is usually not a frequent event like catching a cold and takes time to develop, registries provide an excellent starting point to study risk factors for VTE (see Goldhaber, S. [1] and American Society of Clinical Oncology Guideline [2]).

For example, ICOPER (International Cooperative Pulmonary Embolism Registry), the largest registry that enrolled only patients with PE, had enrolled more than 2,500 subjects over the span of the study period. They found that mortality rate at 3 months after the development of VTE was about 15%. The risk factors associated with an increased likelihood of death include age > 75 years, cancer, congestive heart failure, and chronic obstructive pulmonary disease (COPD). Another registry, RIETE (Registro Informatizado de la Enfermedad Trombo-Embolica venosa), is the largest registry with more than 15,500 VTE patients. The study indicated the clinical risk factors predicting death including immobilization for neurological disease, age > 75 years, and cancer. In addition, the Longitudinal Investigation of Thromboembolism Etiology enrolled 21,680 subjects. The age-standardized incidence of first-time VTE was 1.92/1,000 person-years and the 28-day case-fatality rate was 11% after a first VTE episode, 25% for cancer patients.

4 Risk Factors of VTE

General risk factors of VTE in community identified among the registries and clinical trials include advancing age, cancers, prior VTE, venous insufficiency, pregnancy, trauma, frailty and immobility.

Hospitalization also has the associated risk for VTE. The American College of Chest Physicians had a guideline to ask every hospital to develop a formal strategy to prevent VTE. Nevertheless, it continues to be underused throughout the world. Hospitalized patient with higher risk of VTE includes patients with major surgery, cancer, CHF, COPD, chronic kidney disease, nephrotic syndrome, etc. Review of risk factors often centers on the hospitalized patient; however, about 75% of VTE events occur outside of the hospital setting. There is a relationship between VTE that occurs in the community and previous hospital stay. Many acute VTE cases that occur at home can be linked to a hospital stay or surgical procedure within the preceding 90 days.

VTE can recur. Certain clinical risk factors are associated with recurrent VTE despite the use of anticoagulation medications. A cohort study of 673 patients found that 3.0% of patients suffered recurrence despite anticoagulation and that most (79%) recurrences were fatal. A common clinical problem is deciding whether to prescribe time-limited vs. indefinite-duration anticoagulation. While taking anticoagulant drugs, the risk factors for VTE includes immobilization, cancer, COPD; after anticoagulant drugs are discontinued, the risk factors may involve being a male gender, overweight, obesity, low high-density lipoprotein cholesterol, presenting with symptoms of PE, rather than symptoms of DVT.

Nurses' Health Study cohort investigated risk factors for PE in women. The study enrolled 112,822 women 30 to 55 years of age at baseline, free of cardiovascular disease or cancer. With 16 years of follow-up, three major risk factors for PE were found: obesity, cigarette smoking, and hypertension. Ageno et al. [3] performed a meta-analysis of 63,552 patients from 21 case-control and cohort studies, they found the odds ratio, compared with control subjects, for VTE risk was: 2.3 for obesity, 1.5 for hypertension, 1.4 for diabetes mellitus, and 1.2 for hyper-cholesterolemia. In addition, Heart and Estrogen/Progestin Replacement Study investigated the risk of VTE and post-menopausal hormone therapy in a meta-analysis. Post-menopausal estrogen replacement was associated with an increased risk of VTE. The magnitude of risk was 2 to 3 times the baseline likelihood of VTE. Susceptibility to DVT or PE seemed to be highest during the first year of hormone use.

There seems to be a relationship between markers of inflammation and VTE, but proving a cause-effect relationship remains elusive. The JUPITER (Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosu-

vastatin) study enrolled 17,802 healthy subjects with elevated C-reactive protein levels. They received rosuvastatin 20 mg or its placebo for a median of 2 years. The rosuvastatin group had a 43% reduction in VTE events compared with the control group.

5 Clinical Trial Data in This Analysis

The clinical trial data on multiple myeloma of about 700 penitents is used in this data analysis to identify the risk factors. The patients were treated with either the combination of experimental drug and another active drug (E+A) or placebo and the active drug (P+A). The risk factors considered are similar to the risk factors discussed previously, namely, age, gender, BMI, medical history of VTE, diabetes, cardiac disease, chronic renal disease, and the medications used including antithrombotic prophylaxis, EPO, hormone steroid, and GCSF. Time period of analysis was from the start of study up to the occurrence of VTE for patients with VTE event. For the patients who did not have any VTE during the study, the time period from the start of the study up to the last event day of all the VTEs (from patients with VTE) was used.

VTE can sometimes be detected via laboratory test. The most common laboratory tests for VTE includes D-dimer test. And some clinical indications of VTE including elevated platelet level and elevated hemoglobin level at beginning of trials. Therefore, it is important to examine the longitudinal patterns of the laboratory data for any noticeable trend or abnormality at least at the beginning of the trial.

6 Statistical Analysis to Identify Risk Factors

The commonly used methods to detect risk factors are the logistic regression and Cox regression. However, given the inter-correlation of the risk factors, simple regression approach using these common methods may be overly simplified, and may even produce misleading results as shown in the following. Therefore, more comprehensive methods to assess the goodness of fit, such as marginal models and graphical methods, to account for the non-randomness via matching, as well as mediation effect analysis may be utilized for the more appropriate analysis.

Results from multivariate logistic regression is shown in Table 1 which shows some interesting but clinically counter-intuitive results. For example, GCSF is used primarily to control the level of neutrophiles and does not have much to do with VTE; however, the regression shows GCSF use increase VTE. In addition, use of anticoagulant also increases the possibility of VTE that also contradicts with the conventional clinical thinking. Therefore, further more detailed investigations are

needed. On the other hand, to assess the goodness of fit of the logistic regression model, the traditional residual plot is not very useful. Hence, alternative method such as the marginal modeling, which is described below, can be useful.

Table 1: Estimates from Multivariate Logistic Regression

Parameter	Odds Ratio Estimates			P-value
	Estimate	L. 95% CI	U. 95% CI	
Trt (E+A vs. P+A)	4.215	2.312	7.684	< 0.001
Age	1.044	1.014	1.074	0.0037
Epo	2.268	1.291	3.984	0.0044
Anticoagulant use	0.427	0.211	0.864	0.0179
GCSF use	0.358	0.152	0.846	0.0192
Venous	2.244	1.117	4.506	0.0231

For the marginal modeling, briefly, let y be the dependent variable and X be the covariates. The marginal regression function $E_F(y|a'X)$ is unknown, but can be estimated by smoothing y against $a'X$ for any vector a. The marginal regression function implied by the estimated model $\hat{M}$ can be obtained by averaging over the distribution of $\{y|a'X\}$,

$$E_{\hat{M}}(y|a'X) = E[E_{\hat{M}}(y|X)|a'X], \tag{1}$$

can be estimated by smoothing the fitted values against $a'X$. The variance under $\hat{M}$ can be expressed as

$$\mathrm{Var}_{\hat{M}}(y|a'X) = E[\mathrm{Var}_{\hat{M}}(y|x)|a'X] + \mathrm{Var}[E_{\hat{M}}(y|X)|a'X]. \tag{2}$$

If the model is a close representation of F, we can expect $E_{\hat{M}}(y|a'X) \approx \hat{E}_F(y|a'X)$, and $\mathrm{sd}_F(y|a'X) \approx \mathrm{sd}_{\hat{M}}(y|a'X)$.

Examining the closeness of the blue and red lines of Figures 1 and 2, one can see the model using the significant variables seems to have a reasonably good fit, however, the model using all the available variables does not seem to fit as well as the more parsimonious model.

6.1 Counter-intuitive Dose Effect Findings

Since some subjects received E+A and some subjects received P+A, it is of interest to examine the effect of total E dose and A dose on the effect of VTE. The results from regression are shown in Table 2 and it indicates that the more the dose the less the chance to have VTE since both the coefficients are negative, which is obviously counter-clinical intuitive. The healthier patients tended to stay in the trial longer and received more doses that the patients who had events or dropouts. Therefore, a better analysis is needed to understand the dose effect on VTE.

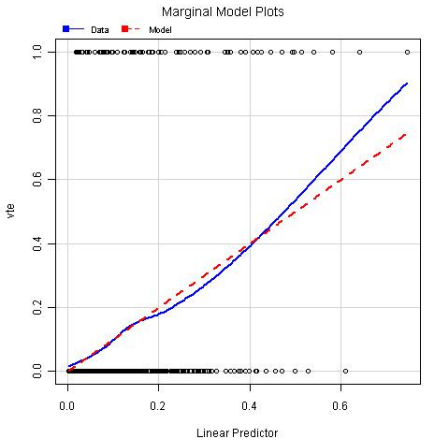

Figure 1: Significant Variables

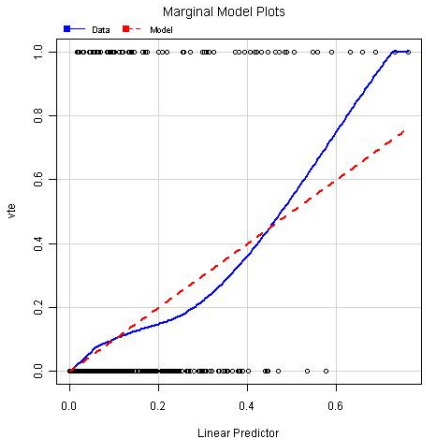

Figure 2: All Variables

Table 2: MLE of Total E Dose and A Dose Effect on VTE Incidence

Parameter	DF	Estimate	Std. Error	Wald χ^2	p
Total E dose	1	-0.00006	0.000037	3.0199	0.0822
Total A dose	1	-0.00056	0.000129	18.6942	$< .0001$

6.2 Analysis of Dose Effect Through Matching

Since the control group received P+A, therefore, the dose effect of A on VTE can be estimated easily. From this estimate, one can estimate the additive effect of E over A on the impact of having VTE for the subjects who received E+A. To estimate this additive effect, one can use the matching method, namely, to match patients with a set of relevant covariates expect for the treatment received. The general rationale can be described as follows.

The individual subject level effect of treatment against control is defined as:

{Subject S's outcome after receiving treatment}−
{Subject S's outcome had the subject received the control}.

The second line above is the so-called counterfactual outcome. These two outcomes cannot be observed simultaneously for subject S over the same time period. Therefore, we seek an alternative subject (subject B), with similar characteristics and observe the outcome under the control exposure in subject B. We use the outcome from subject B as a surrogate outcome for subject S's counterfactual outcome, and allowing us to calculate an individual causal effect. Extending this to all subjects in treatment group, we can calculate the average causal effect for the treated subjects (ATT) (Rosenbaum and Rubin [4]). Similarly, doing this for all subjects in control group, one can calculate the average causal effect for the control subjects (ATC). If one extends this to all subjects in treatment and control groups, one can calculate the average causal effect for all subjects (ATE).

The results are shown in Table 3. As one can easily see that the probability of having VTE incidence increases with the duration of treatment, either A or E, and the addition of drug E on top of A actually further increases the chance of having VTE. These findings are substantially different from the results obtained by simple regressions.

Table 3: Estimate of Counterfactural Dose Effects on VTE Incidence

Effect	N	Cycles						
		≤ 1	≤ 2	≤ 3	≤ 4	≤ 5	≤ 6	All
Treatment A(%)	347	0.58	1.73	2.59	2.88	3.75	3.75	4.32
ATT(%)	352	0.57	1.70	3.13	5.82	5.97	6.82	8.66
(std)		(0.85)	(1.64)	(2.02)	(2.63)	(2.61)	(2.70)	(2.97)
ATE(%)	352	0.57	2.15	4.41	8.32	10.63	11.71	10.16
(std)		(0.91)	(1.35)	(1.85)	(2.39)	(2.65)	(2.72)	(2.72)

7 Estimation of Infection and GCSF Effects

During the course of treatment, some patients may experience abnormally low value of neutrophil counts and that can lead to infection or other serious comoplications. Patients usually receive GCSF to boost the amount of neutrophil and to prevent infection. Therefore, the effects of GCSF and infection on VTE are intertwine and how to entangle this mixed effect is of clinical importance. A simple regression which produced the follow results (Table 4) can be misleading.

Table 4: MLE of Infection and GCSF Effects on VTE Incidence

Parameter	Estimate	Lower 95% CI	Upper 95% CI	p-value
Infection (univ)	1.199	0.680	2.116	0.530
Infection (multiv)	0.634	0.375	1.074	0.090
GCSF (univ)	0.623	0.277	1.400	0.252
GCSF (multiv)	0.358	0.152	0.846	0.019

With univariate regression, infection and GCSF usage were not significant at the 5% level; however, GCSF usage becomes significant and infection becomes marginally significant when the multivariate regression was used. In addition, GCSF usage can be treated as an indicator variable or the number of GCSF usage can be used as covariate because they have different implication clinically. To explore these answers, one can use the mediation effect approach and consider the direct effect of infection and indirect effect of GCSF usage.

Direct effect is the infection effect which is not mediated by a given set of potential mediators such as GCSF, and the indirect or mediator effect is the part of the exposure effect which is mediated by a given set of potential mediators such as GCSF use. The details of this approach can be found in Erikson et al. [5] and references therein.

Briefly, to estimate the effects, one can regress dependent variable (VTE) on both infection and GCSF. Let $O_{x=1,m|x=1}$ be the odds of having VTE for subject with infection ($x = 1$) and GCSF value m. The total infection effect on VTE can be decomposed as following:

$$\underbrace{\ln(O_{x=1,m|x=1}) - \ln(O_{x=0,m|x=0})}_{\text{total effect}} = \underbrace{\ln(O_{x=0,m|x=1}) - \ln(O_{x=0,m|x=0})}_{\text{indirect effect}} + \underbrace{\ln(O_{x=1,m|x=1}) - \ln(O_{x=0,m|x=1})}_{\text{direct effect}} \quad (3)$$

Additional variables, such as dose, laboratory test, and medical history, etc., can also be included in the model.

The model was fit using two different models, the first model uses GCSF usage as an indicator variable, and the second model use the actual number of frequency that GCSF was used. The results from R program are shown below.

```
***** Model #1: GCSF use is treated as indicator variable.

Call: glm(xdat$vte ~ xdat$case + (xdat$gcsfno > 0), family = binomial)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          -2.8894     0.4010  -7.206 5.75e-13
xdat$case             2.8155     0.7463   3.773 0.000162
xdat$gcsfno > 0TRUE   1.3264     0.5663   2.342 0.019181

    Null deviance: 112.547  on 159  degrees of freedom
Residual deviance:  91.761  on 157  degrees of freedom
AIC: 97.761
Number of Fisher Scoring iterations: 5
Total Effect    = 2.84781214283737
Direct Effect   = 2.59281930746949
Indirect Effect = 0.254992835367886


***** Model #2: GCSF useage frequency was used as covariate.

Call: glm(xdat$vte ~ xdat$case + xdat$gcsfno, family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)  -2.5233     0.3208  -7.866 3.67e-15
xdat$case     2.8449     0.7137   3.986 6.72e-05
xdat$gcsfno   0.1062     0.1118   0.950    0.342

    Null deviance: 112.547  on 159  degrees of freedom
Residual deviance:  96.351  on 157  degrees of freedom
AIC: 102.35
Number of Fisher Scoring iterations: 5
Total Effect    = 2.84781214347653
Direct Effect   = 2.83921290678109
Indirect Effect = 0.00859923669544616
```

To evaluate the significance of these estimates, since the asymptotic distribution properties are not readily available, permutation test was used to estimate the significance. Figures 3 to 5 are the permutation distributions from model 1 for the total, direct, and indirect effects, similar information for model 2 are shown in Figures 6 to 8. The red vertical line indicates the 5% critical value of the permutation distribution and the green line indicates the estimated value of the effects. Both of them indicate the significant direct effect and total effect in both models; however, the indirect GCSF effect was significant when it is treated as indicator variable and not significant when the actual frequency of GCSF usage was used as covariate.

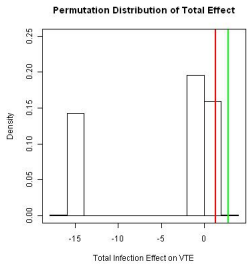

Figure 3: Total Eff.

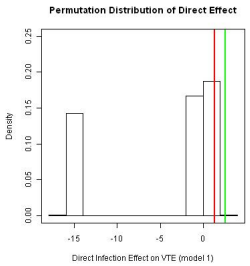

Figure 4: Direct Eff.

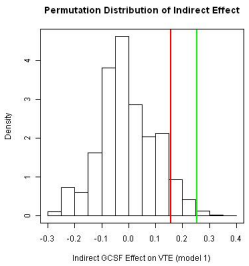

Figure 5: Indirect Eff.

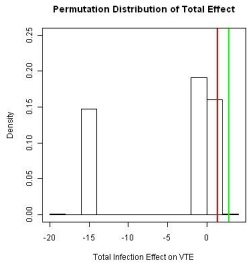

Figure 6: Total Eff.

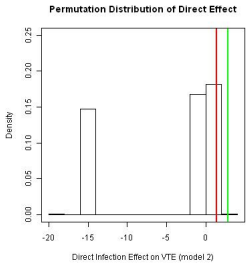

Figure 7: Direct Eff.

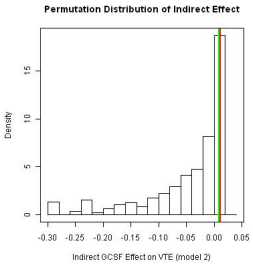

Figure 8: Indirect Eff.

8 Summary

Safety adverse events usually have multiple underlying risk factors and VTE is no exception. It is challenging to identify the risk factors not even mention to confirm them, unless randomized clinical trials are specially designed for the events and that can be extremely costly as in some of the so-called mega studies. Many statistical tools are available beyond the commonly used logistic or Cox regressions. Results from these regressions should only be considered as first order findings. In this article, we showed the danger of stopping further investigation right after these commonly used regression methods.

It is extremely useful and important for the data analysts to have close communications with the subject experts including scientists or clinicians depending on the areas of applications so that statistical techniques can be properly implemented to find the best and relevant results which can be very beneficial to the health of the general public.

References

[1] Goldhaber, S. (2010). Risk Factors for Venous Thromboembolism *Journal of the American College of Cardiology*, v56, p.1-7.

[2] Lyman, G. et al. (2007). American Society of Clinical Oncology Guideline: Recommendations for Venous Thromboembolism Prophylaxis and Treatment in Patients With Cancer *JCO*, p.5490-5505.

[3] Ageno W. et al. (2008). Cardiovascular risk factors and venous thromboembolism: a meta analysis. *Circulation*, v.117, p93-102.

[4] Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, v.79, p.516-524.

[5] Erikson R, et al. (2005). On class differentials in educational attainment. *Proceedings of the National Academy of Science.*, v.102, p.9730-9733.