

Paper 345-2011

## Using SAS® Procedures FREQ, GENMOD, LOGISTIC, and PHREG to Estimate Adjusted Relative Risks – A Case Study

Jiming Fang

Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada

### ABSTRACT

We present nine methods to compute an adjusted relative risk (RR). These methods evolved over the past 25 years (1985–2010) via SAS/STAT® procedures: FREQ, GENMOD, LOGISTIC, and PHREG. We also compare the strengths and limitations of these methods, using an observational cohort study for illustration.

### INTRODUCTION

The relative risk (RR) is a common measure of the effect of treatment or exposure on a dichotomous outcome in cohort studies. Researchers are increasingly using observational studies to estimate the effect of treatment on outcomes. However, unlike randomized controlled trials, treated subjects in non-randomized studies often differ systematically from untreated subjects. The effect of treatment on outcomes cannot be compared directly between groups. Therefore, statistical methods must be used to adjust for systematic differences when estimating the effect of treatment on outcomes. In the present paper, we illustrate 9 methods to compute adjusted relative risks which have been developed in a quarter of a century via 4 different SAS/Stat® procedures: FREQ, GENMOD, LOGISTIC, and PHREG. We will also compare the strengths and limitations of these methods based on an observational cohort study using data from the Registry of Canadian Stroke Network.

#### Study Cohort

We conducted a study to investigate the impact of follow-up at a secondary prevention clinic (SPC) on 1-year mortality in stroke patients. The study cohort was taken from the Registry of Canadian Stroke Network (RCSN), which includes patients seen at all 11 stroke centers in Ontario, Canada between July 2003 and March 2006. Data concerning the date of stroke onset and hospital arrival, stroke type, comorbidities, stroke severity, and outcomes at discharge were abstracted from each patient's chart by trained nurses using custom RCSN data entry software. The risk of 1-year death following stroke onset was determined through linkages to a provincial administrative database. The study cohort consisted of 9074 ischemic or transient ischemic attack (TIA) patients who were alive at discharge. Of these, 4036 patients were referred to a secondary prevention clinic follow-up (SPC=1), and 5038 were not (SPC=0). Patients with SPC were significantly different from those without SPC in terms of their demographic and clinical characteristics (Table 1, P-values <0.05 highlighted in red).

#### Crude RR using Proc Freq

The crude RR provides a measure of the overall association between the risk factor and the outcome, e.g., SPC and 1-year mortality in the present study. It can be obtained easily from Proc Freq using RelRisk option.

```
Proc Freq data=StudyCohort;
    Tables SPC*Death_1year / RelRisk;
Run;
```

The 1-year mortality rates in SPC patients and non-SPC patients were 6.5% and 14.4%, respectively. The crude RR is 0.454 (95% CI: 0.397-0.519), suggesting the 1-year mortality rate for SPC patients was 54.6% lower than for non-SPC patients. However, due to the differences in baseline characteristic (Table 1), we must run multivariate analyses to adjust the RR for the impact of other potential factors that may be related to SPC follow-up.

#### Adjusted RR using Proc Freq – Stratified Mantel-Haenszel

We can use a stratified Mantel-Haenszel Chi-square statistic to control for the other categorical factors, for example, ambulance transportation and hospital admission. This adjusted RR may identify the role of the risk factor of interest (SPC) after the risk from other factors(s) has been statistically removed (Greenland & Robins 1985). Here is Mantel-Haenszel test:

```
Proc Freq data= StudyCohort;
    Tables Ambulance*Admission*SPC*Death_1year / RelRisk;
Run;
```

**Table 1. Baseline comparisons**

Variable	Value	No-SPC	SPC	P-value
Sample size	n	5038	4036	
Male	n (%)	2568 (51.0%)	2171 (53.8%)	0.0076
Lives with others	n (%)	3547 (70.4%)	2953 (73.2%)	0.0037
Urban residence	n (%)	4403 (87.4%)	3755 (93.0%)	0.0000
Neighborhood income quintile 1	n (%)	1246 (24.7%)	788 (19.5%)	0.0000
Neighborhood income quintile 2	n (%)	1076 (21.4%)	803 (19.9%)	0.0877
Neighborhood income quintile 3	n (%)	958 (19.0%)	792 (19.6%)	0.4658
Neighborhood income quintile 4	n (%)	856 (17.0%)	724 (17.9%)	0.2368
Neighborhood income quintile 5	n (%)	902 (17.9%)	929 (23.0%)	0.0000
Presenting to ER from home	n (%)	3635 (72.2%)	3233 (80.1%)	0.0000
Transport by ambulance	n (%)	3176 (63.0%)	2006 (49.7%)	0.0000
Ischemic	n (%)	3468 (68.8%)	2572 (63.7%)	0.0000
Weakness stroke symptom	n (%)	3760 (74.6%)	2784 (69.0%)	0.0000
Stroke classification - TASC	n (%)	367 (7.3%)	165 (4.1%)	0.0000
Emergency consultation	n (%)	2257 (44.8%)	2031 (50.3%)	0.0000
tPA treatment	n (%)	376 (7.5%)	301 (7.5%)	0.9922
Preadmission independence	n (%)	3854 (76.5%)	3588 (88.9%)	0.0000
Charlson score >=2	n (%)	1689 (33.5%)	1098 (27.2%)	0.0000
Admitted to hospital	n (%)	3662 (72.7%)	2102 (52.1%)	0.0000
Deficit at discharge	n (%)	3263 (64.8%)	2322 (57.5%)	0.0000
Discharge to home	n (%)	2732 (54.2%)	3071 (76.1%)	0.0000
Modified Rankin score 3~5	n (%)	2186 (43.4%)	975 (24.2%)	0.0000
Age (Years)	Mean±SD	72.24±13.69	69.09±13.65	0.0000
Stroke severity measured by Canadian Neurological Score	Mean±SD	8.91±2.76	9.77±2.27	0.0000
Length of stay (Days)	Mean±SD	11.42±22.72	6.3±14.87	0.0000
Death in 1-year	n (%)	726 (14.4%)	264 (6.5%)	0.0000

SD=standard deviation

The Mantel-Haenszel adjusted RR is 0.517 (95% CI: 0.451-0.591), suggesting that after controlling for ambulance transport to the stroke centre and hospital admission, the 1-year mortality rate for SPC patients was 48.3% lower than for non-SPC patients. Stratification is attractive when statistical control of a few categorical covariates is required. However, it can be difficult to implement in practice when there are many confounding covariates, especially if some of the confounders are continuous.

### Adjusted RR using Proc GenMod – Log-Binomial regression Model

When we need to adjust for many covariates, including continuous covariates, we can use Log-Binomial regression (McNutt et al. 2003; Wacholder 1986), which is implemented in the GenMod procedure. Here is the SAS program using Log-Binomial regression to adjust for other covariates:

```
Proc GenMod data=StudyCohort descending;
  Class SPC/param=ref ref=first;
  Model Death_1year=SPC Var_1- Var_n / Dist=bin Link=log;
  Estimate 'RR SPC vs. Non-SPC' SPC 1/exp;
Run;
```

Here Var\_1 – Var\_n include 24 covariates besides SPC, such as gender, age, ambulance transportation, admission, etc. However, the program did not run successfully and produced the following error message.

WARNING: The specified model did not converge.

```
ERROR: The mean parameter is either invalid or at a limit of its range for some observations.
```

The probability of an outcome must fall within the bounds [0, 1]. However, the Log link function in Log-Binomial models restricts the probabilities of an outcome to be greater than or equal to zero, that is, to fall within the bounds [0, ∞). Due to this mismatch between the bounds of the model and the allowable outcome, in practice, the Log-Binomial model will routinely fail to converge and will not provide the parameter estimates (Localio et al. 2007). The failure of convergence in the Log-Binomial regression may also indicate that the data do not support the model (Tian & Liu 2006).

We found that if 15 covariates are included in the model instead of desired 24, the model does converge. The adjusted RR is 0.561 (95% CI: 0.490-0.642, with StdErr=0.0389).

### Adjusted RR using Proc GenMod – Log-Binomial regression Model with negative intercept

When all predictors are zero or at their reference levels in the multivariate Log-Binomial regression model, the intercept estimates  $\log(p) < 0$  as  $0 < p < 1$ . So it makes sense to start its estimation in the negative value. It was found that starting value of -4 for the intercept has worked well in practice (Deddens et al. 2003).

```
Proc GenMod data=StudyCohort descending;
  Class SPC/param=ref ref=first;
  Model Death_1year=SPC Var_1- Var_n /Dist=bin Link=log intercept=-4;
  Estimate 'RR SPC vs. Non-SPC' SPC 1/exp;
Run;
```

The adjusted RR from the Log-Binomial regression Model with negative intercept is 0.806 (95% CI: 0.672-0.968), with StdErr = 0.0751. However, the following SAS warning messages suggests that the convergence problem has not been completely solved by the negative intercept given and the model fit is still questionable.

```
WARNING: The relative Hessian convergence criterion of 0.126294346 is greater than the limit of 0.0001. The convergence is questionable.
```

```
WARNING: The procedure is continuing but the validity of the model fit is questionable.
```

### Adjusted RR using Proc GenMod – Poisson regression model

In contrast to the Log-Binomial regression model, the Poisson regression model, using all 24 covariates, has no difficulty with convergence (McNutt et al. 2003). Poisson distribution would be expected to be a good approximation to the binomial distribution when the outcome is low and the sample size is large. Here is the Poisson regression using Proc GenMod:

```
Proc GenMod data=StudyCohort descending;
  Class SPC/param=ref ref=first;
  Model death_1year=SPC Var_1- Var_n / Dist=poisson Link=log;
  Estimate 'RR SPC vs. Non-SPC' SPC 1 /exp;
Run;
```

The adjusted RR from the Poisson regression model is 0.777 (95% CI: 0.667-0.905), with StdErr = 0.0607. However, one limitation in the Poisson approximation is that the estimated probabilities from the Poisson model may be greater than 1, which is invalid (Deddens & Petersen 2004).

### Adjusted RR using Proc GenMod – Modified Poisson regression model

Poisson regression without robust error variances may result in a conservative CI (i.e., wider CI). A “modified Poisson” method has been proposed to estimate the RR using a robust error variance (Zou 2004). This method leads to the robust error variance estimation and produces 95% CIs with the correct coverage. Using SAS, the robust error variances can be obtained by using the repeated statement and the subject identifier (here PatientID), even though there is only one observation per subject. Here is the SAS program (Spiegelman & Hertzmark 2005).

```
Proc GenMod data=StudyCohort descending;
  Class PatientID SPC/param=ref ref=first;
  Model Death_1year=SPC Var_1-Var_n / Dist=poisson Link=log;
  Repeated subject=PatientID / type=Ind;
  Estimate 'RR SPC vs. Non-SPC' SPC 1 /exp;
Run;
```

The RR from the modified Poisson regression is 0.777 (95% CI: 0.675-0.894), with StdErr=0.0555, which is smaller than the StdErr (0.0607) from simple Poisson regression. However, this method may fail when outcomes are common,

for example, the point estimate for risk and the upper CI bounds of the expected probability may exceed 1 because the log link does not constrain expected probabilities (Localio et al. 2007).

### Adjusted RR using Proc Logistic – OR-to-RR formula

We may also obtain the adjusted RR from the adjusted odds ratio (OR) using the simple relationship (Zhang & Yu 1998):

$$OR = \frac{\frac{P_T}{1-P_T}}{\frac{P_C}{1-P_C}} \quad \text{and} \quad RR = \frac{P_T}{P_C} = \frac{OR}{(1-P_C) + (P_C \times OR)}$$

where  $P_C$  and  $P_T$  are the unadjusted risks in the control and treat groups, respectively.

CI's for the RR are estimated by substituting the upper and lower CI's for the OR from the multivariate logistic regression model (Daly 1998). In the present study, the following SAS code was used:

```
Proc Logistic data=StudyCohort descending;
  Class SPC/param=ref ref=first;
  Model Death_1year=SPC Var_1-Var_n /rl lackfit;
Run;
```

Adjusted OR=0.731 (95% CI: 0.618-0.866)

$P_C=0.144$

Adjusted RR=0.761 (95% CI: 0.654-0.883)

This RR is biased away from the null, suggesting a stronger association. However, if the incidence of an outcome of interest is common in the study population (say, >10%, Figure 1), the adjusted OR derived from the logistic regression can no longer approximate the RR (Zhang & Yu 1998). When the outcome is common, the RR estimated using logistic regression will be more extreme (farther from 1.0) than the RR for the same data (Altman et al. 1998; Deeks 1998). In addition, the proposed CI's for the RR will be too narrow because this approach fails to account for variability in the baseline risk (Localio et al. 2007; McNutt et al. 2003).

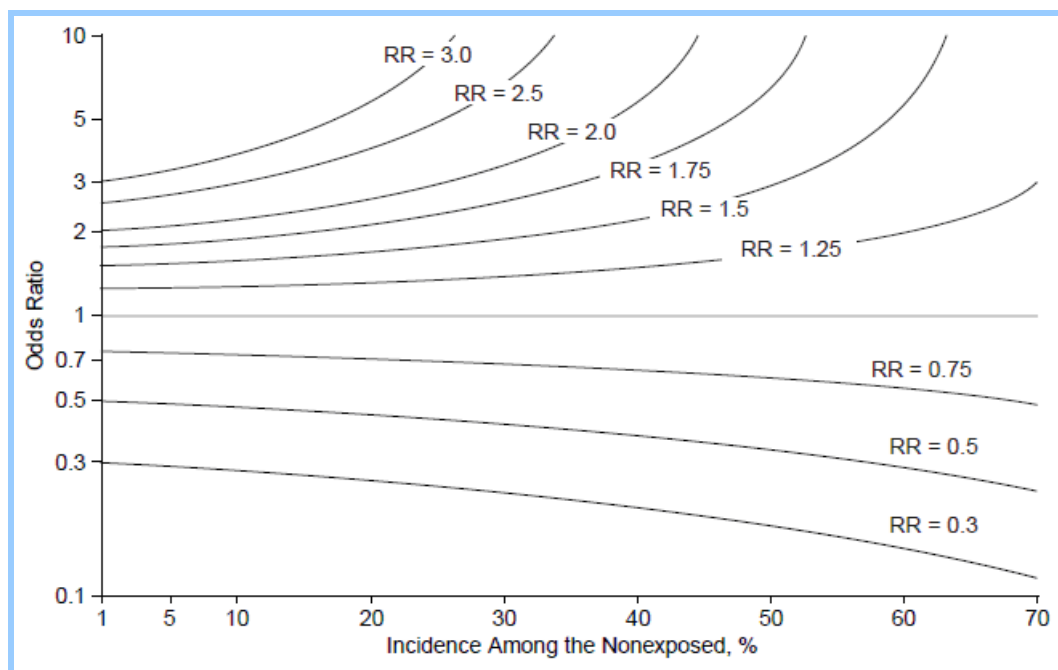


Figure 1. The relationship between relative risk and odds ratios by incidence of the outcome (Cited from Zhang & Yu 1998)

### Adjusted RR using Proc Logistic – Propensity-score matching

Propensity-score matching is frequently used in the medical literature to estimate the effect of treatments and exposures on health outcomes (Austin 2008). The propensity score is defined as a subject's probability of receiving the

treatment or the probability of exposure, conditional on his/her observed baseline characteristics. It is usually estimated using a logistic regression model (Austin et al. 2010). Once the propensity score has been estimated for each subject, treated and untreated subjects are matched on the propensity score. The most commonly-used method is to form pairs with similar propensity scores. The next step is to use the standardized difference to examine the balance in measure baseline variables between treated and untreated subjects. If the balance is acceptable, we then estimate the effect of treatment on the outcome via the appropriate statistical tests (Austin 2007).

Austin and his colleagues illustrate the detailed SAS coding to obtain the RR using propensity-score matching method (Austin et al. 2010). In the present study, 3,114 matched pairs of patients were identified (total n=6,228, less than the original sample size). Figures 2a and 2b show the similarity between SPC and non-SPC patients before (Figure 2a) and after (Figure 2b) propensity-score matching.

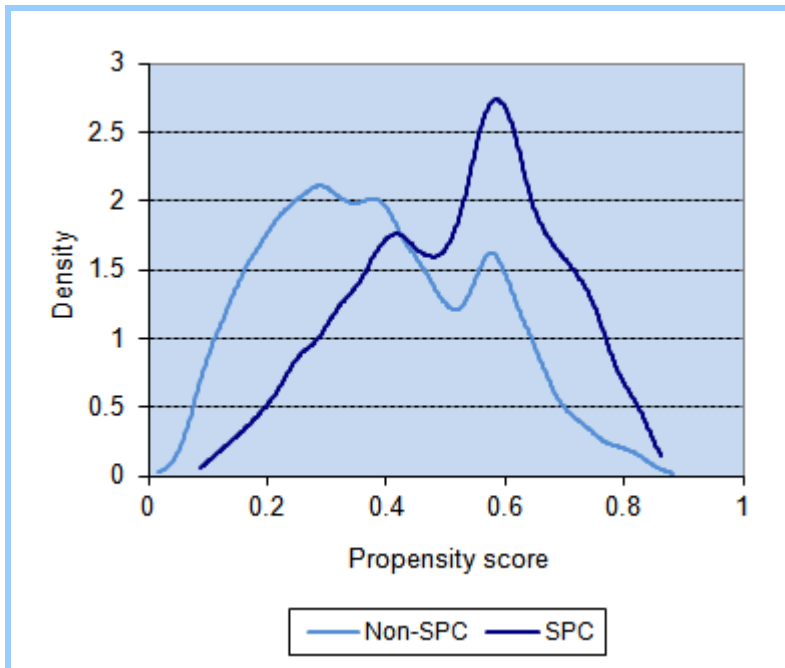


Figure 2a. Density distribution of propensity scores – Original cohort

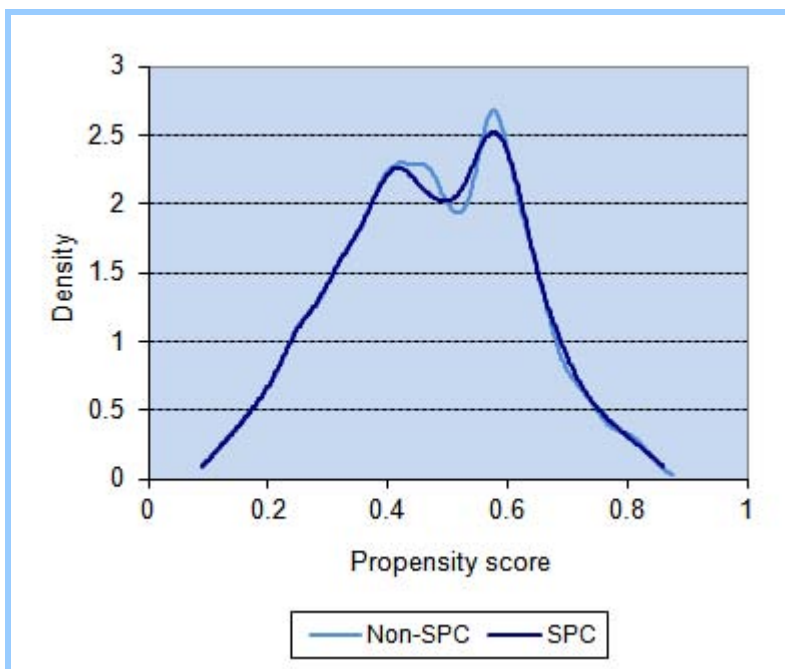


Figure 2b. Density distribution of propensity scores – Propensity-score matched cohort

In the present study, there were 32 matched pairs in which both subjects died within one year of the stroke, 2598 matched pairs in which neither subject died within one year of the stroke, 291 matched pairs in which the Non-SPC patient died and the SPC patient did not die, and 193 matched pairs in which the SPC patient died and the Non-SPC patient did not die. According to the method proposed by Agresti and Min to estimate the RR and its confidence interval for matched data (Agresti & Min 2004), the RR of 1-year mortality for SPC patients compared to Non-SPC patients was 0.697, and the 95%CI was 0.594-0.817.

		SPC	
		Dead	Alive
Non-SPC	Dead	32 (a)	291 (b)
	Alive	193 (c)	2598 (d)

$$RR = \frac{a+c}{a+b}$$

$$RR \text{ 95\% CI given as } e^{\log(RR) \pm 1.96 \sqrt{\frac{b+c}{(a+b) \times (a+c)}}}$$

### Adjusted RR using Proc Logistic – Marginal Probability and Bootstrapping

Austin introduced a new method for deriving the adjusted RR from a logistic regression model (Austin 2010b). This method involves determining the probability of the outcome if each patient in the cohort was treated, and again if each patient was untreated. These probabilities are then averaged across the study cohort to determine the average probability of the outcome in the population if all patients were treated, and if they were untreated.

$$\log\left(\frac{\Pr(Y_i = 1)}{1 - \Pr(Y_i = 1)}\right) = \alpha_0 + \beta T_i + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_k X_{ki}$$

Y=1 denoting outcome success (e.g., dead), Y=0 denoting outcome failure (e.g., alive)

T=1 denoting treatment (SPC), T=0 denoting control (Non-SPC)

X<sub>1</sub> to X<sub>k</sub> denote k confounding covariates

β denotes the log-odds ratio, e<sup>β</sup> denotes the odds ratio

We can estimate the probability of the outcome if a given patient was treated, and if the same patient was untreated.

Probability of the outcome if a patient was treated:

$$\frac{e^{\alpha_0 + \beta + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_k X_{ki}}}{1 + e^{\alpha_0 + \beta + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_k X_{ki}}}$$

Probability of the outcome if a patient was untreated:

$$\frac{e^{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_k X_{ki}}}{1 + e^{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_k X_{ki}}}$$

We then compute the mean probability ( $\bar{P}_{T=1}$ ) of success in the cohort if all patients were treated, and the mean probability ( $\bar{P}_{T=0}$ ) of success in the cohort if all patients were untreated. These are referred to as the marginal

probabilities of success for treated and untreated patients. The adjusted RR is estimated as  $\frac{\bar{P}_{T=1}}{\bar{P}_{T=0}}$ .

Use of marginal probabilities allows one to compare outcomes between two populations whose only difference is the exposure. Because all patients contribute to both  $\bar{P}_{T=1}$  and  $\bar{P}_{T=0}$  there are no systematic differences in baseline

characteristics between the two populations.

The solution to the multivariate logistic regression model of the study cohort is:

$$\text{Log}(p/(1-p)) = -3.4767 - 0.3128 * \text{SPC} + 0.1516 * \text{Male} + \dots - 0.00329 * \text{LOS}$$

$P_{T=1}$  and  $P_{T=0}$  can be computed using the following SAS program:

```
Data PT_PC;
  Set StudyCohort;
  Ln_PT=-3.4767 - 0.3128*SPC + 0.1516*Male + ...-0.00329*LOS;
  Ln_PC=-3.4767          + 0.1516*Male + ...-0.00329*LOS;
  PT=exp(Ln_PT) / (1+exp(Ln_PT));
  PC=exp(Ln_PC) / (1+exp(Ln_PC));
Run;
```

However, typing these SAS codes is tedious and not efficient. We show an easy way to compute  $P_{T=1}$  and  $P_{T=0}$ . First, we generate a population cohort which includes both the treated cohort and the control cohort.

```
Data Population;
  Set StudyCohort (in=a)
      StudyCohort (in=b);
  If a then SPC=1;
  If b then SPC=0;
Run;
```

We then run logistic regression using the score option.

```
Proc Logistic data=StudyCohort descending;
  Class SPC/param=ref ref=first;
  Model Death_1year=SPC Var_1-Var_n/rl;
  Score data=Population out=Pred_risk;
Run;
```

We then compute the mean probability of death for each patient in the population cohort – once for the patients as if they are all untreated (SPC=0), and again for the patients as if they are all treated (SPC=1). The ratio of these two mean probabilities is the estimated RR.

```
Proc Means data=Pred_risk nway;
  Class SPC;
  Var Prob;
  Output out=pop_risk mean=pop_risk;
Run;

Proc Transpose data=pop_risk out=pop_risk prefix=SPC_;
  Id SPC;
  Var pop_risk;
Run;

Data pop_risk;
  Set pop_risk;
  Adjusted_RR=SPC_1/SPC_0;
Run;

Proc Print data=pop_risk;
  Var Adjusted_RR;
Run;
```

The confidence interval of the RR can be estimated using the bootstrap method (Efron & Tibshirani 1993). A bootstrap sample is a random sample drawn with replacement from the original sample such that the random sample has the same size as the original sample. Constructing nonparametric bootstrap 95% CIs requires drawing a large number of bootstrap samples (say 1000 bootstrap samples) and estimating the quantity of interest in each of the bootstrap samples. The endpoints of the nonparametric bootstrap 95% CIs would be the 2.5th and 97.5th percentiles of that quantity across the bootstrap samples.

Applying the method of Austin (Austin 2010b), we got an adjusted RR of 0.789 if all patients were referred to SPC



compared with the case where all patients were not referred to SPC (95% CI: 0.690-0.889). Thus, SPC was associated with a 21.1% relative decrease (95% CI: 11.1%-31.0%) in the risk of 1-year mortality.

### Adjusted RR using Proc Phreg – Time-to-event

Statistically, survival models for time-to-event outcomes are more powerful than logistic models for testing the impact of treatment for event outcomes. Austin described a method to derive the RR of an event occurring within a specific duration of follow-up using an adjusted survival model (Austin 2010a). The method allows for the estimation of measures of treatment effect that may be more clinically meaningful than the adjusted hazard ratio that is obtained directly from the Cox proportional hazards regression model.

The SAS program is similar to that in last section above (see Adjusted RR using Proc Logistic – Marginal Probability and Bootstrapping). The adjusted RR from survival models is 0.769 (95% CI: 0.671-0.865).

```

/* Dataset population generated above */

Proc Phreg data=StudyCohort descending;
  Model SurvivalTime_1y*Death_1year(0)=SPC Var_1-Var_n/rl;
  Baseline      out=Pred_risk
                covariates=Population
                survival=survival/nomean;

Run;

Data Pred_risk;
  Set Pred_risk;
  Event_risk=1-survival;
  Where SurvivalTime_1y=365;

Run;

/* RR calculation omitted (See SAS codes above) */

```

## CONCLUSION

The crude and adjusted RRs of SPC are summarized in Figure 3. Obviously, the crude RR is further away from 1 than the adjusted ones. Thus, the impact of SPC is over-estimated by the crude rate, which suggests that the risk-adjustment is necessary. According to the whole study cohort, the point estimations of the adjusted RRs using Poisson regression (0.777), modified Poisson regression (0.777), Logistic regression (0.789) and Cox proportional hazards model (0.769) are quite close to one another. These adjusted RRs indicated that the ischemic stroke or TIA patients referred to SPCs had greater survival than those without a referral to a SPC. According to the Propensity-score matching cohort, the adjusted RR is 0.697, which suggests even more positive impact of the SPC on patient survival.

In this article, we describe 9 methods to derive adjusted RRs which were developed in the past 25 years, from 1985 to 2010, and illustrate the SAS program codes to estimate the adjusted RR accordingly. Table 2 shows their strengths and limitations. In general, the data structure *per se* may lead to the method that should be used to estimate the adjusted RR.

If there is no convergence problem, we can just use a Log-Binomial model to get the adjusted RR. However, if there is a convergence problem, we should apply Modified Poisson regression instead. Petersen and Deddens compared both Log-Binomial model and Modified Poisson regression and found (1) for very high prevalence and moderate sample size, the Modified Poisson method yields less biased estimate of the prevalence ratios than the Log-Binomial method; (2) However, for moderate prevalence and moderate sample size, the Log-Binomial method yields slightly less biased estimate than the Modified Poisson method; (3) In nearly all cases, the Log-Binomial method yields slightly higher power and smaller standard errors than the Modified Poisson method (Petersen & Deddens 2008).

If computing time is not an issue and both Log-Binomial and Modified Poisson regression models are questionable, then we can obtain the adjusted RR using a Logistic regression model or Cox proportional hazards regression model. Using these two models, we are able to get not only the RR, but also the other meaningful measures of treatment effect, such as the absolute risk reduction, the RR reduction and the number needed to treat (Austin 2010a; Austin 2010b).



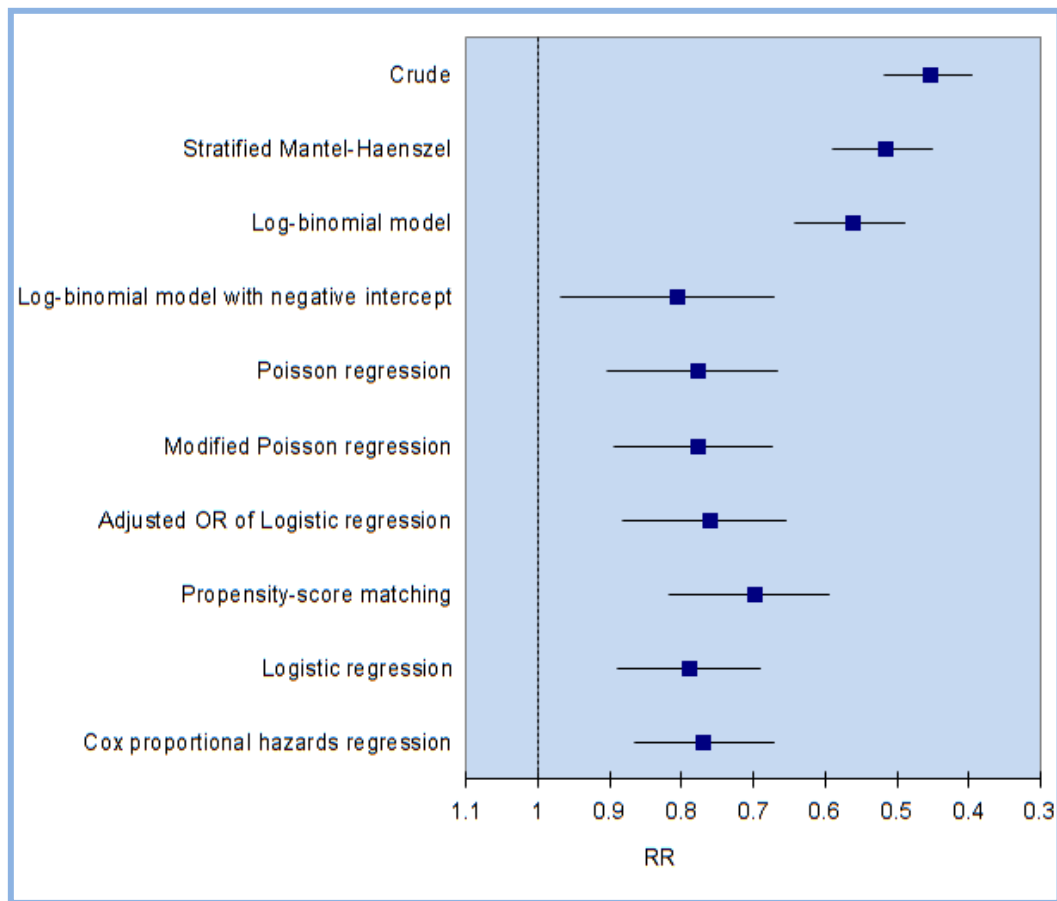


Figure 3. Forest Plot - Comparison between RR computing methods

Table 2. Comparison between different RR computing methods

Methods of Computing Adjusted Relative Risk	References	Estimate of Probability can exceed 1	Biased Estimation	Converge Problem	Wide Confidence Interval	Adjust Continuous Covariates	Time Consuming	Using Whole Study Cohort
Stratified Mantel-Haenszel	(Greenland & Robins 1985)					x		
Log-Binomial model	(Wacholder 1986)			✓		✓		
Log-Binomial model with negative intercept	(Deddens et al. 2003)			✓		✓		
Poisson regression	(McNutt et al. 2003)	✓			✓	✓		
Modified Poisson regression	(Zou 2004)	✓				✓		
Adjusted OR of Logistic regression	(Zhang & Yu 1998)		✓			✓		
Propensity-score matching	(Austin 2008)					✓	✓	x
Logistic regression	(Austin 2010b)					✓	✓	
Cox proportional hazards regression	(Austin 2010a)					✓	✓	

## REFERENCES

- Agresti,A. & Min,Y. 2004. Effects and non-effects of paired identical observations in comparing proportions with binary matched-pairs data. *Stat Med* **23**, 65-75.
- Altman,D.G., Deeks,J.J. & Sackett,D.L. 1998. Odds ratios should be avoided when events are common. *BMJ* **317**, 1318.
- Austin,P.C. 2007. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thorac. Cardiovasc Surg* **134**, 1128-1135.
- Austin,P.C. 2008. The performance of different propensity-score methods for estimating relative risks. *J Clin Epidemiol* **61**, 537-545.
- Austin,P.C. 2010a. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J Clin Epidemiol* **63**, 46-55.
- Austin,P.C. 2010b. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. *J Clin Epidemiol* **63**, 2-6.
- Austin,P.C., Chiu,M., Ko,D.T., Geree,R. & Tu,J.V. 2010. Propensity score matching for estimating treatment effects. In: *Analysis of Observational Health Care Data Using SAS* (Ed. by D.E.Faries, A.C.Leon, J.M.Haro & R.L.Obenchain), pp. 23-49. Cary, NC, SAS Institute Inc.
- Daly,L.E. 1998. Confidence limits made easy: interval estimation using a substitution method. *Am J Epidemiol* **147**, 783-790.
- Deddens,D.A., Petersen,M.R. & Lei,X. 2003. Estimation of prevalence ratios when PROC GENMOD does not converge. *SAS User Group International Proceedings paper* **270-28**.
- Deddens,J.A. & Petersen,M.R. 2004. Re: "Estimating the relative risk in cohort studies and clinical trials of common outcomes". *Am J Epidemiol* **159**, 213-214.
- Deeks,J. 1998. When can odds ratios mislead? Odds ratios should be used only in case-control studies and logistic regression analyses. *BMJ* **317**, 1155-1156.
- Efron,B. & Tibshirani,R.J. 1993. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall.
- Greenland,S. & Robins,J.M. 1985. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* **41**, 55-68.
- Localio,A.R., Margolis,D.J. & Berlin,J.A. 2007. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *J Clin Epidemiol* **60**, 874-882.
- McNutt,L.A., Wu,C., Xue,X. & Hafner,J.P. 2003. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* **157**, 940-943.
- Petersen,M.R. & Deddens,J.A. 2008. A comparison of two methods for estimating prevalence ratios. *BMC Med Res Methodol.* **8**, 9.
- Spiegelman,D. & Hertzmark,E. 2005. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* **162**, 199-200.
- Tian,L. & Liu,K. 2006. Re: "Easy SAS calculations for risk or prevalence ratios and differences". *Am J Epidemiol* **163**, 1157-1158.
- Wacholder,S. 1986. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol* **123**, 174-184.
- Zhang,J. & Yu,K.F. 1998. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* **280**, 1690-1691.

Zou,G. 2004. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* **159**, 702-706.

## ACKNOWLEDGEMENTS

We thank Jennifer Waller, Ruth Croxford and Paul Cascagnette for helpful comments on the manuscript. We wish to acknowledge the helpful comments of Peter Austin, Kelvin Lam and Hong Zheng. This study used data from the Registry of Canadian Stroke Network (RCSN). The RCSN is funded by the Canadian Stroke Network and the Ontario Ministry of Health and Long-Term Care. The Institute for Clinical Evaluative Sciences is supported by an operating grant from the Ontario Ministry of Health and Long-Term Care.

## CONTACT INFORMATION

Jiming Fang, PhD  
Program Lead Analyst - Cardiovascular  
Institute for Clinical Evaluative Sciences  
2075 Bayview Avenue, G106  
Toronto, Ontario M4N 3M5  
Canada  
Works Phone: 416-480-6100 Ext. 3613  
Fax: 416-480-6048  
E-mail: [jiming.fang@ices.on.ca](mailto:jiming.fang@ices.on.ca)  
Web : [www.ices.on.ca](http://www.ices.on.ca)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.  
Other brand and product names are trademarks of their respective companies.