SAS® GLOBAL FORUM 2018

USERS PROGRAM

A SAS Program for a Model-Based Stratification Method

April 8 – 11 | Denver, CO

#SASGF

A SAS Program for a Model-Based Stratification Method

Xiaoli Lu

VA Cooperative Studies Program, Perry Point, MD 21902

ABSTRACT

Some important factors, such as baseline characteristics, are not always balanced between comparison groups in a clinical study. The imbalance often leads to biased estimations. In this report, a SAS programmer was developed for a new approach that combines both simple stratification and model-based covariate adjustment methods, i.e. adjust unbalanced factors using propensity score and stratify the factor by subgroup information when interaction between the stratified factor and the treatment is absent. The treatment effect coefficients estimated from covariate adjusted regression models for each stratum are combined using a weighted method based on either the sample sizes of each treatment group for each stratum or the inverse square of standard error for each stratum coefficient. When interaction is present, subgroup analysis results are reported. The analytical approach is illustrated using different statistical models including analysis of covariance, survival analysis, and logistic regressions based on clinical data. The potential biased estimation and testing efficiency in the imbalance adjustment are evaluated based on different types of clinical outcomes.

METHODS

An approach that combines both stratification and covariate adjustment methods is proposed:

- Define subgroups for the analysis
- Check imbalance between the intervention groups
- Estimate propensity scores based on imbalanced covariates
- Fit regression models for each subgroup stratum
- Find common estimate if there is no interaction between the subgroup strata and the intervention groups

Cochran-Mantel-Haenszel Method:

H0: $\theta_{AB(1)} = \theta_{AB(2)} = \ldots = \theta_{AB(K)} = 1$

Test statistic : $M^2 = \frac{\left[\sum_k (n_{k11} - \mu_{k11})\right]^2}{\sum_k \sigma^2(n_{k11})}$ $\quad \sigma^2(n_{k11}) = \frac{n_{k1\bullet} n_{k2\bullet} n_{k\bullet 1} n_{k\bullet 2}}{n_k^2 (n_k - 1)}$

Common odds ratio estimate : $OR_{MH} = \frac{\sum_k (n_{k11} n_{k22})/n_k}{\sum_k (n_{k12} n_{k21})/n_k}$

100(1-α/2) confidence interval for OR_{MH}: $(OR_{MH} \times \exp(-z\hat{\sigma}),\ OR_{MH} \times \exp(z\hat{\sigma}))$

Variance :

$$\hat{\sigma}^2 = \mathrm{var}(\ln(OR_{MH})) = \frac{\sum_k (n_{k11} + n_{k22})(n_{k11} n_{k22})/n_k^2}{2\left(\sum_k n_{k11} n_{k22}/n_k\right)^2} + \frac{\sum_k \left[(n_{k11} + n_{k22})(n_{k12} n_{k21}) + (n_{k12} + n_{k21})(n_{k11} n_{k22})\right]/n_k^2}{2\left(\sum_k n_{k11} n_{k22}/n_k\right)\left(\sum_k n_{k12} n_{k21}/n_k\right)} + \frac{\sum_k (n_{k12} + n_{k21})/n_k^2}{2\left(\sum_k n_{k12} n_{k21}/n_k\right)^2}$$

Woolf Logit Estimate:

Common odds ratio estimate: $OR_L = \exp\left(\sum_k w_k \ln(OR_k) / \sum_k w_k\right)$

100(1-α/2) confidence interval for OR_L: $\left(OR_L \times \exp\left(-z/\sqrt{\sum_k w_k}\right), OR_L \times \exp\left(z/\sqrt{\sum_k w_k}\right)\right)$ $\quad w_k = 1/\mathrm{var}(\ln(OR_k))$

METHODS CONTINUED

Stratification Methods :

1. Weight on Inverse Variance

Common estimate: $\hat{\theta} = \left(\sum_k w_k \hat{\theta}_k / \sum_k w_k\right)$ $\quad w_k = 1/\hat{\sigma}^2(\theta_k)$

Test statistic: $Z = \frac{\hat{\theta}}{\sqrt{\sum_k w_k}}$

100(1-α/2) confidence interval for θ: $\left(\hat{\theta} - z/\sqrt{\sum_k w_k}, \hat{\theta} + z/\sqrt{\sum_k w_k}\right)$

100(1-α/2) confidence interval for OR: $\left(OR \times \exp\left(-z/\sqrt{\sum_k w_k}\right), OR \times \exp\left(z/\sqrt{\sum_k w_k}\right)\right)$

2. Weight on N

Common estimate: $\hat{\theta} = \sum_k w_k \theta_k$ $\quad w_k = \frac{n_{k1} n_{k2} / (n_{k1} + n_{k2})}{n}$

Test statistic: $Z = \frac{\hat{\theta}}{SE}$ $\quad SE = \sqrt{\sum_k w_k^2 \hat{\sigma}_k^2}$

100(1-α/2) confidence interval for θ: $\left(\hat{\theta} - z \times SE, \hat{\theta} + z \times SE\right)$

100(1-α/2) confidence interval for OR: $\left(OR \times \exp(-z \times SE), OR \times \exp(z \times SE)\right)$

Stratification Models:

Model Strategy	Binary (Logistic Regression)	Time-to Event (Cox Model)	Continuous (ANCOVA)
No adjustment	$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \theta v_i$	$\lambda(t_j) = \lambda_0(t_j)\exp(\alpha + \theta v_i)$	$y_i = \alpha + \theta v_i$
Adjusted on subgroup	$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta x_i + \theta v_i + \tau(x_i v_i)$	$\lambda(t_j) = \lambda_0(t_j)\exp(\alpha + \theta v_i + \beta x_i + \tau(x_i v_j))$	$y_i = \alpha + \theta v_i + \beta x_i + \tau(x_i v_j)$
Stratified on subgroup	$Logit(p_{ki}) = \ln\left(\frac{p_{ki}}{1-p_{ki}}\right) = \alpha_k + \theta_k v_{ki}$	$\lambda_k(t_j) = \lambda_{k0}(t_j)\exp(\alpha_k + \theta_k v_{ki})$	$y_k = \alpha_k + \theta_k v_{ki}$
Adjusted for other covariates but not subgroup	$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \theta v_i + \gamma_1 z_{1i} + \ldots + \gamma_m z_{mi}$	$\lambda(t_j) = \lambda_0(t_j)\exp(\alpha + \theta v_i + \gamma_1 z_{1i} + \ldots + \gamma_m z_{mi})$	$y_i = \alpha + \theta v_i + \gamma_1 z_{1i} + \ldots + \gamma_m z_{mi}$
Adjusted for subgroup and other covariates	$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta x_i + \theta v_i + \tau(x_i v_i) + \gamma_1 z_{1i} + \ldots + \gamma_m z_{mi}$	$\lambda(t_j) = \lambda_0(t_j)\exp(\alpha + \beta x_i + \theta v_i + \tau(x_i v_i) + \gamma_1 z_{1i} + \ldots + \gamma_m z_{mi})$	$y_i = \alpha + \beta x_i + \theta v_i + \tau(x_i v_i) + \gamma_1 z_{1i} + \ldots + \gamma_m z_{mi}$
Stratified on subgroup with adjusting other covariates	$Logit(p_{ki}) = \ln\left(\frac{p_{ki}}{1-p_{ki}}\right) = \alpha_k + \theta_k v_{ki} + \gamma_{k1} z_{k1i} + \ldots + \gamma_{km} z_{kmi}$	$\lambda_k(t_j) = \lambda_{k0}(t_j)\exp(\alpha_k + \theta_k v_{ki} + \gamma_{k1} z_{1ki} + \ldots + \gamma_{km} z_{mki})$	$y_{ki} = \alpha_k + \theta_k v_{ki} + \gamma_{k1} z_{1ki} + \ldots + \gamma_{km} z_{mki}$
Adjusted for propensity score but not subgroup	$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \theta v_i + \phi s_i$	$\lambda(t_j) = \lambda_0(t_j)\exp(\alpha + \theta v_i + \phi s_i)$	$y_i = \alpha + \theta v_i + \phi s_i$
Adjusted for subgroup and propensity score	$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta x_i + \theta v_i + \tau(x_i v_i) + \phi s_i$	$\lambda(t_j) = \lambda_0(t_j)\exp(\alpha + \beta x_i + \theta v_i + \tau(x_i v_i) + \phi s_i)$	$y_i = \alpha + \beta x_i + \theta v_i + \tau(x_i v_i) + \phi s_i$
Stratified on subgroup with adjusting propensity score	$Logit(p_{ki}) = \ln\left(\frac{p_{ki}}{1-p_{ki}}\right) = \alpha_k + \theta_k v_{ki} + \phi_k s_{ki}$	$\lambda_k(t_j) = \lambda_{k0}(t_j)\exp(\alpha_k + \theta_k v_{ki} + \phi_k s_{ki})$	$y_{ki} = \alpha_k + \theta_k v_{ki} + \phi_k s_{ki}$

A SAS Program for a Model-Based Stratification Method

Xiaoli Lu

VA Cooperative Studies Program, Perry Point, MD 21902

RESULTS

Data 1:

A randomized clinical trial was performed to quantitatively describe the relationship between alcohol consumption and cardiovascular diseases via an intervention program. Among 641 non-alcohol dependent moderate to heavy drinkers enrolled in the study, 320 were randomized to the intervention group and 321 to the control group. The outcomes in the analysis were ECG voltages (mV) in the precordial leads with the voltage dichotomized into low (< 35 mV) and high (≥ 35 mV) classes; the testing variable was alcohol consumption; the subgroup factor was subject hypertension status; other covariates included important demographic and clinical baseline factors.

Alcohol Consumption Intervention on Cardiac Conditions

Analysis Approach	Logistic Regression	θ	SE	Statistic	p	OR	95% C.I.
Model with only subgroup factor	Unadjusted	0.4741	0.3584	1.75	0.1859	1.607	0.796 – 3.243
	Hypertension Adjusted	**0.4838**	**0.3592**	**1.81**	**0.1780**	**1.622**	**0.802 – 3.280**
Model stratified on subgroup factor	Hypertension Stratum	0.0194	0.5190	0.00	0.7902	1.020	0.369 – 2.820
	No Hypertension Stratum	0.9200	0.5296	3.02	0.0823	2.509	0.889 – 7.085
	Weighted on Variance	**0.4606**	**0.3707**	**1.24***	**0.2140**	**1.585**	**0.766 – 3.278**
	Weighted on N	**0.5427**	**0.3768**	**1.44***	**0.1498**	**1.721**	**0.822 – 3.601**
Model with subgroup factor and other covariates	Hypertension Unadjusted	0.3808	0.5915	0.41	0.5197	1.463	0.459 – 4.665
	Hypertension Adjusted	**0.3851**	**0.6142**	**0.39**	**0.5307**	**1.470**	**0.441 – 4.898**
Model with subgroup factor and other covariates stratified on subgroup factor	Hypertension Stratum	0.9849	1.0741	0.84	0.3592	2.678	0.326 – 21.979
	No Hypertension Stratum	0.1433	1.0031	0.02	0.8864	1.154	0.162 – 8.243
	Weighted on Variance	**0.5354**	**0.7331**	**0.73***	**0.4652**	**1.708**	**0.406 – 7.187**
	Weighted on N	**0.4664**	**0.7429**	**0.63***	**0.5301**	**1.594**	**0.372 – 6.839**
Model with subgroup factor and propensity score	Hypertension Unadjusted	0.2783	0.4706	0.35	0.5544	1.321	0.525 – 3.322
	Hypertension Adjusted	**0.2896**	**0.4706**	**0.38**	**0.5383**	**1.336**	**0.531 – 3.360**
Model with subgroup factor and propensity score stratified on subgroup factor	Hypertension Stratum	- 0.1656	0.6906	0.06	0.8105	0.847	0.219 – 3.280
	No Hypertension Stratum	0.7097	0.6621	1.15	0.2838	2.033	0.555 – 7.444
	Weighted on Variance	**0.2905**	**0.4779**	**0.61***	**0.5433**	**1.337**	**0.524– 3.412**
	Weighted on N	**0.3736**	**0.4865**	**0.77***	**0.4425**	**1.453**	**0.560 – 3.770**
Cochran-Mantel-Haenszel Method				**1.84**	**0.1754**	**1.630**	**0.803 – 3.307**

Alcohol Consumption Intervention on ECG Voltages

Analysis Approach	ANCOVA	θ	SE	Statistic	p
Model with only subgroup factor	Unadjusted	0.0967	0.0769	1.26	0.2096
	Hypertension Adjusted	**0.1060**	**0.0766**	**1.39**	**0.1668**
Model stratified on subgroup factor	Hypertension Stratum	0.0508	0.1266	0.40	0.6890
	No Hypertension Stratum	0.1420	0.0960	1.48	0.1403
	Weighted on Variance	**0.1087**	**0.0765**	**1.42***	**0.1553**
	Weighted on N	**0.1060**	**0.0766**	**1.38***	**0.1664**
Model with subgroup factor and other covariates	Hypertension Unadjusted	0.0530	0.0617	0.86	0.3910
	Hypertension Adjusted	**0.0554**	**0.0616**	**0.90**	**0.3694**
Model with subgroup factor and other covariates stratified on subgroup factor	Hypertension Stratum	0.1743	0.1166	1.49	0.1380
	No Hypertension Stratum	- 0.0416	0.0722	0.58	0.5651
	Weighted on Variance	**0.0182**	**0.0614**	**0.30***	**0.7664**
	Weighted on N	**0.0413**	**0.0631**	**0.65***	**0.5129**
Model with subgroup factor and propensity score	Hypertension Unadjusted	0.0614	0.0889	0.69	0.4903
	Hypertension Adjusted	**0.0652**	**0.0882**	**0.74**	**0.4598**
Model with subgroup factor and propensity score stratified on subgroup factor	Hypertension Stratum	0.0332	0.1451	0.23	0.8195
	No Hypertension Stratum	0.0912	0.1116	0.82	0.4151
	Weighted on Variance	**0.0696**	**0.0885**	**0.79***	**0.4311**
	Weighted on N	**0.0689**	**0.0885**	**0.78***	**0.4360**

RESULTS CONTINUED

Data 2:

A registry data was used to identify the effect of thyroid hormone supplement with or without insulin by controlling for other covariates. Because controlled clinical trials were not available in this population, 40,124 brain-dead organ donors and 12,461 heart recipients in the 10-year period, 2000 to 2009. The outcomes for the analysis were heart procurement success rate, five year heart transplant survival, and number of organs procured and transplanted; the testing variable was thyroid hormone therapy; the subgroup factor was insulin use; other covariates included important demographic and clinical factors.

Thyroid Hormone Therapy on Heart Procurement

Analysis Approach	Logistic Regression	θ	SE	Statistic	p	OR	95% C.I.
Model with only subgroup factor	Unadjusted	- 0.4837	0.0223	387.5	<1E-8	0.645	0.617 – 0.674
	Insulin Adjusted	**- 0.4512**	**0.0233**	**376.4**	**<1E-8**	**0.637**	**0.608 – 0.667**
Model stratified on subgroup factor	Insulin Stratum	- 0.4548	0.0312	212.5	<1E-8	0.635	0.597 – 0.675
	No Insulin Stratum	- 0.4466	0.0349	163.7	<1E-8	0.640	0.597 – 0.685
	Weighted on Variance	**- 0.4512**	**0.0233**	**19.4***	**<1E-8**	**0.637**	**0.608 – 0.667**
	Weighted on N	**- 0.4513**	**0.0233**	**19.4***	**<1E-8**	**0.637**	**0.608 – 0.667**
Model with subgroup factor and other covariates	Insulin Unadjusted	- 0.2423	0.0280	74.8	<1E-8	0.785	0.743 – 0.829
	Insulin Adjusted	**- 0.2305**	**0.0289**	**63.6**	**<1E-8**	**0.794**	**0.750 – 0.840**
Model with subgroup factor and other covariates stratified on subgroup factor	Insulin Stratum	- 0.2104	0.0375	31.4	2.1E-8	0.810	0.753 – 0.872
	No Insulin Stratum	- 0.2479	0.0469	27.9	1.3E-7	0.780	0.712 – 0.856
	Weighted on Variance	**- 0.2250**	**0.0293**	**7.7***	**<1E-8**	**0.799**	**0.754 – 0.846**
	Weighted on N	**- 0.2264**	**0.0294**	**7.7***	**<1E-8**	**0.797**	**0.753 – 0.845**
Model with subgroup factor and propensity score	Insulin Unadjusted	- 0.1962	0.0224	64.5	<1E-8	0.822	0.783 – 0.862
	Insulin Adjusted	**- 0.2176**	**0.0251**	**75.1**	**<1E-8**	**0.804**	**0.766 – 0.845**
Model with subgroup factor and propensity score stratified on subgroup factor	Insulin Stratum	- 0.2747	0.0329	69.8	<1E-8	0.760	0.712 – 0.810
	No Insulin Stratum	- 0.1257	0.0396	10.0	0.0016	0.882	0.816 – 0.953
	Weighted on Variance	**- 0.2139**	**0.0253**	**8.5***	**<1E-8**	**0.807**	**0.768 – 0.849**
	Weighted on N	**- 0.2111**	**0.0253**	**8.3***	**<1E-8**	**0.810**	**0.770– 0.851**
Cochran-Mantel-Haenszel Method				**379.0**	**<1E-8**	**0.637**	**0.608– 0.667**

Thyroid Hormone Therapy on Heart Graft Five Year Survival

Analysis Approach	Cox PH Regression	θ	SE	Statistic	p	HR	95% C.I.
Model with only subgroup factor	Unadjusted	- 0.0284	0.0443	0.4104	0.5218	0.972	0.891– 1.060
	Insulin Adjusted	**- 0.0386**	**0.0454**	**0.6998**	**0.4028**	**0.962**	**0.879– 1.053**
Model stratified on subgroup factor	Insulin Stratum	- 0.1316	0.0646	4.1490	0.0417	0.877	0.772 – 0.995
	No Insulin Stratum	0.0652	0.0677	0.9268	0.3357	1.067	0.935 – 1.219
	Weighted on Variance	**- 0.0378**	**0.0467**	**0.8090***	**0.4185**	**0.963**	**0.879 – 1.055**
	Weighted on N	**- 0.0417**	**0.0468**	**0.8916***	**0.3726**	**0.959**	**0.875 – 1.051**
Model with subgroup factor and other covariates	Insulin Unadjusted	- 0.0173	0.0463	0.1388	0.7095	0.983	0.898 – 1.076
	Insulin Adjusted	**- 0.0266**	**0.0481**	**0.3071**	**0.5794**	**0.974**	**0.886 – 1.070**
Model with subgroup factor and other covariates stratified on subgroup factor	Insulin Stratum	- 0.1020	0.0669	2.3208	0.1277	0.903	0.792 – 1.030
	No Insulin Stratum	0.0558	0.0707	0.6230	0.4299	1.057	0.921 – 1.215
	Weighted on Variance	**- 0.0275**	**0.0486**	**0.5650***	**0.5721**	**0.973**	**0.885 – 1.070**
	Weighted on N	**- 0.0299**	**0.0486**	**0.6154***	**0.5383**	**0.971**	**0.882 – 1.068**
Model with subgroup factor and propensity score	Insulin Unadjusted	0.4815	0.0887	29.4815	5.6E-8	1.618	1.360 – 1.926
	Insulin Adjusted	**0.4701**	**0.0894**	**27.6499**	**1.5E-7**	**1.600**	**1.343 – 1.907**
Model with subgroup factor and propensity score stratified on subgroup factor	Insulin Stratum	0.3901	0.1151	11.4891	0.0007	1.477	1.179 – 1.851
	No Insulin Stratum	0.5764	0.1452	15.7596	7.4E-5	1.780	1.339 – 2.366
	Weighted on Variance	**0.4620**	**0.0902**	**5.1220***	**3.4E-7**	**1.588**	**1.330 – 1.894**
	Weighted on N	**0.4750**	**0.0633**	**7.5098***	**<1E-8**	**1.608**	**1.421 – 1.820**

CONCLUSIONS

- A SAS program was developed to implement the model based stratification method.
- The models showed different results with or without adjusting covariates; further the intervention effects varied from the propensity score adjusted models to the vector of covariates adjusted ones.
- Model based subgroup adjustment showed similar results as the stratified analysis based on either the Cochran-Mantel-Haenszel method or the model-based stratification method when sample size was large as shown in the registry data analysis
- Stratification weights based on inverse variance or sample size yielded similar common estimates across the subgroup strata.
- The results from the model-based stratification method were consistent with those from the model based covariate adjustment and traditional stratification method; similar findings were also observed using Cox regression and ANCOVA models.

Disclaimer : The views expressed in this report are those of the author and do not necessarily reflect the position or policy of the U.S. Department of Veterans Affairs or the United States Government

Contact Info: Xiaoli.Lu@va.gov

SAS® GLOBAL FORUM 2018

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF