Comparison of Statistical Methods (Univariate vs. Multivariate) Used to Evaluate Biomarker Changes Following Reduction in Exposure to Cigarette Smoke

Qiwei Liang, Jingzhu Wang, Yuxi Zhao, and Ryan Black

Altria Client Services LLC, Center for Research and Technology, 601 East Jackson Street, Richmond, VA 23219, USA

SRNT 24th Annual Meeting, February 21 - 24, 2018, Baltimore, MD, USA

This poster may be accessed at www.altria.com/ALCS-Science

Altria
Altria Client Services

Abstract

A single primary clinical outcome measure is generally used to evaluate the effectiveness of pharmaceutical products. In contrast, several clinical outcome measures (biomarkers) are needed to assess cigarette smoking exposure reduction. In most cases however, each biomarker difference is independently determined with a univariate statistical model. Since changes in biomarkers are generally correlated, it may be more appropriate to use a multivariate model under certain circumstances. This may increase the power to detect group difference.

We developed a multivariate linear mixed model to examine multiple biomarkers. We analyzed data from a clinical study designed to assess exposure reduction in adult cigarette smokers who switched to an electronically heated cigarette smoking system (EHCSS). The smokers were randomized into five study groups and several biomarkers of exposure including nicotine equivalents (NE), carboxyhemoglobin (CoHb), total 1-hydroxypyrene (1-OHP), 3-hydroxypropylmercapturic acid (3-HPMA) and S-phenylmercapturic acid (SPMA) were measured daily for 8 days.

We began our assessment with a model designed for two biomarkers (NE and COHb) with and without covariates. In the model without covariates, the dependent variables were NE and COHb, with the fixed effect of study group, the random effect of time and subject, and the interaction between study group and time. In the model with covariates, age, gender, daily cigarette consumption and BMI were considered. We then expanded the model to assess five biomarkers (NE, COHb, 1-OHP, 3-HPMA and SPMA). The results of the single univariate model were compared to the results of the multivariate model.

The multivariate linear mixed model provides additional insight on biomarker relationships and factors. Unlike the single univariate model, the multivariate model estimates study effect on multiple biomarkers simultaneously and therefore do not need p value adjustment for multiplicity. This model can also estimate either the study effect for each biomarker or an overall effect on all biomarkers.

Study Purpose

- Develop a multivariate statistical approach to examine multiple biomarkers simultaneously in clinical trials for tobacco harm reduction.
- Compare results between univariate and multivariate approaches

Introduction & Background

In clinical studies for drugs and devices, the primary objective is generally to look at a single clinical endpoint. In clinical studies for tobacco harm reduction, however, the primary objective is often to look at multiple outcome measures (e.g. several biomarkers)

When each outcome of multiple outcomes is analyzed separately, the probability of obtaining statistically significant results by chance may increase [2,3]. The probability of finding at least one false significant result is called the familywise error rate (FWER). The FWER should be maintained at an acceptable level, usually 5% for the statistical to be appropriate. To maintain the FWER at the 5% level, adjustments needs to be made to the p-values if each outcome is analyzed separately.

Because multivariate approach allows the test of multiple outcomes simultaneously, no p value adjustment for multiple outcomes is needed. It also take into account the correlation between multiple outcomes and therefore may increase the statistical power to detect group difference [4].

Definitions

Univariate model: Model with one outcome measure

Multivariate model: Model with more than one outcome measure

Linear mixed mode: Linear model with both fixed and random effects

Fixed effect: the levels of an effect are not for inference of a larger population

Random effect: the levels of an effect are for inference of a larger population

Data

The data are from a previous Altria study [1] in which the reduction of exposure to several cigarette smoking constituents was examined in 100 adult cigarette users. The cigarette smokers were randomized to five study groups:

A: continued to smoke conventional cigarette Type I

B: continued to smoke conventional cigarette Type II

C: used an electrically heated cigarette smoking system I

D: used an electrically heated cigarette smoking system II

E: stopped use of tobacco product

Several biomarkers of exposure were measured every day for 8 days after randomization. The biomarkers include nicotine equivalents (NE), carboxyhemoglobin (COHb),1-hydroxypyrene (1-OHP), S-phenylmercapturic acid (S-PMA) and 3-hydroxypropylmercapture acid (3-HPMA).

Analyses

Univariate Linear Mixed Model Approach

- $Y_{1it}=\mu_{10}+\mu_{1B}I(T_i=B)+\mu_{1C}I(T_i=C)+\mu_{1D}I(T_i=D)+\mu_{1E}I(T_i=E)+a_{1i}t+\beta_{1B}I(T_i=B)t+\beta_{1C}I(T_i=C)t+\beta_{1D}I(T_i=D)t+\beta_{1E}I(T_i=E)t+Y_{1it}+\varepsilon_{1it}$
- $Y_{2it}=\mu_{20}+\mu_{2B}I(T_i=B)+\mu_{2C}I(T_i=C)+\mu_{2D}I(T_i=D)+\mu_{2E}I(T_i=E)+a_{2i}t+\beta_{2B}I(T_i=B)t+\beta_{2C}I(T_i=C)t+\beta_{2D}I(T_i=D)t+\beta_{2E}I(T_i=E)t+Y_{2it}+\varepsilon_{2it}$

Where Y_{1it} is the measure of biomarker 1 in Subject i at time t, μ_{10} is the model intercept for the biomarker, T_i is the fixed effect for study group, with μ_{1B}, μ_{1C}, μ_{1D} and μ_{1E} for product group B, C, D and E compared with Group A, respectively; $a_{1i}t$ is the time effect, β_{1B}, β_{1C}, β_{1D} and β_{1E} is the interaction for each group with time, Y_{1it} is the random effect and ε_{1it} is the error term. Y_{2it} is the measure of biomarker 2 in Subject i at time t, other model terms can be described similarly as those in Y_{1it}.

In a univariate approach, the above two models will be run separately.

Multivariate Linear Mixed Model Approach

- $Y_{1it}=\mu_{10}+\mu_{1B}I(T_i=B)+\mu_{1C}I(T_i=C)+\mu_{1D}I(T_i=D)+\mu_{1E}I(T_i=E)+a_{1i}t+\beta_{1B}I(T_i=B)t+\beta_{1C}I(T_i=C)t+\beta_{1D}I(T_i=D)t+\beta_{1E}I(T_i=E)t+Y_{1it}+\varepsilon_{1it}$
- $Y_{nit}=\mu_{n0}+\mu_{nB}I(T_i=B)+\mu_{nC}I(T_i=C)+\mu_{nD}I(T_i=D)+\mu_{nE}I(T_i=E)+a_{ni}t+\beta_{nB}I(T_i=B)t+\beta_{nC}I(T_i=C)t+\beta_{nD}I(T_i=D)t+\beta_{nE}I(T_i=E)t+Y_{nit}+\varepsilon_{nit}$

Where Y_{1it} is the measure of biomarker 1 in Subject i at time t, μ_{10} is the model intercept for the biomarker, T_i is the fixed effect for study group, with μ_{1B}, μ_{1C}, μ_{1D} and μ_{1E} for product group B, C, D and E compared with Group A, respectively; $a_{1i}t$ is the time effect, β_{1B}, β_{1C}, β_{1D} and β_{1E} is the interaction for each group with time, Y_{1it} is the random effect and ε_{1it} is the error term. Y_{nit} is the measure of biomarker n in Subject i at time t, other model terms can be described similarly as those in Y_{1it}.

The multivariate approach accounts for the correlation between the two biomarkers. The biomarkers are analyzed simultaneously. Y_{1it} and Y_{nit} are correlated and so are ε_{1it} and ε_{nit}. We need to create a new class variable to describe the relationship between the n biomarkers. In addition, a Kronecker product covariance structures is used to model the multivariate covariance structure. Age, gender and BMI were included as covariates in some models.

Summary of Test of Fixed Effect

	Model Type			
Effect	Univariate		Multivariate (2 Biomarkers)	Multivariate (5 Biomarkers)
	NE (mg/24 h)	COHb (AUC 7-23)		
Product	<0.0001	<0.0001	<0.001	<0.001
Study Day	<0.0001	0.0652	<0.001	<0.001
Product x Day	<0.0001	0.2604	<0.001	<0.001

Study Product Effect: Univariate vs. Multivariate

	Univariate Model[1]				Multivariate Model[2] (2 Biomarkers)			
	NE (mg/24 h)		COHb (%xh)		NE (mg/24 h)		COHb (%xh)	
Effect	Est. (95%CI)	P value	Est. (95%CI)	P value	Est. (95%CI)	P value	Est. (95%CI)	P value
B vs A	-6.96	<0.0001	-2.02-	<0.0001	-6.85	<0.0001	-2.06	<0.0001
	-8.60,-5.31		-2.53,-1.51		-9.04,-4.66		-2.55,-1.57	
C vs A	-8.94	<0.0001	-4.96	<0.0001	-7.32	<0.0001	-5.00	<0.0001
	-10.61,-7.26		-5.47,-4.45		-9.53,-5.11		-5.50,-4.50	
D vs A	-8.95	<0.0001	-4.84	<0.0001	-7.86	<0.0001	-4.88	<0.0001
	-10.60,-7.29		-5.35,-4.33		-9.98,-5.57		-5.38,-4.38	
E vs A	-12.18	<0.0001	-5.38	<0.0001	-10.08	<0.0001	-5.37	<0.0001
	-13.83,-10.53		-5.89, -4.87		-12.3,-7.87		-5.87,-4.87	

[1]Unstructured covariance
[2]Unstructured@compound covariance

Product Effect under Multivariate Model of 5 Biomarkers without Covariates

	Biomarker									
	NE (mg/24 h)		COHb (%xh)		1-OHP (ng/24 h)		SPMA (µg/24 h)		3-HPMA (µg/24 h)	
Effect	Est. (95%CI)	P value	Est. (95%CI)	P value	Est. (95%CI)	P value	Est. (95%CI)	P value	Est. (95%CI)	P value
B vs. A	-9.62	<0.001	-1.69	0.0001	-74.16	<0.0001	-1.82	0.0478	-7.44	0.0001
	-13.61, -5.63		–2.51,-0.87		-109.22,-39.1		-3.60,-0.04		-11.2,-3.86	
C vs. A	-12.47	<0.001	-4.95	<0.0001	-120.56	<0.0001	-7.07	<0.0001	-18.49	<0.0001
	-16.02, -8.92		-5.81,-4.09		-154.27,-86.89		-8.89,-5.25		-22.31,-14.67	
D vs. A	-10.76	<0.001	-4.68	<0.0001	-110.77	<0.0001	-5.53	<0.0001	-15.20	<0.0001
	-13.89, -7.67		-5.50,-3.86		-146.14,-75.4		-7.33.-3.73		-19.00,-11.40	
E vs. A	-17.65	<0.001	-5.60	<0.0001	-107.92	<0.0001	-5.58	<0.0001	-20.53	<0.0001
	-20.86, -14.44		-6.48,-4.72		-144.41, -71.44		-7.4,-3.76		-24.45,-16.61	

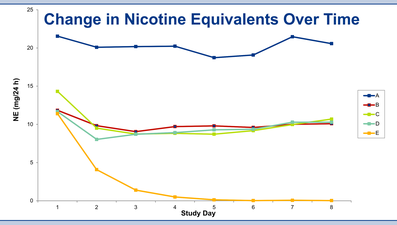
Unstructured@compound covariance

Product Effect under Multivariate Model of 5 Biomarkers with Covariates

	Biomarker									
	NE (mg/24 h)		COHb (%xh)		1-OHP (ng/24 h)		SPMA (µg/24 h)		3-HPMA (µg/24 h)	
Effect	Est.	P value	Est.	P value	Est.	P value	Est.	P value	Est.	P value
B vs. A	-9.62	<0.001	-1.67	0.0001	-71.57	<0.0001	-1.80	0.0478	-7.37	0.0001
C vs. A	-12.57	<0.001	-5.01	<0.0001	-121.95	<0.0001	-7.05	<0.0001	-18.64	<0.0001
D vs. A	-10.70	<0.001	-4.64	<0.0001	-109.00	<0.0001	-5.49	<0.0001	-15.08	<0.0001
E vs. A	-17.76	<0.001	-5.63	<0.0001	-110.37	<0.0001	-5.69	<0.0001	-20.82	<0.0001

Unstructured@compound covariance

Change in Nicotine Equivalents Over Time



Summary

- Multivariate linear mixed model was used to examine reduction in levels of biomarkers of cigarette smoking, with models for two biomarkers and more than two biomarkers. The statistical estimates of the product effect were very close between the univariate models and the multivariate models for 2 biomarkers. The estimates were comparable between the univariate models and the multivariate models of 5 biomarkers. The product effects were all highly statistically significant in all models.
- The multivariate approach bypass multiplicity adjustment of p values for testing multiple biomarkers in univariate models. It may also improve statistical power for detecting product group difference.
- This is a preliminary work. Computer simulation will be used to further compare the performance of the two approaches in the next step. In addition to linear mixed model, alternative models such as generalized estimation equation will also be used.

References

[1] Roethig J., Zedler B., Kinser R., Feng S., Nelson B., Liang Q., 2007. Short-term clinical exposure evaluation of a second-generation electrically heated cigarette smoking system. Journal of Clinical Pharmacology, 47(4), 518-530.

[2] Alosh M., Bretz F., Huque M., 2014. Advanced multiplicity adjustment methods in clinical trials, Stat. Med. 33 (4) 693–713.

[3] Bagiella E., 2009. Clinical trials in rehabilitation: single or multiple outcomes? Arch. Phys. Med. Rehabil. 90 (11) S17–S21.

[4] Reyes A., De Los, S.M.A. Kundey, M. Wang, 2011. The end of the primary outcome measure: a research agenda for constructing its replacement, Clin. Psychol. Rev. 31 (5)

No conflict of interest