Paper PR04

Assessing Within- and Between-Subject Effects with Correlated Clinical Data

Marla Husnik, Fred Hutchinson Cancer Research Center, Seattle, WA
Eugene Huang, Emory University, Atlanta, GA
Grant Colfax, San Francisco Department of Health, San Francisco, CA

ABSTRACT

Various statistical methods can be used to analyze correlated data from a clinical study. At baseline of the EXPLORE study, a multi-site HIV prevention clinical trial conducted in the U.S., data from the most recent sexual episode with up to three of the most recent partners were collected. We were interested in the relationship between substance use and risky sexual behavior. To address the within-individual effects, we first used non-linear mixed models (Proc NLMIX) and found computational difficulties. Then, we used conditional logistic regression (Proc PHREG) for the within-individual effects, and generalized estimating equations (Proc GENMOD) for the between-individual effects. In this paper, we show examples of these techniques using SAS® and compare the methods.

INTRODUCTION

The EXPLORE study was a multi-site, two-arm randomized controlled screening trial. It tested the efficacy of a ten-session behavioral counseling program in preventing the acquisition of HIV among 4,295 HIV-negative men who have sex with men. Study recruitment occurred between January 1999 and February 2001. Participants completed semi-annual follow-up visits through July 2003. At each follow-up visit staff collected behavioral risk data and blood specimens that were tested for HIV. The behavioral risk data collected included frequency of substance use and sexual behaviors in the past 6 months as well as the last sexual episodes with up to three partners. See Koblin et al. for additional details [1].

From this study, we conducted an analysis examining the relationship between baseline substance use and sero-discordant (with an HIV-positive or HIV-unknown partner) unprotected anal sex (SDUA). Previous studies have found associations between episode-specific use of substances such as amphetamines, poppers, alcohol, cocaine, and ecstasy with sexual risk behavior [2]. Few studies, however, have examined the effects of both background substance use and episode-specific use on sexual risk behavior [2]. With the EXPLORE study data we were able to address these relationships by modeling the within- and between-individual effects of substance use on SDUA.

This paper discusses the approaches considered for our analysis. The first approach uses non-linear mixed-effects models to address within-individual effects based on participants' sexual episode-specific responses. We eventually abandoned this approach due to long computation times and difficulties with numerical estimations. The second approach employs conditional logistic regression and generalized linear models. These methods provided the opportunity to address both within- and between-individual effects, respectively. First, we describe the variables and data structure used in the analysis. Second, we show an example of the SAS code used for a non-linear mixed-effects model (using proc NLMIX). Third, we show the SAS code and output for the final multivariate conditional logistic regression model (using proc PHREG) for the within-individual effects. Fourth, we show the SAS code and output used for the final multivariate generalized estimating equation model (using proc GENMOD) for the between-individual effects. Last, we compare the methods.

EXPLORE STUDY DATA

Participants reported their demographic data such as age, race, ethnicity, education level, and household income at the baseline visit. In addition, each participant completed a risk assessment via audio-computer assisted interview format. This assessment asked questions pertaining to the amount and frequency of alcohol use, frequency of drug use including marijuana, poppers or inhaled nitrites, smoked crack or cocaine, snorted or sniffed cocaine, swallowed, snorted or smoked amphetamines such as speed, crystal, or crank, snorted or smoked heroin, hallucinogens, and any other non-injectable drugs in the last 6 months. Sexual behavior, including the number of male sex partners, and depression were also assessed. For the episode-specific portion of the questionnaire, participants reported their last three most recent partners' HIV status, whether he was a steady partner, the attractiveness of the partner, his age, location of sex with the partner, number of times having had sex with the partner in the last 6 months and length of time having sex with the partner. For each partner, the participant also reported information about sexual behavior with his partner, whether or not he or his partner was drinking alcohol within 2 hours before or during having sex, and specific drug usage immediately before or during having sex with his partner. We constructed a risky sex (SDUA) variable from participants' reports of insertive and receptive anal sexual behavior, without a condom.

NON-LINEAR MIXED-EFFECTS MODELS

Our first approach entailed constructing a variety of logistic mixed-effects models (using SAS procedure proc NLMIX) for the outcome of risky sex (SDUA). The fixed effects in these models included both background demographic data as well as the episode-specific covariates. We assumed normally distributed random effects to accommodate the correlation across episodes within an individual. Such a model is called logistic-normal with random effects [3,4]. In the models, risksexm denotes a 0/1 (no/yes) indicator for whether the participant (i=1,…,4,295) had SDUA with his partner. The data is structured so that there are up to three records per participant; one record for each reported partner (j = 1,2,3). The logistic model is as follows:

$$risksexm_{ij}|u_i \sim \text{Binomial}(1,p_{ij})$$

$$\eta_{ij} = log\ (p_{ij}/(1- p_{ij})) = \beta_0 + \beta_1 c_1 + \ldots + \beta_n c_n + u_i$$

$c_1,\ldots,c_n$ denotes the 1,…,n fixed covariates, p_{ij} is the probability of SDUA for participant (i) with his partner (j), and u_i denotes the random effect assumed to be iid $N(0,\sigma^2)$.

SAS SYNTAX

A macro for a multivariate model generated is shown below.

```
%macro
nlmix21(c1,c2,c3,c4,c5,c6,c7,c8,c9,c10,c11,c12,c13,c14,c15,c16,c17,c18,c19,c20,c21,
covname1, covname2, covname3, covname4, covname5, covname6, covname7,
covname8, covname9, covname10, covname11, covname12, covname13,
covname14, covname15, covname16, covname17, covname18, covname19,
covname20, covname21);


title1 "EXPLORE - Baseline Paper F";
title2 "Non-linear Mixed Model of Risky Sex on";
title3 "&covname1, &covname2, &covname3, &covname4,";
title4 "&covname5, &covname6, &covname7, &covname8,";
title5 "&covname9, &covname10, &covname11, &covname12,";
title6 "&covname13, &covname14, &covname15, &covname16,";
title7 "&covname17, &covname18, &covname19, &covname20,";
title8 "and &covname21";

proc nlmixed data = riskpart;
parms     beta0=-2.99
          beta1=0.099  beta2=0.877
          beta3=0.547  beta4=0.202
          beta5=0.115  beta6=-0.018
          beta7=-0.072 beta8=-0.035
          beta9=0.086  beta10=0.026
          beta11=0.118 beta12=0.326
          beta13=0.035 beta14=0.638
          beta15=0.308 beta16=0.952
          beta17=0.429 beta18=-1.42
          beta19=0.188 beta20=0.560
          beta21=0     s2u=1.393;
eta =       beta0
          + beta1*&c1 + beta2*&c2
          + beta3*&c3 + beta4*&c4
          + beta5*&c5 + beta6*&c6
          + beta7*&c7 + beta8*&c8
          + beta9*&c9 + beta10*&c10
          + beta11*&c11 + beta12*&c12
          + beta13*&c13 + beta14*&c14
```

```
            + beta15*&c15 + beta16*&c16
            + beta17*&c17 + beta18*&c18
            + beta19*&c19 + beta20*&c20
            + beta21*&c21 + u;

expeta = exp(eta);
p = expeta/(1+expeta);
model risksexm ~ binomial(1,p);
random u ~ normal(0,s2u) subject=ptid;
run;
%mend;
```

The PARMS statement defines the parameters and their starting values. (Note: Zero was used for the initial starting values for smaller models. As the models grew in complexity, however, we used starting values found from simpler models). P corresponds to p_{ij}, and MODEL defines the conditional distribution as Binomial, where n=1. RANDOM defines the distribution of the random effect (u), with SUBJECT=ptid as the random effect variable.

We faced several barriers using this procedure. First, as the models became more complex (with additional covariates), we observed increasing computation time. Second, we experienced difficulties with convergence for models including covariates with low prevalence or those where the initial values were not close to the final estimates. Third, in order to deal with the convergence problems, it was necessary to run less complex models, obtain estimates for initial values, and then re-run the composite model. Due to long computation time, and the unpredictability of the procedure in meeting the convergence criterion, constructing the best multivariate mixed-effects model became extremely time consuming and challenging. These issues, particularly the last one, motivated us to consider alternative analysis methods to investigate the within-and between-individual effects.

WITHIN-INDIVIDUAL EFFECTS USING CONDITIONAL LOGISTIC REGRESSION

As an alternative to mixed-effects modeling, we used conditional logistic regression (using proc PHREG) to model the within-individual effects of SDUA on substance use. Note that there is no SAS procedure designed specifically to carry out conditional logistic regression. The proc PHREG procedure can be utilized if given an appropriately constructed data set. To this end, we restricted the analysis to only those participants who reported SDUA at least once, and reported no SDUA at least once. In effect, each individual served as his own control. We used the discrete logistic model and created a stratum for each participant's matched set of records [5]. For example, if he reported SDUA with one partner and no SDUA with only one other partner, then his matched set would be 1:1. If he reported SDUA with 2 partners and no SDUA with one partner, then his matched set would be 2:1. We constructed a dummy survival time variable to reflect the same value for the SDUA records in the matched set and a later survival time for the non-SDUA records.

LISTING OF DATA USED IN CONDITIONAL LOGISTIC REGRESSION ANALYSIS

The indicator variable names (0=no, 1=yes) and labels for the data set used in this analysis are shown below.

Table 1: Indicator Variable Names and Labels in Conditional Logistic Regression Analysis

VARIABLE	VARIABLE LABEL
ptid	Participant ID
risksexm	Risky sex (SDUA)
drinks12m	1–2 drinks just before or during sex
drinks35m	3-5 drinks just before or during sex
drinks6ormorem	>= 6 drinks just before or during sex
du_osm	Popper (amyl nitrites), snorted or sniffed cocaine, or amphetamines immediately before or during sex
drugyespartm	Participant reported that he did know that his partner used drugs immediately before or during sex
drugunkpartm	Participant reported that he did not know if his partner used drugs immediately before or during sex
alcyespartm	Participant reported that he did know that his partner consumed alcohol within 2 hours before or during sex
alcunkpartm	Participant reported that he did not know if his partner consumed alcohol within 2 hours before or during sex

VARIABLE	VARIABLE LABEL
page2635m	Partner age: 26–35 years
page3645m	Partner age: 36–45 years
page46ormorem	Partner age: >= 46 years
steadym	Partner type: steady
nsteadym	Partner type: non-steady
locsexclubbathm	Location of sex with partner: sex club or bath house
locother2m	Location of sex with partner: other (includes hotel, bar, dance club, porn theater, video arcade, or other public place, including the street)
timesex2to5m	Number of times had sex with partner in the last 6 months: 2–5
timesex6ormorem	Number of times had sex with partner in the last 6 months: >= 6
time_sdua	Dummy time variable (1 if SDUA, 2 if non-SDUA)

Below is a portion of the first 20 records of the data set used for analysis.

Table 2: Data Set Example for Conditional Logistic Regression Analysis

ptid	risksexm	drinks12m	drinks35m	drinks6ormorem	…	timesex6ormorem	time_sdua
001	0	0	0	0	…	0	2
001	0	0	0	0	…	0	2
001	0	0	0	0	…	0	2
002	0	0	1	0	…	0	2
002	0	0	1	0	…	0	2
002	0	0	1	0	…	0	2
003	1	0	0	1	…	1	1
003	.	.	.	.	…	.	.
003	.	.	.	.	…	.	.
004	0	0	0	0	…	1	2
004	0	0	0	0	…	1	2
004	0	0	0	0	…	0	2
005	0	0	0	0	…	1	2
005	1	0	0	1	…	0	1
005	0	0	0	1	…	0	2
006	0	1	0	0	…	0	2
006	0	1	0	0	…	0	2
006	0	1	0	0	…	1	2
007	0	0	0	0	…	1	2
007	0	0	0	0	…	0	2

SAS SYNTAX

The SAS macro used for the analysis is shown below.

```
%macro condlog(outcome,dataset=,variables=,time=,strata=);
ods output ParameterEstimates = phreg_est;

proc phreg data =&dataset;
model &time*&outcome(0) = &variables / ties=discrete risklimits;
strata &strata;
run;

*** Print off the parameter estimates and ORs;
```

```
title "PHREG Parameter Estimates with ORs";
proc print data = phreg_est;
run;
ods output close;
%mend;
```

The &time variable (time_sdua) = 1 for the record(s) where the participant reported SDUA; 2 for no SDUA. &outcome denotes the outcome variable, risksexm. &variables are the covariates of interest. Ties=discrete indicates the discrete logistic model; risklimits request output of confidence limits for the hazards ratios. &strata indicates the variable for the stratum for each matched set; in this case the participant ID (ptid). The call for this macro is as follows:

```
%condlog(risksexm, dataset=riskpart2_condlog,
variables=
drinks12m drinks35m drinks6ormorem du_osm
drugyespartm drugunkpartm
alcyespartm alcunkpartm
page2635m page3645m page46ormorem
steadym nsteadym
locsexclubbathm locother2m
timesex2to5m timesex6ormorem,
time=time_sdua, strata= ptid);
```

RESULTS OF WITHIN- INDIVIDUAL EFFECTS

Results of the analysis using the within-individual covariates are shown below.

Table 3: Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
drinks12m	1	0.00387	0.15470	0.0006	0.9800	1.004	0.741	1.359
drinks35m	1	0.20786	0.15025	1.9139	0.1665	1.231	0.917	1.653
drinks6ormorem	1	0.88161	0.21976	16.0940	<.0001	2.415	1.570	3.715
du_osm	1	0.38074	0.16379	5.4037	0.0201	1.463	1.062	2.017
drugyespartm	1	0.41365	0.13290	9.6877	0.0019	1.512	1.166	1.962
drugunkpartm	1	0.43851	0.12653	12.0102	0.0005	1.550	1.210	1.987
alcyespartm	1	0.29101	0.13449	4.6823	0.0305	1.338	1.028	1.741
alcunkpartm	1	0.19713	0.14337	1.8906	0.1691	1.218	0.920	1.613
page2635m	1	-0.21930	0.10805	4.1193	0.0424	0.803	0.650	0.993
page3645m	1	-0.30868	0.13762	5.0307	0.0249	0.734	0.561	0.962
page46ormorem	1	-0.29328	0.21590	1.8452	0.1743	0.746	0.488	1.139
steadym	1	0.15663	0.14393	1.1842	0.2765	1.170	0.882	1.551
nsteadym	1	0.26791	0.16533	2.6259	0.1051	1.307	0.945	1.807
locsexclubbathm	1	0.07208	0.17545	0.1688	0.6812	1.075	0.762	1.516
locother2m	1	-0.24315	0.11995	4.1091	0.0427	0.784	0.620	0.992
timesex2to5m	1	0.17602	0.10012	3.0906	0.0787	1.192	0.980	1.451
timesex6ormorem	1	0.14686	0.15930	0.8499	0.3566	1.158	0.848	1.583

BETWEEN-INDIVIDUAL EFFECTS USING GENERALIZED ESTIMATING EQUATIONS

To model the between-individual effects of SDUA on substance use, we used generalized estimating equations (GEE) (using proc GENMOD) [3,6–7].

LISTING OF DATA USED IN GENERALIZED ESTIMATING EQUATIONS ANALYSIS

The indicator variable names (0=no, 1=yes) and labels for the data set used in this analysis are shown below.

All variables pertain to the participant enrolled in the study.

Table 4: Indicator Variable Names and Labels in Generalized Estimating Equations Analysis

VARIABLE	VARIABLE LABEL
ptid	Participant ID
risksexm	Risky sex (SDUA)
age2635m	Age: 26–35 years
age3645m	Age: 36–45 years
age46ormorem	Age: >= 46 years
blackm	Race: black
hispanicm	Race: hispanic
otherm	Race: other
hsm	Education: highschool or less
scm	Education: some college
cdm	Education: college degree
lt12m	Annual income: < $12,000
to30m	Annual income: $12,000–$29,999
to60m	Annual income: $30,000–$59,999
depressm	Depressed
numpart25m	Number of male sex partners in the last 6 months: 2–5
numpart69m	Number of male sex partners in the last 6 months: 6–9
numpart10m	Number of male partners in the last 6 months: >= 10
alclight6m	Alcohol use in the last 6 months: Light
alcmoderate6m	Alcohol use in the last 6 months: Moderate
heavy6m	Alcohol use in the last 6 months: Heavy
marj6lt1wkm	Marijuana use in the last 6 months: < 1 time/week
marj61wkmorem	Marijuana use in the last 6 months: >= 1 time/week
pop6lt1wkm	Popper (amyl nitrites) use in the last 6 months: < 1 time/week
pop61wkmorem	Popper (amyl nitrites) use in the last 6 months: >= 1 time/week
pcp6lt1wkm	Hallucinogen (PCP, “angel dust”, Special K, LSD, ecstasy) use in the last 6 months: < 1 time/week
pcp61wkmorem	Hallucinogen (PCP, “angel dust, Special K, LSD, ecstasy) use in the last 6 months: >= 1 time/week
coc6lt1wkm	Sniffed cocaine use in the last 6 months: < 1 time/week
coc61wkmorem	Sniffed cocaine use in the last 6 months: >= 1 time/week
spd6lt1wkm	Amphetamine (“speed”, “crystal”, “crank”) use in the last 6 months: < 1 time/week
spd61wkmorem	Amphetamine (“speed”, “crystal”, “crank”) use in the last 6 months: >= 1 time/week

VARIABLE	VARIABLE LABEL
crk6lt1wkm	Smoked crack cocaine: < 1 time/week
crk61wkmorem	Smoked crack cocaine: >= 1 time/week
her6lt1wkm	Smoked heroin: < 1 time/week
her61wkmorem	Smoked heroin: >= 1 time/week
injm	Any injectable drugs

Below is a portion of the first 20 records of the data set used for analysis.

Table 5: Data Set Example for Generalized Estimating Equations Analysis

ptid	risksexm	age2635m	age3645m	age46ormorem	…	her61wkmorem	injm
001	0	0	1	0	…	0	0
001	0	0	1	0	…	0	0
001	0	0	1	0	…	0	0
002	0	1	0	0	…	0	0
002	0	1	0	0	…	0	0
002	0	1	0	0	…	0	0
003	1	1	0	0	…	0	0
003	.	1	0	0	…	0	.
003	.	1	0	0	…	0	.
004	0	0	1	0	…	0	0
004	0	0	1	0	…	0	0
004	0	0	1	0	…	0	0
005	0	0	1	0	…	0	0
005	1	0	1	0	…	0	0
005	0	0	1	0	…	0	0
006	0	0	1	0	…	0	0
006	0	0	1	0	…	0	0
006	0	0	1	0	…	0	0
007	0	0	0	1	…	0	0
007	0	0	0	1	…	0	0

SAS SYNTAX

The SAS macro used was as follows:

```
%macro
gee(outcome=,dataset=,variables=,classvar=,dist=,link=,subject=,type=);
ods output GEEEmpPEst = geeparams;
proc genmod data = &dataset descending;
class &classvar;
model &outcome = &variables / dist = &dist link = &link;
repeated subject = &subject / type = &type;
ods listing;

*** Create data set that outputs the ORs for the parameter estimates;
title "GEE Model Parameters";
data geeparam_or;
set geeparams;
OR_est = exp(estimate);
```

```
OR_lower95 = exp(lowercl);
OR_upper95 = exp(uppercl);
run;
title "GEE Model Parameters with ORs";
proc print data = geeparam_or;
var       parm estimate lowercl uppercl
          OR_est OR_lower95 OR_upper95
          probz;
where parm ne "Intercept";
 run;
%mend;
```

The call for this macro is as follows:

```
%gee(outcome=risksexm, dataset=riskpart,
variables=
age2635m age3645m age46ormorem
blackm hispanicm otherm
hsm scm cdm
lt12m to30m to60m
depressm
numpart25m numpart69m numpart10m
alclight6m alcmoderate6m heavy6m
marj6lt1wkm marj61wkmorem
pop6lt1wkm pop61wkmorem
pcp6lt1wkm pcp61wkmorem
coc6lt1wkm coc61wkmorem
spd6lt1wkm spd61wkmorem
crk6lt1wkm crk61wkmorem
her6lt1wkm her61wkmorem
injm, classvar= ptid,dist=binomial,link=logit,subject=ptid, type=exch);
```

Since the outcome of interest, risksexm, is binary (0=no SDUA, 1=SDUA), we modeled the GEE using the logit link function; dist=binomial, link=logit. Likewise, the class variable = ptid since there were repeated measures on the participant level. &variables represent the covariates of interest. For this part of the analysis, the covariates were demographics, risk behaviors and substance use reported by the participant in the previous 6-month period. These data represented background levels of substance use and other personal characteristics related to the participant's sexual behavior. The REPEATED statement specifies the covariance structure of multivariate responses for the GEE model. In our case, SUBJECT=ptid. TYPE=exch indicates an exchangeable working structure for the correlation matrix on the responses of the subjects [3].

RESULTS OF BETWEEN-INDIVIDUAL EFFECTS

The results of the analysis for between-individual effects of substance use on SDUA are shown below.

Table 6: GEE Model Parameters with ORs

Obs	Parm	Estimate	LowerCL	UpperCL	OR_est	OR_lower95	OR_upper95	ProbZ
2	age2635m	0.1997	0.0229	0.3766	1.22108	1.02315	1.45729	0.0269
3	age3645m	0.2597	0.0625	0.4569	1.29653	1.06446	1.57919	0.0099
4	age46ormorem	0.3668	0.1152	0.6184	1.44316	1.12215	1.85601	0.0043
5	blackm	-0.0458	-0.2981	0.2064	0.95522	0.74226	1.22928	0.7219
6	hispanicm	0.0191	-0.1498	0.1880	1.01927	0.86089	1.20679	0.8247
7	otherm	-0.0319	-0.2910	0.2271	0.96858	0.74755	1.25495	0.809
8	hsm	0.6145	0.3821	0.8469	1.84874	1.46538	2.33239	<.0001
9	scm	0.3601	0.1900	0.5302	1.43347	1.20930	1.69919	<.0001

Obs	Parm	Estimate	LowerCL	UpperCL	OR_est	OR_lower95	OR_upper95	ProbZ
10	cdm	0.1429	-0.0130	0.2988	1.15360	0.98711	1.34817	0.0724
11	lt12m	0.2691	0.0307	0.5075	1.30876	1.03113	1.66113	0.0270
12	to30m	0.2658	0.0813	0.4503	1.30447	1.08473	1.56872	0.0047
13	to60m	0.2282	0.0622	0.3943	1.25639	1.06420	1.48329	0.0070
14	depressm	0.2674	0.1465	0.3884	1.30657	1.15772	1.47456	<.0001
15	numpart25m	0.0220	-0.3276	0.3716	1.02226	0.72063	1.45012	0.9018
16	numpart69m	0.0759	-0.2831	0.4349	1.07887	0.75344	1.54485	0.6786
17	numpart10m	0.2801	-0.0668	0.6270	1.32331	0.93543	1.87204	0.1135
18	alclight6m	-0.0429	-0.2544	0.1686	0.95801	0.77536	1.18369	0.6910
19	alcmoderate6m	-0.0050	-0.2280	0.2180	0.99504	0.79615	1.24361	0.9651
20	heavy6m	0.3195	0.0587	0.5803	1.37650	1.06051	1.78664	0.0163
21	marj6lt1wkm	-0.0258	-0.1705	0.1189	0.97453	0.84327	1.12621	0.7266
22	marj61wkmorem	-0.1525	-0.3409	0.0360	0.85859	0.71114	1.03663	0.1128
23	pop6lt1wkm	0.2798	0.1414	0.4181	1.32283	1.15193	1.51909	<.0001
24	pop61wkmorem	0.1719	-0.0464	0.3902	1.18753	0.95462	1.47727	0.1228
25	pcp6lt1wkm	-0.0499	-0.2190	0.1191	0.95128	0.80332	1.12649	0.5625
26	pcp61wkmorem	-0.0563	-0.4329	0.3203	0.94528	0.64863	1.37759	0.7696
27	coc6lt1wkm	0.1773	0.0062	0.3484	1.19398	1.00620	1.41681	0.0423
28	coc61wkmorem	0.5417	0.1744	0.9089	1.71887	1.19057	2.48159	0.0038
29	spd6lt1wkm	0.3386	0.1399	0.5373	1.40304	1.15021	1.71145	0.0008
30	spd61wkmorem	0.7031	0.2909	1.1153	2.01999	1.33767	3.05035	0.0008
31	crk6lt1wkm	-0.0585	-0.3814	0.2643	0.94313	0.68289	1.30254	0.7223
32	crk61wkmorem	0.3706	-0.2262	0.9674	1.44859	0.79759	2.63097	0.2235
33	her6lt1wkm	-0.3783	-1.1087	0.3521	0.68503	0.32998	1.42211	0.3101
34	her61wkmorem	-1.2325	-3.2532	0.7882	0.29156	0.03865	2.19938	0.2319
35	injm	0.4343	-0.2668	1.1355	1.54393	0.76582	3.11265	0.2247

COMPARISON OF METHODS

Since we were confronted with computational difficulties using non-linear mixed-effects modeling, a fair comparison of the data analysis results is not possible. However, conceptually both logistic mixed-effects and conditional logistic regression methods attempt to address the within-individual effects. That is, they address the effects of substance use on risky sex conditional on characteristics of an individual. The first achieves this by imposing a specific distribution on the random effect. The second eliminates the influence of between-individual predictors and accounts for within-individual correlation by limiting the analysis to participants who reported at least one, but not all episodes of risky sex. Furthermore, the logistic mixed-effects model implies the model for conditional logistic regression; thus making the latter approach more robust. Likewise, if the additional distribution assumptions of the random effects do hold, the logistic mixed-effects model is expected to provide more efficient estimates.

Most importantly, our experience in this analysis showed that the estimation procedure for logistic mixed-effects models could become numerically unstable, especially with a large number of covariates. Conditional logistic regression has a clear advantage in this regard.

Between-individual effects are marginal effects, which are different from within-individual effects. Generalized estimating equations are an effective and efficient approach to characterize this relationship. We found that the proc GENMOD procedure in SAS is easy to use and numerically reliable.

CONCLUSION

Examining risky sexual behavior and substance use in the EXPLORE data set presented some challenging data analysis issues. Given the complexity of the data structure, a thorough picture of the within- and between-individual effects was best seen by the use of conditional logistic regression and generalized estimating equations. The SAS procedures (proc PHREG and proc GENMOD) proved to be the most powerful tools in this endeavor.

REFERENCES

1. The EXPLORE Study Team. Effects of a behavioural intervention to reduce acquisition of HIV infection among men who have sex with men: the EXPLORE randomised controlled study. Lancet 2004;364:41-50.
2. Colfax G, Vittinghof E, Husnik MJ, et al. Substance use and sexual risk: a participant- and episode-level analysis among a cohort of men who have sex with men. Am J Epidemiol 2004;159:1002-1012.
3. SAS Institute Inc. SAS/STAT User's Guide, Version 8, Volume 2, Cary, NC: SAS Institute Inc., 1999. 3884 pp.
4. Wolfinger RD, Fitting nonlinear mixed models with the new NLMIXED procedure. SAS Institute Inc. paper 287.
5. SAS Institute Inc. SAS/STAT User's Guide, Version 8, Volume 3, Cary, NC: SAS Institute Inc., 1999. 3884 pp.
6. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13-22.
7. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121-30.

CONTACT INFORMATION

Marla Husnik, MS
Statistical Research Associate
Fred Hutchinson Cancer Research Center
Statistical Center for HIV/AIDS Research & Prevention
1100 Fairview Avenue N, LE-400
P.O. Box 19024
Seattle, WA 98109-1024
E-mail: marla@scharp.org

Yijian (Eugene) Huang, PhD
Associate Professor of Biostatistics
Department of Biostatistics
Rollins School of Public Health
Emory University
1518 Clifton Road NE
Atlanta, GA 30322
E-mail: yhuang5@emory.edu

Grant Colfax, MD
Director of HIV Prevention Studies
HIV Research Branch
San Francisco Department of Public Health
25 Van Ness Avenue, Suite 710
San Francisco, CA 94102
E-mail: grant.colfax@sfdph.org