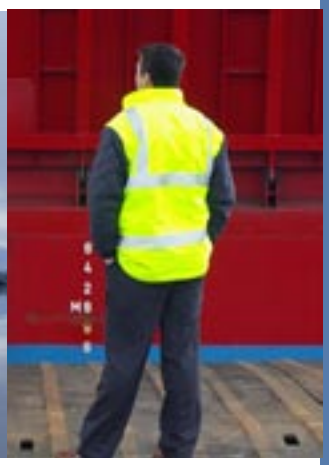
How to use SAS® Proc Traj and SAS® Proc Glimmix in Respiratory Epidemiology

Victoria Arrandale[1], Mieke Koehoorn[2], Ying MacNab[2], Susan M. Kennedy[1]

[1]School of Occupational & Environmental Hygiene
[2]Department of Health Care & Epidemiology
University of British Columbia, Vancouver, Canada

December, 2006

CHER

SOEH School of Occupational & Environmental Hygiene

UBC

Table of Contents

How to use SAS® Proc Traj and SAS® Proc Glimmix in Respiratory Epidemiology

Introduction

This document outlines the use of two procedures capable of modeling repeated respiratory symptom data in the software package SAS®: Proc Traj and Proc Glimmix. SAS® Proc Traj is a discrete mixture model which models the patterns of change over time in multiple subgroups within the population. SAS® Proc Glimmix is a procedure that fits a generalized linear model to non-linear outcome data either with or without random effects.

Goal

The goal of this document is to provide a concise user's guide for applying discrete mixture models (Proc Traj) and generalized linear mixed models (Proc Glimmix) in the analysis of longitudinal respiratory symptom data using SAS® software. This document does not attempt to describe the statistical theory behind either of these techniques.

How to Use This Document

This document presents an outline for setting up models in both Proc Traj and Proc Glimmix for analyzing repeated respiratory symptom outcomes. Data organization is explained, the modeling procedure is outlined, the basic syntax (appropriate for binary respiratory symptom outcomes, although Proc Traj will handle other outcomes) is described and the relevant modeling possibilities are discussed.

This document should be a starting point for modeling using Proc Traj and Proc Glimmix. Readers are advised to refer to the SAS® documentation as well as the noted reference texts for further explanations and for confirmation that the models are appropriate for the data in use.

Please Note: This document was produced as part of an MSc thesis.[1] It is correct to the best of our knowledge; however, the authors take no responsibility for for the performance of the procedures or for the correctness of any research results. This document is liscenced under the Creative Commons Attribution-Noncommercial-Share Alike 2.5 Canada License. Please feel free to share this document under the following conditions:

- You must provide attribution to the authors.
- You may not use this work for commercial purposes.
- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the authors.

To request a paper copy of this document, please send a stamped, self-addressed envelope to: The Centre for Health and Environment Research, 2206 East Mall, Vancouver, BC, V6T 1Z3.
Correspondence of a scientific nature should be sent to: Dr Susan M. Kennedy, 2206 East Mall, Vancouver, BC V6T 1Z3; e-mail kennedy@interchange.ubc.ca

[1]Arrandale VH. An Evaluation of Two Existing Methods for Analyzing Longitudinal Respiratory Symptom Data [M.Sc. Thesis]. Vancouver: University of British Columbia; 2006.

SAS® Trajectory Procedure

Overview

The SAS® Trajectory Procedure (Proc Traj) is a user-friendly finite mixture model procedure designed to run easily on the SAS® platform. Proc Traj is a specialized mixture model that estimates multiple groups within the population, in contrast to a traditional regression or growth curve model that models only one mean within the population. Designed by researchers, Proc Traj is not part of the base SAS® program and must be downloaded separately.

Proc Traj is designed to address research questions focused on describing the trajectory, or pattern, of change over time in the dependent variable, specifically questions concerned with multiple distinct patterns of change over time and modeling unobserved heterogeneity in the data. Proc Traj estimates a regression model for each discrete group within the population.

The focus of the Proc Traj procedure is on group membership and identifying distinct subgroups within the population. Proc Traj does not provide any individual level information on the pattern of change over time; subjects are grouped and it is assumed that every subject in the group follows the same trajectory. There is no random effect capability within the Proc Traj model.

The documentation for SAS® Proc Traj is a peer-reviewed publication by Jones, Nagin and Roeder (1). A recent text authored by D. Nagin (2) is a valuable reference for users of Proc Traj and should be reviewed by those interested in the statistical theory behind Proc Traj.

Requirements

To apply Proc Traj to your data, you need (at a minimum) multiple measures of the outcome of interest and information on the timing of the repeated measures. It would also be helpful to have multiple measures on a number of covariates you are also interested in.

You must also download the Proc Traj application from B. Jones' website[2] and have copied the files to the folders as directed on the website.

[2] http://www.andrew.cmu.edu/user/bjones

Data Organization

In order to use Proc Traj you must organize your data in a multivariate, or "wide" format, where there is only one row of data for each subject and multiple observations included in one line of data. An example of data ready for use with Proc Traj is shown in Table 1, a description of each variable is provided in Table 2. You can see in Table 1 that the outcome variable "wheeze" is denoted by the variables Wez01, Wez02 and Wez03. These three variables correspond to three repeated measurements taken at three different times. The time at which each of these measurements was collected is represented by the variables Yr01, Yr02 and Yr03. If a subject did not complete a visit, all variables corresponding to that visit are blank, in this case a "." is used to indicate missing data.

Table 1 Mock data set up for analysis with Proc Traj

ID	Sex	Byr	Csmk01	Csmk02	Csmk03	Wez01	Wez02	Wez03	Yr01	Yr02	Yr03
001	0	1947	0	0	0	0	0	0	1992	1994	1999
002	1	1953	1	.	0	0	.	1	1992	1994	1999
003	0	1951	0	1	1	1	0	1	1992	1994	1999
004	0	1946	0	0	.	1	1	.	1992	1994	1999
005	1	1950	1	0	1	1	1	1	1992	1994	1999

Table 2 Description of variables in mock data (Table 1)

Variable Name	Description	Values
ID	Subject ID	as assigned
Sex	Sex of subject	0= male 1= female
Byr	Year of birth	continuous, in years
Csmk01	Current smoker at visit 1	0= never or former smoker 1= current smoker
Csmk02	Current smoker at visit 2	
Csmk03	Current smoker at visit 3	
Wez01	Response to wheeze question at visit 1	0= no wheeze 1= wheeze
Wez02	Response to wheeze question at visit 2	
Wez03	Response to wheeze question at visit 3	
Yr01	Date of visit 1	values corresponding to data
Yr02	Date of visit 2	
Yr03	Date of visit 3	

The variables that describe repeated measures of the same outcome must be numbered consecutively (i.e. Csmk01, Csmk02, Csmk03 etc.) before Proc Traj will accept them; this will usually require some recoding. SAS® will not accept the data if the variables are labeled alternatively (i.e. smk1992, smk1994, smk1999) even if this is logical given your data set. By identifying the variables that contain information on the date of each repeat measure (i.e. Yr01, Yr02, Yr03) you are specifying the space between repeated measures. Time varying covariates (i.e. Csmk01, Csmk02, Csmk03 for smoking information at each visit) must also be named with consecutive numbers corresponding to the visit.

Dummy Variables

It is advisable to create dummy variables for each of your covariates that you plan to input into a Proc Traj model, as was done for Current Smoking in Table 1. Covariates can be input in a binary (dummy) form or a continuous form but Proc Traj does not handle categorical covariates.

Missing Data

Proc Traj is able to handle data that is missing completely at random (MCAR), but is unable to handle data that is missing for more complex reasons (2). Missing data can be entered in the dataset as shown in Table 1.

Types of Research Questions

In terms of respiratory symptom data, Proc Traj should be used when your research question is similar to one of the following:

- Are there multiple patterns of change in the outcome?
- How many patterns of change are there in the outcome?
- What is the shape of the change over time?
- What predicts membership in each of these groups?
- What are the characteristics that differ (or are similar) between the different groups?

Syntax

The entire Proc Traj syntax is outlined on B. Jones' website[1] and should be referenced for any further questions regarding syntax.

A simple Proc Traj syntax for a two group model of the respiratory symptom wheeze is presented here:

```
proc traj data=a.mockdata out=out outstat=os
outplot=op;
var wez01-wez03;
indep year01-year03;
model logit;
ngroups 2;
order 0 1;
id ID;
run;
%trajplot (OP, OS, "Title of graph",
"Subtitle", "Y-axis label", "X-axis label");
```

As in all SAS® procedures, the Proc Traj statement outlines the data set to be used and in this case also defines the output from the procedure.

The ‘var’ statement defines the binary symptom outcome of interest. ‘Indep’ defines the time variables that you are modeling the outcome over. The ‘model’ statement identifies the outcome as binary and the ‘ngroups’ states how many groups you want to model. ‘Order’ assigns the order of each equation that will describe the change over time in each group. ‘ID’ identifies the subjects in your population and also denotes which variable you want to use to uniquely assign subjects to a specific group in the output data set (in this case, “out”).

The ‘%trajplot’ is a macro statement that results in the graphical output from Proc Traj. This macro includes references to the outplot and outstat statements in the ‘Proc Traj’ statement. If you make any changes in the ‘Proc Traj’ statement be sure to adjust the trajplot macro accordingly.

When including time independent covariates into a Proc Traj model, the ‘risk’ or ‘tcov’ statements can be added to the syntax for time stable covariates and time varying covariates respectively. For example, a time stable covariate for sex could be added:

```
proc traj data=a.mockdata out=out outstat=os
outplot=op;
var wez01-wez03;
indep year01-year03;
model logit;
ngroups 2;
order 0 1;
risk female;
id ID;
run;
%trajplot (OP, OS, “Title of graph”, “Subtitle”, “Y-
axis label”, “X-axis label”);
```

Or, a time varying covariate for current smoking could be added:

```
proc traj data=a.mockdata out=out outstat=os
outplot=op;
var wez01-wez03;
indep year01-year03;
model logit;
ngroups 2;
order 0 1;
tcov csmk01-csmk03;
id ID;
run;
%trajplot (OP, OS, “Title of graph”, “Subtitle”, “Y-
axis label”, “X-axis label”);
```

Selecting the Best Model

The model fitting procedure with Proc Traj is iterative and requires a priori decisions based on substantive knowledge. In the most basic process, the following **steps** should be followed:

1. Decide on the maximum number of groups using a priori knowledge
2. Fit number of groups to data (start by fitting a one group model, and then fit up to the maximum logical number of groups in a step wise manner)
3. Select the shape of the pattern of change for each group over time (e.g. linear)
4. Perform further modeling if required (addition of covariates, inclusion of second outcome etc.)

Steps 1 & 2: To decide on the optimum number of groups for your data you must begin by fitting a basic one group model with all groups set to a second order (quadratic) equation. Then fit a two group, then three group model etc. until you have fit the maximum number of groups based on your a priori decision. Nagin suggests setting all group orders to second order during this process (2).

For each model you fit in this first step you will be given two Bayesian Information Criterion (BIC) values in the output: one relates to the overall sample size (total number of observations), and the other relates to the subject sample size (number of subjects). The true BIC for the model lies between these values (2). The BIC is the log-likelihood adjusted for the number of parameters and the sample size (1). In the Proc Traj procedure the BIC values given in the output are negative; the best fit model is the one with the smallest negative number.

Model selection in Proc Traj uses the BIC to select the best fitting model via two different methods. The first, described by Jones, Roeder and Nagin (1) uses the change in the BIC between two models to measure the weight of evidence against the null model. For each increasingly complex model that is tested, the BIC of the more complex (larger number of groups, or higher order equation) less the BIC of the less complex model is used to select the model that better fits the data.

$$\Delta BIC = BIC_{(complex)} - BIC_{(null)}$$

The difference in BIC between the two models is a measure of the evidence against the null model. Jones, Nagin and Roeder (1) suggest criteria for strength of evidence against the null model (Table 3). Using the difference in the logged Bayes factor between successive models, the difference between the alternate and the null model can be qualified. The null model is always the simpler model (i.e. less groups, or lower order equations). The interpretation of the logged Bayes factor ($2\Delta BIC$) in terms of model preference is shown in Table 3.

Table 3 Interpretation of logged Bayes factor (2*ΔBIC) for model selection (Adapted from Table 2 in (1))

2*ΔBIC	Evidence against Ho
0 to 2	Not worth mentioning
2 to 6	Positive
6 to 10	Strong
> 10	Very Strong

The second method is called Jeffreys's scale of the evidence and is described by Nagin (2). Jeffreys's scale of the evidence uses the exponentiated difference between the BIC values of models, i and j:

$$\text{Bayes Factor} \approx e^{BICi-BICj}$$

In this case is does not matter which model is the null model; only that the researcher remembers which model is which. The interpretation of Jeffreys's scale of the evidence is outlined in Table 4. Further description and explanation can be found in Chapter 4 of Nagin (2005) (2).

Table 4 Interpretation of Bayes Factor (eBICi-BICj) for model selection (Adapted from Table 4.2 in (2))

Bayes Factor (Bij)	Interpretation
Bij < 1/10	Strong evidence for model j
1/10 < Bij < 1/3	Moderate evidence for model j
1/3 < Bij < 1	Weak evidence for model j
1< Bij < 3	Weak evidence for model i
3< Bij < 10	Moderate evidence for model i
Bij > 10	Strong evidence for model i

When selecting the 'best' model it is important to base decisions on substantive knowledge about the research area, and remember the rule of parsimony to select the simplest model that best describes the data.

Again, in reference to the example with respiratory symptoms, if we tested five models (one group up to five groups) we would have five BIC values to review. The comparisons are completed in a step-wise manner so that the two-group model is compared to the one-group model, and the three-group model to the two-group model and so on. In each case, the model with the smaller number of groups is the null model.

Step 3: The next step in fitting a model using Proc Traj is selecting the shape of each group's trajectory over time. Proc Traj can model up to a fourth order polynomial and can model both linear and non-linear trajectories within the same model. This can be done using substantive knowledge (i.e. we expect one group to never report symptoms so this group's trajectory will be a zero-order equation, or a straight line) or it can be done using the ΔBIC. It seems ideal to use a combination of substantive knowledge and statistical inference to make the decision regarding the shape of each group's trajectory.

Output

The output from Proc Traj includes the parameter estimates for each group (with standard errors), group membership probabilities (population level) and model fit statistics. The output data set (out= in 'Proc Traj' statement) includes all the variables included in the analysis (not all the variables in the original dataset), the variable identified in the 'id' statement, posterior subject specific group membership probabilities and a group assignment for each individual.

The parameter estimates can be used to construct regression equations for each group and a system of equations to describe the population. The relative differences between the estimates for the same covariate between groups can be used to make inferences about differences between the groups.

The posterior group membership probabilities and the group assignment variables in the output data set can be used to explore between group differences in covariates not included in the model and potentially as predictor variables in separate analyses. The posterior group probabilities are calculated for each individual based on the estimated parameters, and the individual is assigned to a group based on their highest posterior group probability (2).

The output from a basic model (no covariates) is shown in Figure 1. The intercept parameters represent the estimated intercept for each group. For Group 2 the linear parameter represents the estimated coefficients for the linear time component of the regression equation. The group membership probabilities indicate what proportion of the population is estimated to belong to each group. And, the BIC values are the final portion of the output. Note the BIC values are shown for two sample sizes; first for all the data points and second for the number of subjects.

Figure 1 Text output from basic Proc Traj model with no covariates

Output - (Untitled)

The SAS System 10:25 Wednesday

Maximum Likelihood Estimates
Model: Logistic (LOGIT)

Group	Parameter	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
1	Intercept	-2.89924	0.24364	-11.900	0.0000
2	Intercept	-11.80231	2.63866	-4.473	0.0000
	Linear	0.13221	0.02944	4.491	0.0000
	Group membership				
1	(%)	73.14700	3.71808	19.673	0.0000
2	(%)	26.85300	3.71808	7.222	0.0000

BIC= -1113.12092 (N=2489) BIC= -1111.14124 (N=925) AIC= -1101.48165

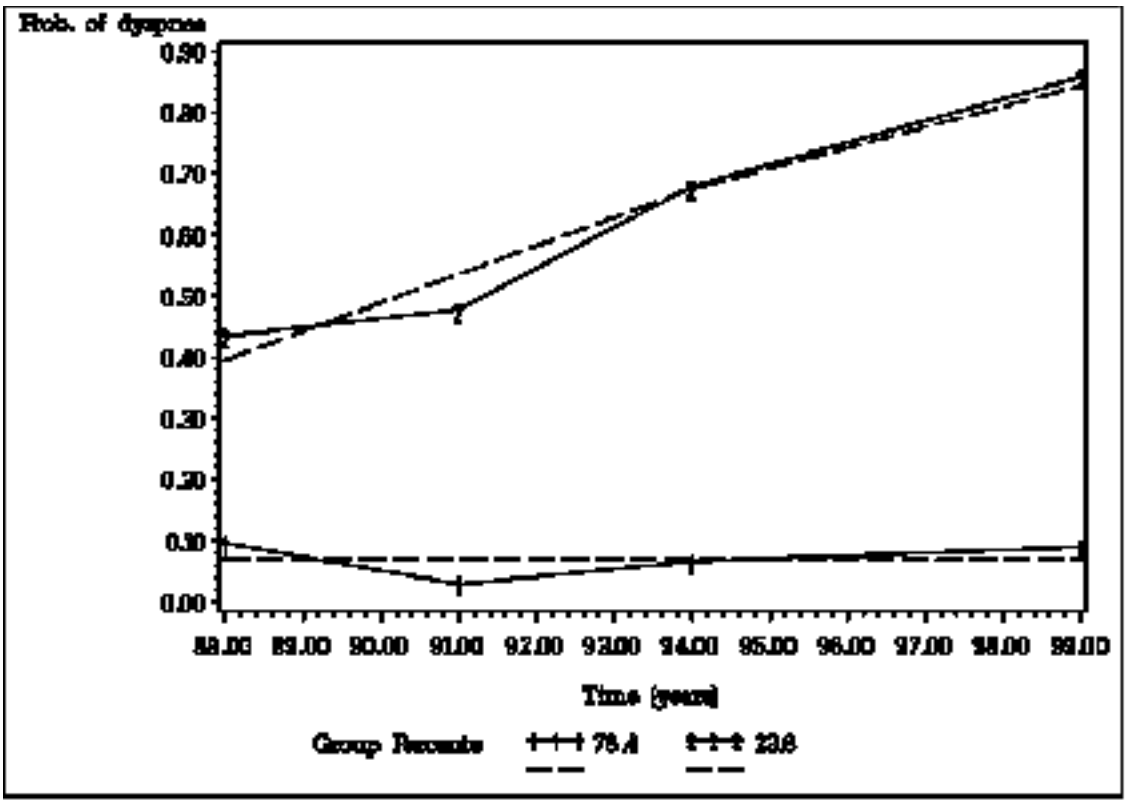


Figure 2 Graphical output from Proc Traj showing a two-group model with group membership probability and shape of change over time

User Information

Base SAS® is required to run Proc Traj. You can download the procedure from the B. Jones' Proc Traj Home Page and copy the downloaded files into the appropriate folders on your hard drive (instructions provided on web page). Proc Traj will then be installed and functional in the SAS® platform.

Because Proc Traj is an add-on to SAS®, there is no formal SAS® documentation in the traditional SAS® format. Users are advised to thoroughly review the reference texts listed below.

Cautions

Researchers using Proc Traj are advised to remember that the multiple groups estimated are not reified groups. The identified groups are estimations of multiple patterns of change within the population, and we must be careful not to think of group membership and trajectory shape as absolute certainties.

Reference Texts

Nagin, Daniel S. Group-based Modeling of Development. Harvard University Press: Massachusetts (2005).

Jones, B., Nagin, D., & Roeder, K. A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. Sociological Methods & Research (2001) 29: 374-393.

SAS® Proc Traj Home. http://www.andrew.cmu.edu/user/bjones.

SAS® Glimmix Procedure

Overview

Within the SAS® program there are several procedure for constructing mixed models. The most common procedure is Proc Mixed, which models continuous outcomes. A newer procedure, Proc Glimmix (general linear mixed models), models non-linear data. Proc Nlmixed (non-linear mixed models) also models non-linear data but is primarily for use with advanced modeling and is programmatically complex.

This guide is focused on modeling binary respiratory symptom data using Proc Glimmix. Proc Glimmix is a fast, flexible procedure capable of running linear models (fixed effects), generalized linear models (fixed effects), linear mixed models (fixed and random effects) as well as generalized linear mixed models (fixed and random effects). The focus of Proc Glimmix for this guide is the generalized linear mixed model capability. Proc Glimmix does not ship with the SAS® v.9, but the add-on and the documentation are both available for download on the SAS® support website.[2]

Mixed models are particularly useful in the modeling of longitudinal data because repeated measurements are collected over time on subjects and are inevitably correlated. A fixed effects model requires all of the measurements to be independent; in a longitudinal repeated measures data set this assumption is violated. In a longitudinal mixed effects regression model the autocorrelation between repeated measures on individual subjects is accounted for, one overall group mean is modeled and subject specific deviations from the group mean are estimated.

When to Use Mixed Effects

In reference to longitudinal study designs, random effects should be introduced into a regression model when there are correlated outcome measures (i.e. repeated measures on individuals) and when you want to allow the effect of a particular covariate on the outcome to vary randomly among your subjects.

Requirements

For a mixed model you should have outcome measures that you expect are correlated; this occurs when you have collected repeated measures on individuals. To use Proc Glimmix, you should also have a non-linear outcome variable, in this case binary symptom data. If you are using a continuous measure of a respiratory symptom, you should consult SAS® Proc Mixed.

[2] http://support.sas.com

Data Organization

For Proc Glimmix models, data must be organized in a univariate, or “long”, format where there is one observation per line of data and multiple lines of data per subject. An example of data organized this way is shown in Table 5.

Table 5 Mock data set up for analysis with Proc Glimmix

ID	Age	Vyr	First Visit	Second Visit	Sex	Fsmk	Csmk	Wez
001	53	1992	0	0	0	0	0	0
001	55	1994	1	0	0	0	0	0
001	60	1999	0	1	0	0	0	0
002	49	1992	0	0	1	0	1	0
002	56	1999	1	0	1	1	0	1
003	40	1992	0	0	0	1	0	1
003	42	1994	1	0	0	0	1	0
003	47	1999	0	1	0	0	1	1
004	60	1992	0	0	0	0	0	1
004	62	1994	1	0	0	0	0	1
005	36	1992	0	0	1	0	1	1
005	38	1994	1	0	1	1	0	1
005	43	1999	0	1	1	0	1	1

Table 6 Description of variables in mock data (Table 5)

Variable Name	Description	Values
ID	Subject ID	as assigned
Age	Subject’s Age	age in years
Vyr	Year of Visit	date in years
Vis2	Two complete visits	yes/no
Vis3	Three complete visits	yes/no
Sex	Sex of subject	0= male 1= female
Fsmk	Former Smoker	yes/no
Csmk	Current Smoker	yes/no
Wez	Response to wheeze question	0= no wheeze 1= wheeze

The data setup is quite straightforward, but note that the data includes dummy variables for otherwise categorical variables (smoking, visit number). It is easier to deal with dummy variables, rather than categorical variables, in Proc Glimmix.

Dummy Variables

Proc Glimmix does have a ‘class’ statement in the syntax, and therefore theoretically you can input categorical variables without any recoding. However, it is not easy to adjust the reference groups using the class statement. Instead, researchers are advised to create dummy variables for each categorical variable.

Missing Data

Proc Glimmix does handle missing data. Observations are not excluded if variable values are missing within the observation. However, if the amount of missing data is substantial the specified models may not converge. In this case, you can limit your dataset to subjects with less missing data in an attempt to run the models successfully, but this will result in a smaller sample size and a loss of power.

Types of Research Questions

In terms of longitudinal respiratory symptom data, Proc Glimmix should be used when your research question is similar to one of the following:

- Considering the repeated measures on individuals, what are the risk factors that predict the outcome?
- How much variation exists between individuals for a given main effect?
- How are the repeated measures on individuals correlated?
- Does the probability of the outcome change over time?

Syntax

The syntax for a basic Proc Glimmix model is outlined here. For further discussion of the Proc Glimmix syntax, including the specification of a marginal model for estimating correlation structures, readers should refer to the official SAS® documentation (3). The first model presented is a mixed model estimating the risk factors for wheeze:

```
proc glimmix data=a.mockdata ;
model wez (event='1')= age sex vis2 vis3 fsmk csmk /
s dist=binary link=logit or ;
random intercept / subject=case ;
ods output oddsratios=a.oddratio ;
run;
```

Again, the procedure statement specifies the dataset to be used. The model statement indicates that the outcome is 'wez' and that Proc Glimmix is modeling the probability of 'wez=1'. Beyond that, the model statement lists the covariates to include in the model (in this case they are all dummy variables except age) and the model options. The included model options in this example are 's' (can also be written as 'solution') to provide the fixed effects parameter estimates, 'dist' to specify the distribution of the outcome, 'link' to specify the link function and 'or' to provide the odds ratios for the fixed effects. An explanation of the 'dist' and 'link' options as well as a table of possible values is provided in the Proc Glimmix documentation (3). When the outcome is a binary respiratory symptom, the 'dist' option will be binary and the 'link' will always be logit.

The random statement specifies the random variables. In this case only a random intercept was specified, but any other random variables would be listed before the forward slash. The random statement options used here are 'subject', which identifies the variable for which there are repeated measurements. The entry here will always be the subject or case identification variable when the repeated measures are on individuals.

The odds ratio option in the model statement gives a very long output table that is difficult to interpret from the SAS® window. For this reason it is advisable to use the 'ods output' statement to specify that the odds ratio table be output as a dataset. Once the odds ratio table is seen as a dataset file it is much easier to interpret. For more information on ODS output and how to limit the output to specific portions (using the 'ods select' statement) or output specific tables to a new dataset, refer to the SAS/STAT® documentation (4).

Selecting the Best Model

Unfortunately there is no easy way to select the best fitting model using Proc Glimmix. Proc Glimmix does not provide a likelihood value for the estimated models, instead pseudo-likelihood is calculated and this value cannot be used in a likelihood ratio test. Instead, users are advised to construct their model in a stepwise manner using substantive knowledge. A priori hypotheses should drive decision making while constructing the model. Once the model is assembled, the significance of individual estimates and prior knowledge should guide what remains in the model.

Additional fit statistics can be requested in the Proc Glimmix statement by including the following the command:

```
IC = PQ
```

When this command is included, pseudo-AIC and pseudo-BIC values will be included in the output fit statistic table. In the case of both pseudo-AIC and pseudo-BIC values, a smaller value indicates a better model fit.

More information on the complexities of fitting models in Proc Glimmix can be found in the documentation (3). There is also an on-going discussion of this and other pertinent SAS® issues on the SAS® user's list serve (5).

Output

Proc Glimmix provides extensive text output in SAS®. A sample of Proc Glimmix output, limited to the key pieces of output, is shown in Figure 2. The fit statistics are shown, including the pseudo-likelihood mentioned previously. Covariance parameter estimates are the estimates of variance in each of the specified random effects, in this case only a random intercept was included in the random statement. The covariance parameter estimates provide a measure of the between subject variability in the random variable. The next table shown is the estimates of the fixed effects included in the model statement. These are the regression coefficients describing the effect of each independent variable on the probability of reporting the symptom outcome.

If odds ratios had been requested in the output they would follow after the fixed effects parameter estimates.

The complete Proc Glimmix output is extensive, including information on the model optimization, iterative process of model fitting and the convergence criteria. Specific portions of the default output can be selected for viewing in the output window using the 'ods select' statement (4) as was done in Figure 2.

User Information

Proc Glimmix does not ship with SAS®, instead the procedure and documentation can be downloaded from the SAS® Support website. The files are self-extracting and will copy all necessary files to the correct location (unlike Proc Traj, where you have to manually move the downloaded files into the correct folders).

The Glimmix procedure is supported by SAS® and has traditional SAS® documentation (3). In addition, the book SAS® for Mixed Models contains an intensive chapter (with examples) on generalized linear mixed models that should be reviewed.

Figure 3 Sample output from mixed model using Proc Glimmix.

led)

The SAS System 10:25 Wednesday, October 18, 2006

The GLIMMIX Procedure

Number of Observations Read	2472
Number of Observations Used	2472

Fit Statistics

-2 Res Log Pseudo-Likelihood	11820.21
Generalized Chi-Square	1483.84
Gener. Chi-Square / DF	0.60

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	CASE	1.3567	0.1763

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-4.3395	0.4093	921	-10.60	<.0001
AGE2	0.05232	0.006807	1539	7.69	<.0001
FEMALE	0.4411	0.2342	1539	1.88	0.0599
threevis	-0.3112	0.1622	1539	-1.92	0.0552
fourvis	0.1189	0.1903	1539	0.62	0.5321
FSMOKE	0.1215	0.1709	1539	0.71	0.4773
CSMOKE	0.5036	0.1877	1539	2.68	0.0074
wkartern	0.7909	0.2056	1539	3.85	0.0001
wkardeas	-0.2562	0.2365	1539	-1.08	0.2789
asmachck	1.1135	0.2320	1539	4.80	<.0001
wkexsonc	-0.1777	0.1890	1539	-0.94	0.3473
wkexoft	-0.3564	0.2051	1539	-1.74	0.0825

Cautions

Researchers using Proc Glimmix should be aware that there are acknowledged issues with the estimation technique used in Proc Glimmix and that the procedure may result in biased coefficient estimates. A statistician can assess the magnitude of this problem using simulation techniques. Researchers should consult a statistician to ensure that their results are not biased.

Reference Texts

Arrandale VH. An Evaluation of Two Existing Methods for Analyzing Longitudinal Respiratory Symptom Data [M.Sc. Thesis]. Vancouver: University of British Columbia; 2006.

Diggle PJ HP, Liang K, Zeger SL. Analysis of Longitudinal Data. 2nd ed. New York: Oxford University Press; 2003.

Fitzmaurice GM LN, Ware JH. Applied Longitudinal Analysis. New Jersey: John Wiley & Sons; 2004.

Littell R, Milliken G, Stroup W, Wolfinger R, Schabenberger O. SAS for Mixed Models. 2nd ed. Cary NC: SAS Institute Inc.; 2006.

SAS Institute. SAS/STAT 9.1 User's Guide: SAS Publishing; 2004.

SAS Institute. The Glimmix Procedure: SAS Publishing; 2006.

Reference List

(1) Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. Sociological Methods & Research 2001;29(3):374-393.

(2) Nagin DS. Group-based Modeling of Development. Cambridge, Massachusetts: Harvard University Press; 2005.

(3) SAS Institute. The GLIMMIX Procedure: SAS Publishing; 2006.

(4) SAS Institute. SAS/STAT 9.1 User's Guide: SAS Publishing; 2004.

(5) Archives of SAS-L@LISTSERV.UGA.EDU. Retrieved on multiple occasions, from http://listserv.uga.edu/archives/sas-l.html.