Adaptive Modeling of Longitudinal HIV Viral Load Data

George J. Knaf1, Jean P. O’Malley1, Kristopher P. Fennie2, Carol A. Bova3, Kevin D. Dieckhaus4, Gerald H. Friedland2, Ann B. Williams2

1 Oregon Health & Science University 2 Yale University 3 University of Massachusetts Worcester 4 University of Connecticut Health Center

Background
We have previously developed methods for adaptive linear, logistic, and Poisson regression modeling of univariate outcomes. We have previously used adaptive Poisson regression to model adherence adherence over time for HIV patients using data collected electronically with Medication Event Monitoring System (MEMS) caps, including data for each subject separately as well as for all subjects combined. We have recently extended these methods to handle repeated measures data. These adaptive modeling methods use k-fold likelihood cross-validation along with heuristic search through fractional polynomial models to identify nonparametric regression models for expected outcomes while accounting when appropriate for correlation within repeated outcome measurements. We describe adaptive repeated measures modeling and demonstrate it with analyses of HIV viral load over time in terms of antiretroviral adherence based on MEMS caps.

Overview of Topics
- the adaptive modeling process
- the data to be analyzed
  - including electronic adherence data and viral load data
- modeling the dependence of the vector y of log (base 10) viral loads at s times t on prior adherence $X$

Adaptive Repeated Measures Modeling

Modeling the Expectation
- used nonparametric fractional polynomial model
- By modeled with polynomial in power transforms $x^p$ of possibly time-varying predictors $X$ with coefficient vector $p$
- represented by predictor matrix $X$ combining $p$ values for a subject over all $s$
- for a given $X$, maximum likelihood used to estimate $p$
- $X$ determined by adaptively selecting number of terms and powers $p$ for transforms $x^p$
- subject indexes for $X$ left off to simplify notation

Modeling the Covariance Matrix $\Sigma$
- assuming multivariate normal distribution
- first used standard repeated measures approach
  - invariance assumption
    - variances the same for all $s$ and covariances the same for all pairs $(t, t')$
    - recently extended to autoregression
      - variances still constant, but correlations weakening the further apart $t$ and $t'$ are
- need way to evaluate and select models

Model Evaluation
LCV = $\ln L_p + I_f(\pi) L_p(f)$
- $k$-fold likelihood cross-validation (LCV)
- $k$ is number of folds
- subjects randomly assigned to folds $F$
- $I_f(x)$ = integral assignment probabilities
- $\pi$ = is number of outcome measurements for all subjects
- compute multivariate normal likelihood $L$ for folds $F$
- $L$ = $\sum_{i=1}^{k} L_i(x)$
- $L_i(x)$ = likelihood for fold $F_i$
- parameter values used in likelihood terms for fold $F_i$
- $L_i(x)$ = for all subjects $i$ of outcome vectors $y_i$ and predictor matrices $X_i$
- geometric average deleted likelihood scores
- larger scores indicate better models for $X$
- $k$ set to the first local maximum for a benchmark analysis

Computational Support
- models computed using specialized SAS macro written primarily in matrix language of PROC IML
- macro supports nonparametric regression modeling
- using heuristic search controlled by LCV scores
- including linear, logistic, and Poisson regression for univariate outcomes
- can be extended to recently adapted repeated measures
- with compound symmetric or autoregressive covariance

Log Viral Load vs. Prior Adherence

Example Data
- from study of 172 HIV+ subjects on antiretroviral (ARV) meds
- adherence through Home Education and Nursing Assessment (ATHENA) Project, PI A. Williams, Yale University
- 50.6% (87) randomized to adherence intervention
- control group received standard care

Electronic Adherence Data
- subjects given pill bottles equipped with MEMS caps
- caps recorded openings and presumably medication taking
- all medications in pill bottles prescribed at 2 doses per day
- usage MEMS data for 181 subjects (53.6%)
- based on 75,000 cap openings for 66,500 days of use of 186 cases over about 4.5 years from 6/1999 to 3/2002
- used standard summary adherence measure
- % prescribed administrations taken (PAT)

Self-Reported Adherence Data
- self-reported adherence for 3 days prior to interview
- from interview data collected up to 7 times 3 months apart

Viral Load Data
- viral loads obtained from medical records and matched in time to interview dates
- 843 measurements up to 7 times for 160 subjects
- for whom MEMS adherence data also available
- range 1.30-5.88 log copies/ml (20-750,000 copies/ml)

Controlling for Initial Viral Load
- modeled log viral loads at 6 later time points in terms of initial log viral load as well as $\epsilon$
- used measurements for 128 subjects with some later data
- initial log viral loads fall from 3.1 to 3.8 log copies/ml
- same as for all time points combined
- get similar sharp decrease in mean viral load with increasing $\epsilon$ after controlling for initial log viral load
- perhaps relationship different within initial log viral load categories

Adaptive Initial Log Viral Load Levels
- modeled later log viral load versus $\epsilon$ within separate intervals for initial log viral load levels
- adaptively selected cut point for low versus high initial log viral load levels
- fitting separate curves for each level and comparing LCV scores
- no distinct benefit to more than 2 initial levels

Characteristics for Initial Log Viral Load Levels
- 61% of subjects and 63% of later log viral loads fell within the low initial viral load level
- subjects with low initial log viral loads were significantly more likely to be white (53% vs. 24%, p=0.045) and older at baseline (43.7 years vs. 41.0 years, SD 7.4 years, p=0.047)
- not significantly different on a variety of other characteristics

Low Versus High Initial Log Viral Load

Summary
- have demonstrated adaptive repeated measures modeling
- need to extend it to other covariance structures
- need way to evaluate and select models
- need methods for adaptively searching through both fixed and random components
- modeling variances for repeated conditions using fractional polynomials
- selecting fractional polynomials with random as well as fixed coefficients
- need methods for handling repeated measures in logistic and Poisson regression situations

Acknowledgements
Supported in part by NIAID grant R01 AI072451 (OHHS), Yale University’s Center for Interdisciplinary Research on AIDS (CIRA) through NIMH grant P30 MH62294, and OHHS’s Oregon Clinical and Translational Science Institute (OCTSI) through NCLRR grant 1UL1 RR024144. Collection of data supported in part by NIMH Grant R01 NR04744 (Yale University), NCMR SCIR Program M01 RR0123 (Vice-Urban University), and NCBR GCRC Program M01 RR06162 (University of Connecticut Health Center).