# Adaptive Modeling of Longitudinal HIV Viral Load Data

George Knafl[1], Jean O'Malley[1],
Kristopher Fennie[2], Carol Bova[3]
Kevin Dieckhaus[4], Gerald Friedland[2], Ann Williams[2]

[1] Oregon Health & Science University (OHSU)
[2] Yale University
[3] University of Massachusetts Worcester
[4] University of Connecticut Health Center

Supported in part by NIAID Grant R01 AI057043, by Yale University's Center for Interdisciplinary Research on AIDS (CIRA) through NIMH Grant P30 MH62294), and the OHSU Oregon Clinical & Translational Institute (OCTRI) through NCRR Grant UL1 RR024140.

# Overview of Topics

- the adaptive modeling process for repeated measurements $y_t$ over conditions t
  - modeling how the expectation E**y** for the vector **y** with entries $y_t$ depends on predictor variable(s) x
  - may be nonlinear in x which may depend on t
  - accounting for within-subject correlation in t
- the study and its data to be analyzed
  - electronic adherence data and viral load data
- modeling how log (base 10) viral loads $y_t$ at times t depend on prior adherence x

# Modeling the Expectation

- used nonparametric fractional polynomial models
  - E**y** modeled with polynomials in q power transforms $x^p$ of predictors x with associated coefficient vectors $\beta$
  - represented by predictor matrices **X** combining $x^p$ values for a subject over all conditions t
  - for a given **X**, maximum likelihood used to estimate $\beta$
  - **X** determined by adaptively selecting the number q of terms and the powers p for associated transforms $x^p$
  - subject indexes for **y** and **X** left off to simplify notation

for details, see papers in Statistics in Medicine, 23, 783-801, 2004 and in Proceedings of Second IASTED International Conference on Computational Intelligence, ed. B. Bovaruchuk, ACTA Press: Anaheim, CA, 2006, 422-427

# Modeling Variances/Correlations

- need model for within-subject covariance matrix $\boldsymbol{\Sigma}$
- assuming multivariate normal distribution
- first used standard repeated measures approach
  - compound symmetry (CS)
  - with variances the same for all conditions t and correlations the same for all pairs of conditions t and t'
- recently extended to autoregression (AR)
  - correlations weakening the further apart t and t' are
- need way to evaluate alternative models

# Model Evaluation

$$LCV=\Pi_{\text{folds } F}\Pi_{\text{subjects in } F}L(\mathbf{y},\mathbf{X};\beta(F^c),\mathbf{\Sigma}(F^c))^{1/m}$$

- k-fold likelihood cross-validation (LCV)
  - k is number of folds
  - subjects randomly assigned to folds F
    - with equal assignment probabilities 1/k
  - m is number of outcome measurements for all subjects
- compute multivariate normal likelihood L for folds
  - using deleted parameter estimates $\beta(F^c)$ and $\mathbf{\Sigma}(F^c)$
    - parameter values used in likelihood terms for fold F estimated using data for its complement $F^c$
  - for subjects' response vectors **y** and predictor matrices **X**
- larger scores indicate better models **X** for **y**

# Model Selection Process

- model selection occurs in two phases
  - expansion
    - starting from base model, add power transforms $x^p$ of x
  - contraction
    - remove any extraneous transforms from expanded model
    - remaining powers adjusted after each removal
- search controlled by tolerance parameters
  - how much of a decrease in LCV scores can be tolerated
  - continue search as long as penalty in LCV not too high
  - produces model with nearly optimal LCV score
    - usually parsimonious with all its coefficients significant

# 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 outcome variables
  - linear case has been extended to repeated measurements
    - with CS or AR covariance structures

# Example Data

- from study of 172 HIV+ subjects on ARV meds
  - 50.6% (87) randomized to adherence intervention
  - control group received standard care
- adherence data collected electronically
  - with Medication Event Monitoring System (MEMS) caps
- viral loads obtained from medical records
- interview data collected up to 7 times 3 months apart
  - including self-reported adherence

the Adherence through Home Education and Nursing Assessment (ATHENA) Project, PI. A. Williams

collection of data was supported in part by NINR Grant R01 NR04744, NCRR GCRC Program M01 RR00125 (Yale University), and NCRR GCRC Program M01 RR06192 (University of Connecticut Health Sciences Center)

# Electronic Adherence Data

- subjects given pill bottles with MEMS caps
  - caps recorded dates and time for openings and presumably for medication taking
  - medications in pill bottles prescribed at 2 per day
- usable MEMS data for 161 subjects (93.6%)
  - 75,000+ cap openings for 66,000+ days of cap use
  - over about 2½ years from 8/1999 to 3/2002
- used standard summary adherence measure
  - % prescribed doses taken (PDT)

used MEMS IV caps and MEMS Version 2.61 software

# Viral Load Data

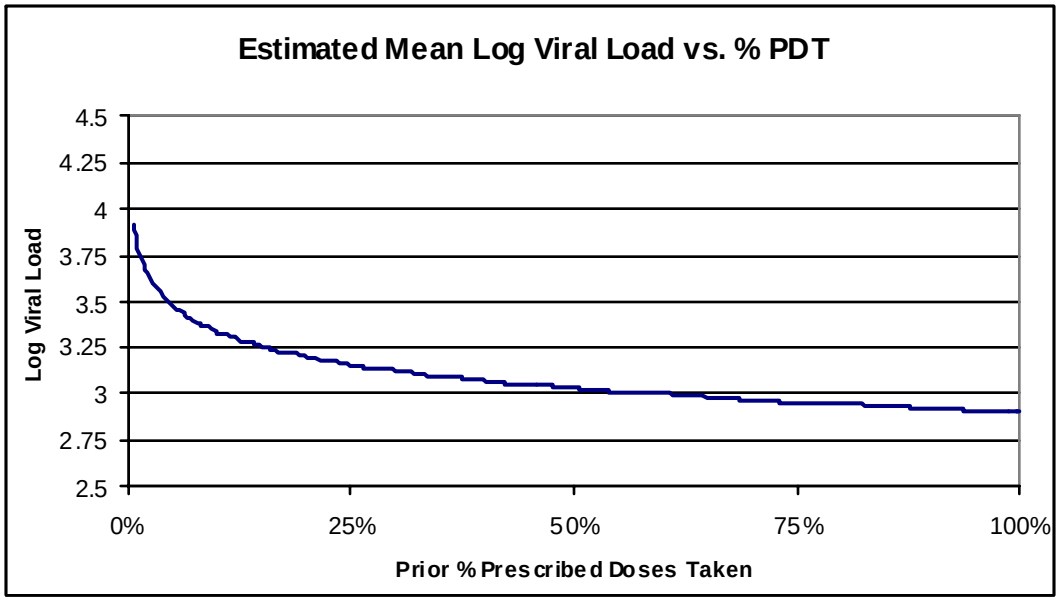
- viral loads obtained from medical records and matched in time to interview dates
  - 643 measurements at up to 7 times for 160 subjects
    - for whom MEMS adherence data also available
  - from 1.30 to 5.88 log copies/mL (20-750,000 copies/mL)
    - viral loads designated as below the detectable limit set to that limit (either 50 or 400 copies/mL)
- modeled log (base 10) viral load in terms of adherence prior to interview dates using % PDT
  - with CS covariance structure

# Log Viral Load vs. Time

- dependence of log viral load on time was completely explained through correlations
  - mean log viral load did not change with time
  - estimated at 3.02 log copies/mL (1047 copies/mL)
- log viral loads were highly correlated but also highly variable
  - estimated $\rho$ was .61, estimated SD was 1.02
- log viral load also did not depend on intervention group membership

# Log Viral Load vs. Prior % PDT

- mean log viral load decreased very quickly
  - from 3.91 to 2.91 log copies/mL as % PDT increased from its observed values of 0.7% to 100%
  - at 25% PDT, decreased 75% of the way to the minimum
  - at 50% PDT, decreased 88% of the way to the minimum
  - at 75% PDT, decreased 95% of the way to the minimum
- did not change much once adherence reached modest levels
- suggests that high levels of adherence may not have much of an impact on viral load
  - unexpected result



# Controlling for Initial Viral Load

- modeled log viral loads at 6 later time points in terms of initial log viral load plus % PDT
  - 485 measurements for 128 subjects with some later data
  - initial log viral loads vary from 1.30 to 5.88 log copies/mL (same as for all time points combined)
- get similar sharp decrease in mean log viral load with increasing % PDT even after controlling for initial log viral load

# Adaptive Initial Viral Load Levels

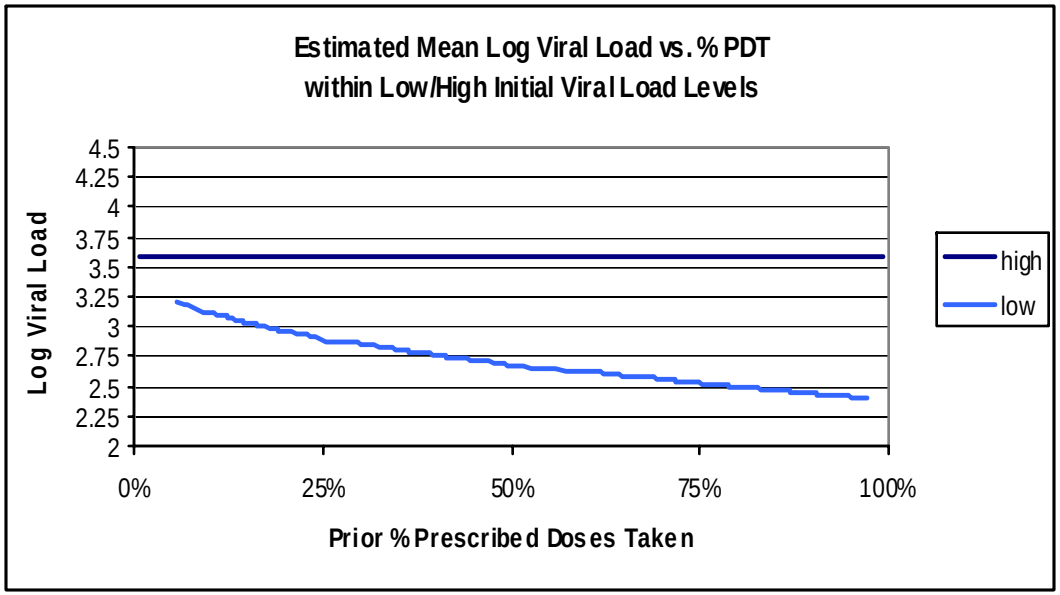
- modeled later log viral load vs. % PDT, but over different initial log viral load intervals
- adaptively selected cut point for low versus high initial log viral load groups
  - fitting separate curves for each group and comparing LCV scores
  - best choice was 2.8 log copies/mL (631 copies/mL)
- no distinct benefit to more than 2 initial levels

# Low/High Initial Viral Load

- the low initial viral load group consisted of 61% of the subjects and 63% of the later log viral loads
- subjects with low initial log viral loads were
  - significantly more likely to be white (p=.040)
    - 53% vs. 34%
  - not significantly more likely to be: male; Hispanic; ever on an NRTI, an NNRTI, or a PI; or have education of at most a high school degree or not
  - significantly older at baseline (p=.047)
    - mean of 43.7 years vs. 41.0 years (SD of 7.4 years)
  - not significantly different on mean HIV duration or mean time on meds

# Low/High Initial Viral Load

- mean log viral load constant for high initial levels
  - at 3.58 log copies/mL (3,802 copies/mL)
- but decreased steadily for low initial levels
  - from 3.20 to 2.39 log copies/mL (1,585 to 245 copies/mL)
- when log viral loads were not too large to start with, there were distinct benefits to high levels of adherence on later log viral loads
- adaptive modeling can uncover novel insights



# Comments

- standard repeated measures model with constant correlations an effective choice for these data
  - provided better depiction (with higher LCV score) of correlations for log viral loads over time than AR
- identification of adherence effects required electronic adherence measure
  - mean log viral load over all 7 time points was constant in self-reported adherence
    - as measured by the % of prescribed medications the subject reported taking in the 3 days prior to an interview

# Summary

- have demonstrated adaptive modeling
- working on extension to other covariance structures
  - e.g., ARMA and spatial autoregression
- and for adaptively searching through both fixed and random components
  - modeling variances for repeated conditions using fractional polynomials (recently completed)
  - adaptively selecting fractional polynomials with random as well as fixed coefficients
- and for handling repeated measures in logistic and Poisson regression situations
  - i.e., for repeated categorical or count measurements