Nonparametric tests for two group comparisons of dependent observations obtained at varying time points with application to RNA viral load decline

Susanne May
Division of Biostatistics and Bioinformatics, University of California San Diego
SMay@ucsd.edu

In collaboration with
Victor DeGruttola, Harvard University

Outline

- Motivation
- Two new tests
- Example data / Simulations
- Asymptotics
- Comparison to other tests
- Summary

Motivation

- AIEDRP
  - Acute HIV Infection and Early Disease Research Program
- Research question:
  - RNA decline slower with transmitted drug resistance?
- Study group: Tx naïve HIV+ patients who start ARV

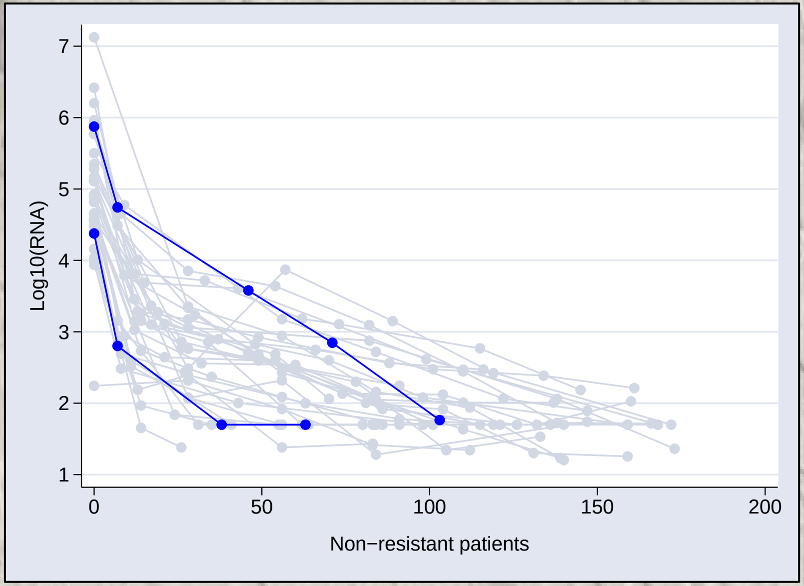
Motivation

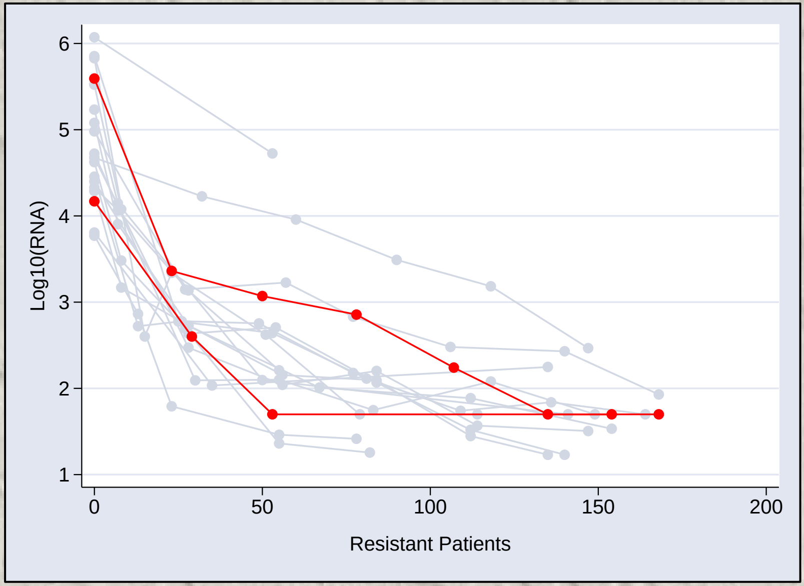
- 15-20 drugs available, 3 drug classes
- Regimen of 3-4 drugs
- Virus mutating
- Resistance to drug(s), drug classes
- Transmitted to uninfected individual
- Newly infected has drug resistant virus

- Outcome: Decline in RNA viral load over time
  Viral load can be censored, above and below
- Groups: Resistant vs Sensitive

Motivation

- **AIEDRP Data, Los Angeles and San Diego**





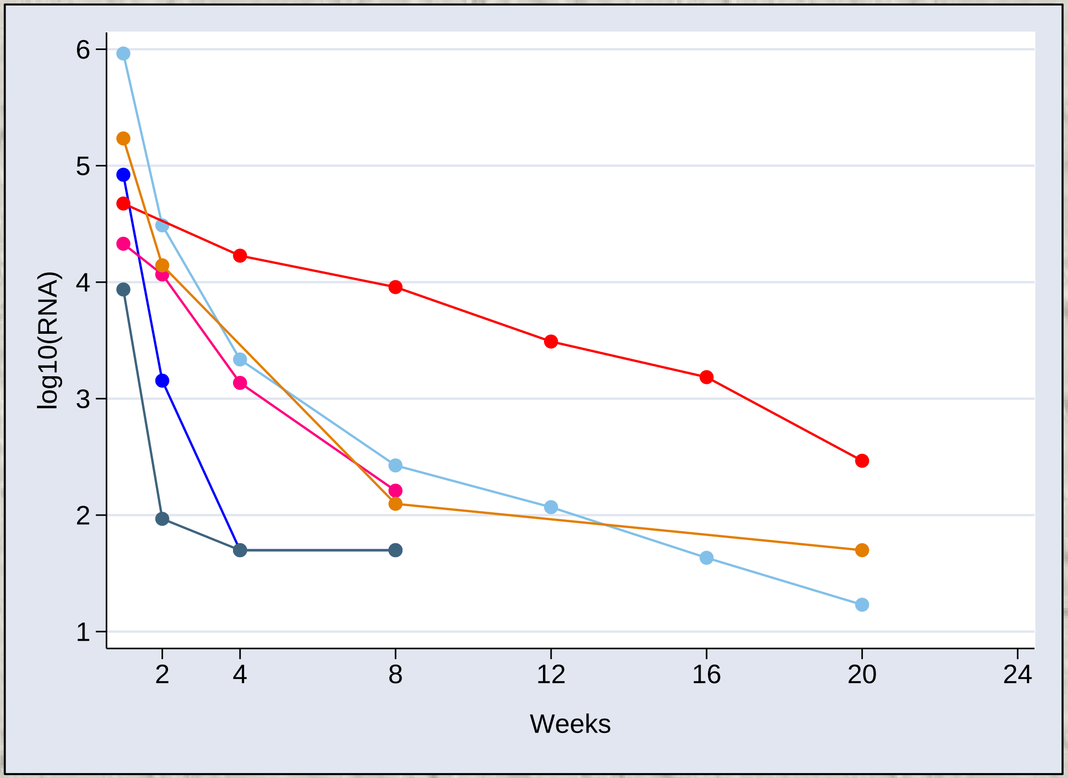
Sensitive **Resistant**

Motivation

- Wei and Johnson (1985, Biometrika)
  - Same follow-up schedule, all patients
  - Test at each time point
  - Combine across time points
- Yao, Wei and Hogan (1998, Biometrika)
  - Shift model
  - Incomplete repeated measures
  - Informative censoring
  - Does not require same follow-up schedule
- Others ...

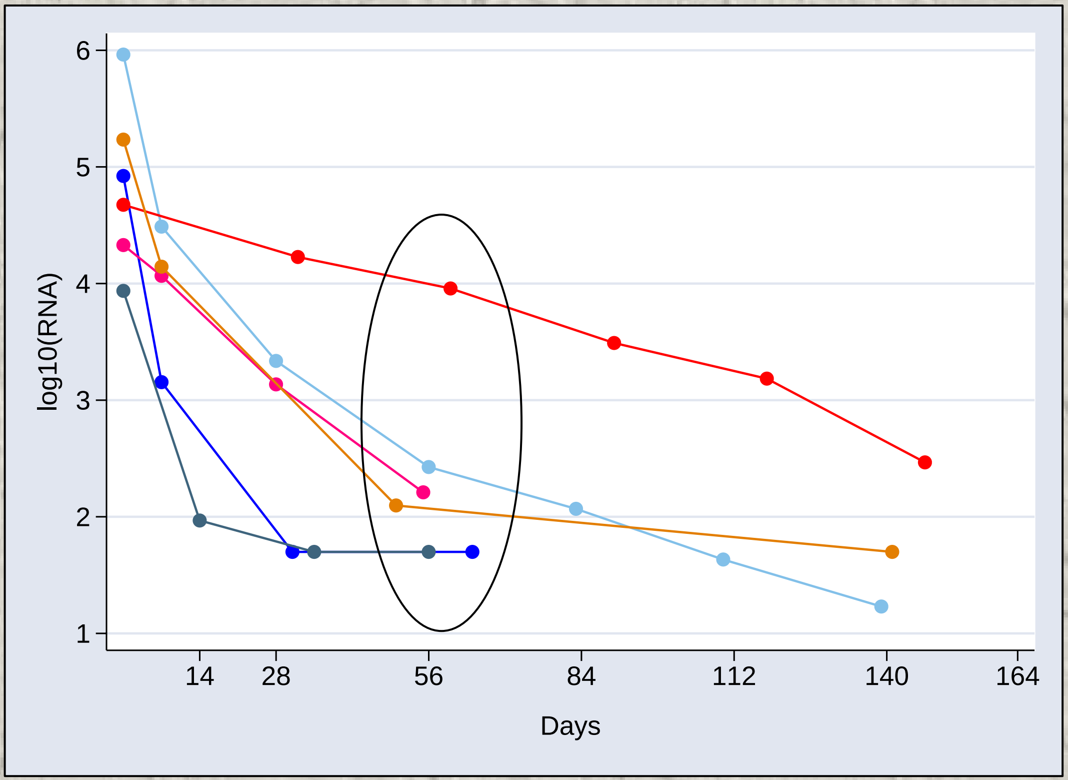
Motivation

- **Same follow-up schedule**



Motivation

- ... in reality



Outline

- Motivation
- Two new tests
- Example data, Simulations
- Asymptotics
- Comparison to other tests
- Summary

New Tests

- **Assume for sensitive and resistant groups**

$$X_{ik} = \mu(t_{ik}) + \varepsilon_i(t_{ik}) \qquad i = 1,\ldots,m \quad k = 1,\ldots,c_i$$

$$Y_{j\ell} = \eta(t_{j\ell}) + \delta_j(t_{j\ell}) \qquad j = 1,\ldots,n \quad \ell = 1,\ldots,c_j$$

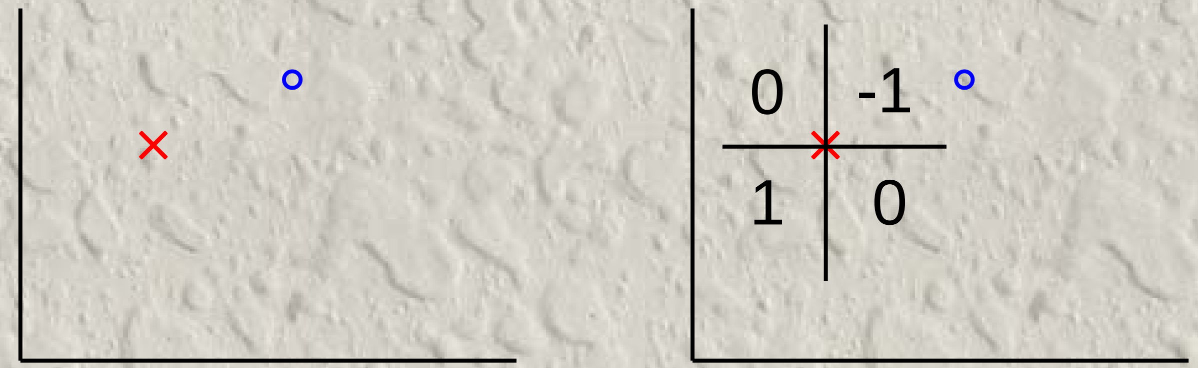
- **Hypothesis**

  H_0: $\mu(t) = \eta(t)$

  H_A: $\mu(t) = \eta(t) + \rho(t)$ **,** $\rho(t) > 0$ **or** $\rho(t) < 0$

New Tests

- **General idea**



- **Score**

$$= \begin{cases} 1 & \text{if } X_{ik} < Y_{j\ell} \text{ and } t_{ik} \le t_{j\ell} \\ -1 & \text{if } X_{ik} > Y_{j\ell} \text{ and } t_{ik} \ge t_{j\ell} \\ 0 & \text{otherwise} \end{cases}$$

Test 1

- **Test statistic**

$$U_1 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{k=1}^{c_i}\sum_{j=1}^{n}\sum_{\ell=1}^{c_j}\Theta\Big(\big(X_{ik},t_{ik}\big),\big(Y_{j\ell},t_{j\ell}\big)\Big)-\hat{\theta}_{ikj\ell}$$

where $\hat{\theta}_{ikj\ell}$ **estimates** $E\Big[\Theta\Big(\big(X_{ik},t_{ik}\big),\big(Y_{j\ell},t_{j\ell}\big)\Big)\Big]$

$$\Theta\Big(\big(X_{ik},t_{ik}\big),\big(Y_{j\ell},t_{j\ell}\big)\Big)=\begin{cases}1 & \text{if } X_{ik}<Y_{j\ell} \text{ and } t_{ik}\le t_{j\ell}\\ -1 & \text{if } X_{ik}>Y_{j\ell} \text{ and } t_{ik}\ge t_{j\ell}\\ 0 & \text{otherwise}\end{cases}$$

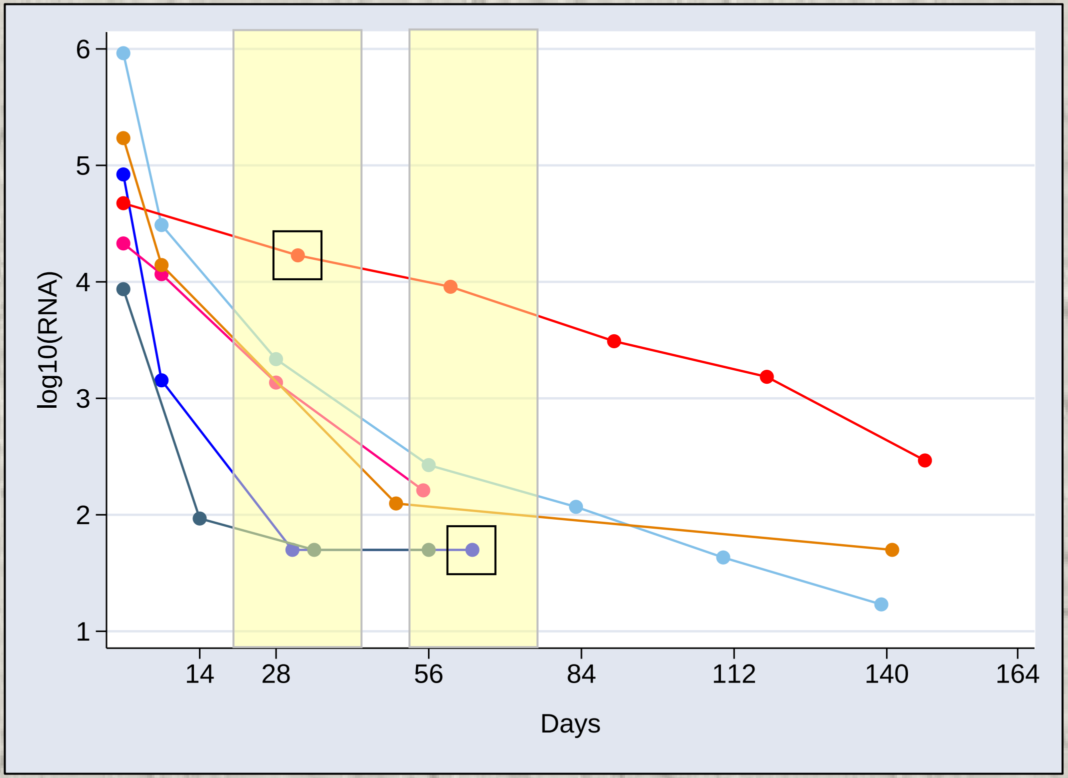
Test 1

- **How to estimate** $E\left[\Theta\left(\left(X_{ik},t_{ik}\right),\left(Y_{j\ell},t_{j\ell}\right)\right)\right]$ **?**
  - **Form intervals** I_{ik} **and** $I_{j\ell}$ **around** t_{ik} **and** $t_{j\ell}$
  - **Calculate scores**
  - **Use all observations in intervals**
  - **Divide by number of scores**
  - **Separately for each group, then combine**

$$\hat{\theta}_{ikj\ell} = \sum_{t_{i^*k^*}\in I_{ik}} \sum_{t_{j^*\ell^*}\in I_{j\ell}} \Theta\left(\left(Z_{i^*k^*},t_{i^*k^*}\right),\left(Z_{j^*\ell^*},t_{j^*\ell^*}\right)\right) \Big/ d_{I_{ikj\ell}}$$

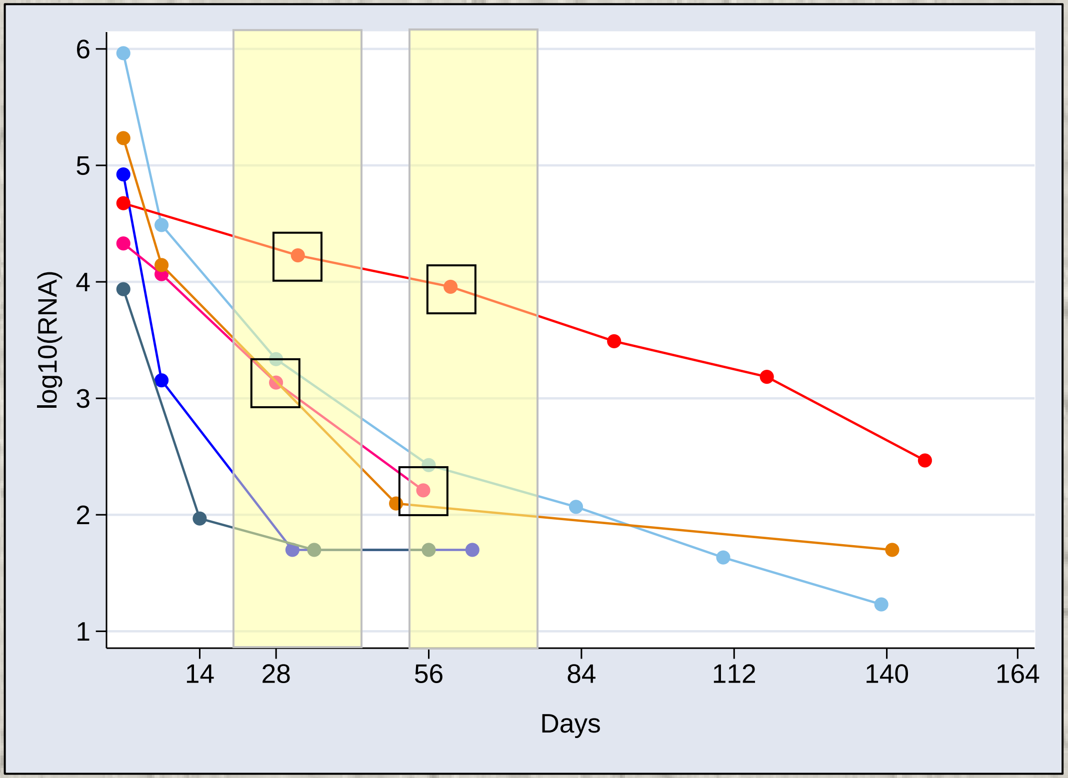
Test 1

- **How to calculate the expected score**



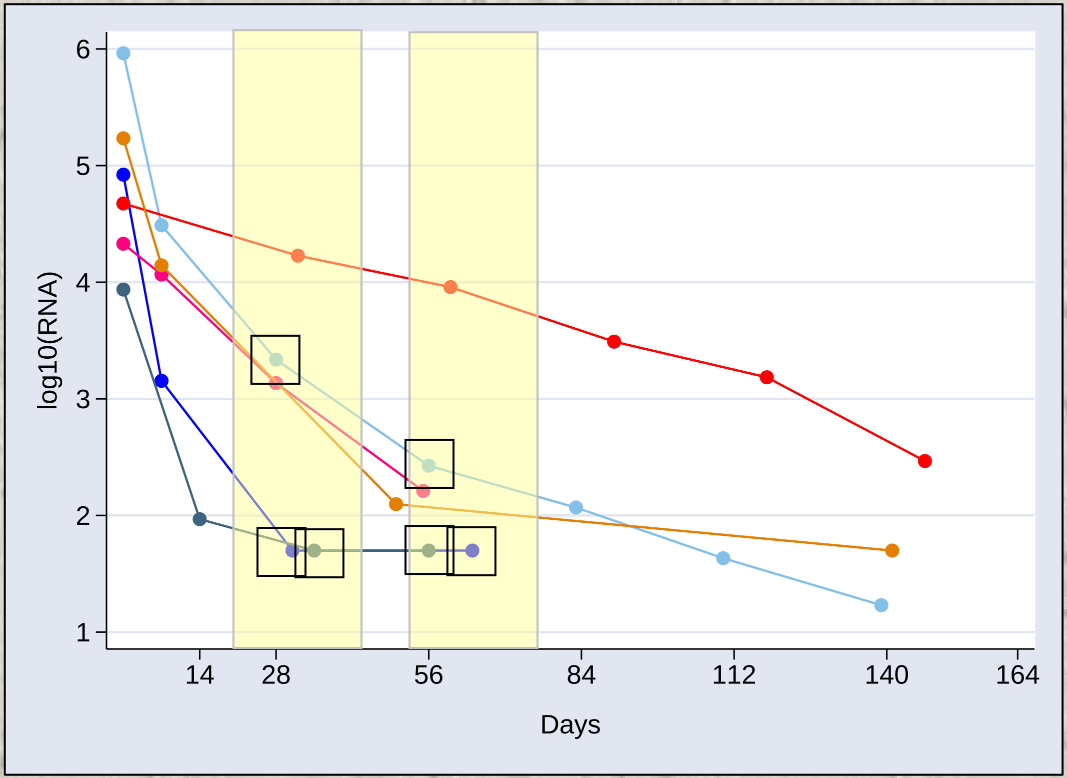
Test 1

- **How to calculate the expected score**



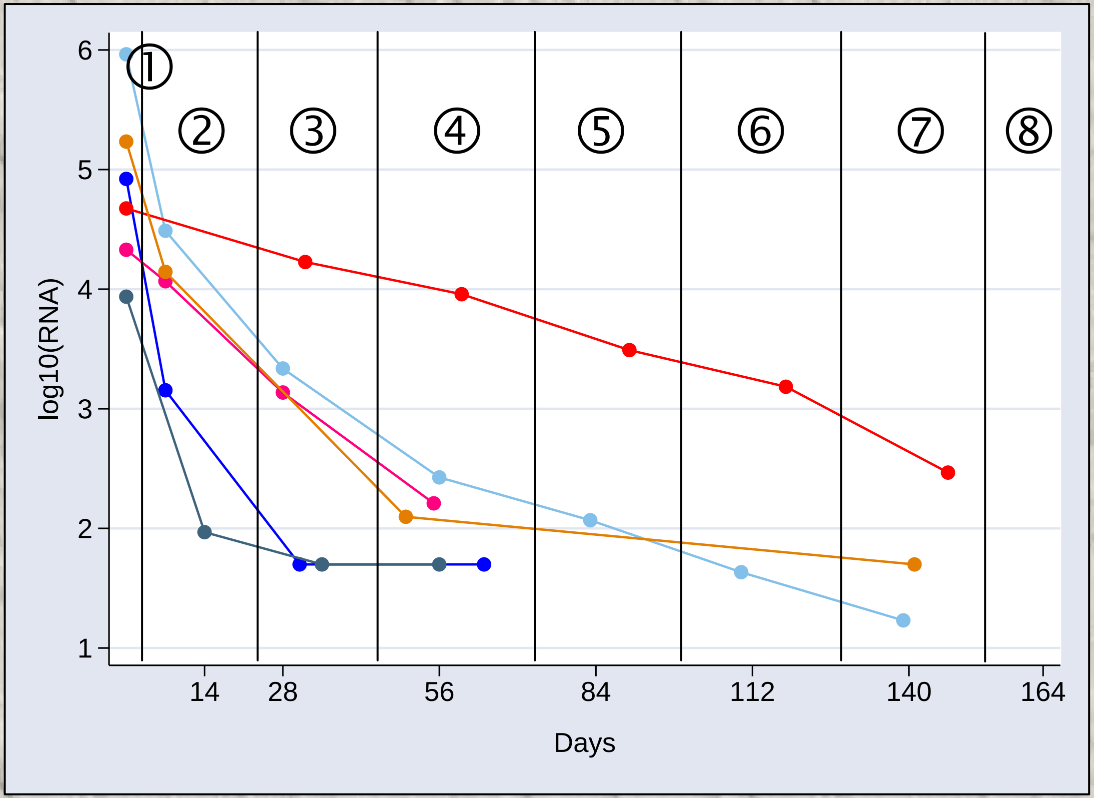
Test 1

- **How to calculate expected score**



Test 2

- **Form "bins" around follow-up visits**



Test 2

- **Score within bins**
- **Weight by inverse of covariance matrix**
- **Assume discrete number of time points**

$$U_3 = \frac{\sqrt{m+n}}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\Theta_{i1j1} - \hat{\theta}_{i1j1}, \ldots, \Theta_{iBjB} - \hat{\theta}_{iBjB} \right) \Sigma^{-1} \begin{pmatrix} \Theta_{i1j1} - \hat{\theta}_{i1j1} \\ \vdots \\ \Theta_{iBjB} - \hat{\theta}_{iBjB} \end{pmatrix}$$

$$\sigma_{pq}^2 = \frac{m+n}{(mn)^2} \sum_{m_p \times n_p} \sum_{m_q \times n_q} \left(\Theta\left(X_{ip}, Y_{jp}\right) - \hat{\theta}_{ipjp} \right) \left(\Theta\left(X_{i'q}, Y_{j'q}\right) - \hat{\theta}_{i'qj'q} \right)$$

- **Obtain *p*-values via re-sampling**

Censored observations

- **Due to measurement limits**

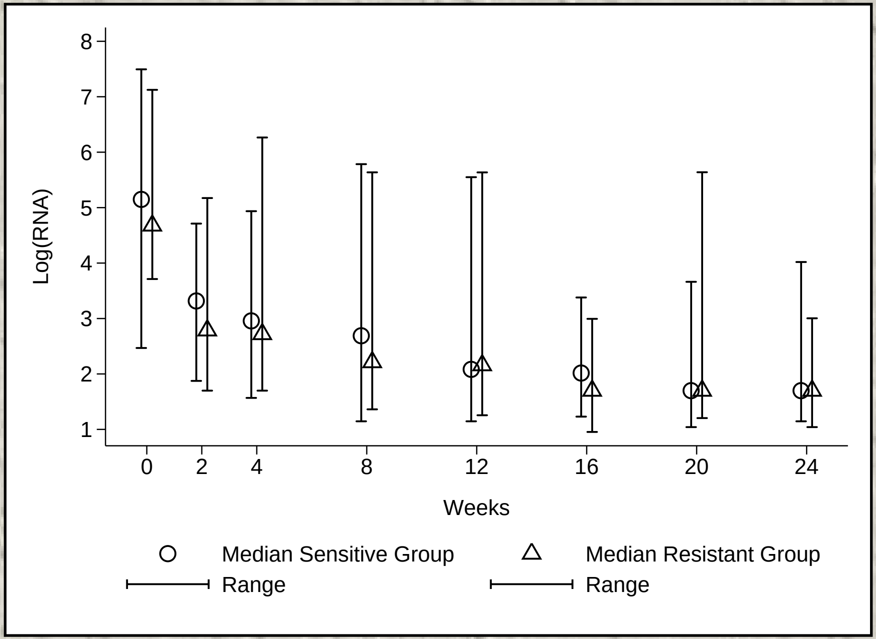
$$\textbf{Score} = \begin{cases} 1 & \text{if } X_{ik} < Y_{j\ell} \text{ and } t_{ik} \le t_{j\ell} \\ & \textbf{and } \boldsymbol{X} \textbf{ is not censored from above} \\ & \textbf{and } \boldsymbol{Y} \textbf{ is not censored from below} \\ -1 & \text{if } X_{ik} > Y_{j\ell} \text{ and } t_{ik} \ge t_{j\ell} \\ & \textbf{and } \boldsymbol{X} \textbf{ is not censored from below} \\ & \textbf{and } \boldsymbol{Y} \textbf{ is not censored from above} \\ 0 & \text{otherwise} \end{cases}$$

Outline

- Motivation
- Two new tests
- Example data / Simulations
- Asymptotics
- Comparison to other tests
- Summary

Example data – AIEDRP

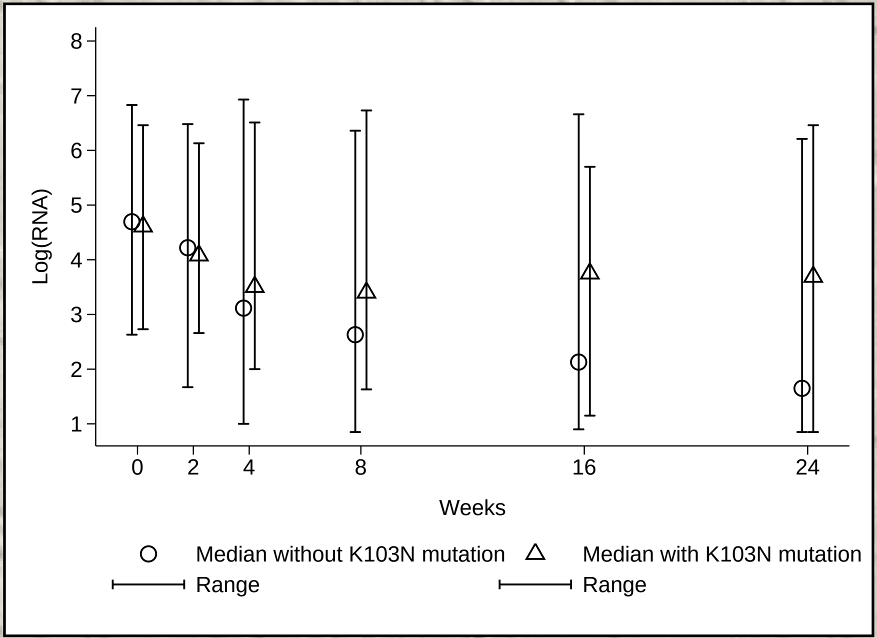
- **93 sensitive, 33 resistant patients**



- **U_1: p-value 0.22; U_3: p-value 0.20**
- **Wei-Johnson: p-value 0.30**

Example data – ACTG398

- **292 without, 64 with K103 mutation**



- ***U_3*: *p*-value 0.004**
- **Wei-Johnson: *p*-value 0.03**

Simulations

- Power: U_3 versus U_1
  - Highly correlated responses over time
  - U_3 higher power than U_1
- Power: U_3
  - Week 2, 4, 6, linear decline
  - Re-sampling: 1000
  - Simulations: 2000
  - Differences in slope -15.0, -10.0, -7.5, -5.0
  - SD: 11.8
  - Autoregressive(1) covariance, $\rho = 0.7$

Simulations – cont.

- **Power: various effect sizes**

Diff in β	Total sample size		
	N = 20	*N* = 60	*N* = 80
-15.0	0.44	0.68	0.86
-10.0	0.32	0.53	0.70
- 7.5	0.27	0.42	0.58
- 5.0	0.19	0.29	0.46

Outline

- Motivation
- Two new tests
- Example data / Simulations
- **Asymptotics**
- Comparison to other tests
- Summary

Asymptotics

- *"aU"*-statistics (*a*lmost *U*-statistics)
- Kernel includes unknown parameter
- Randles (1982) or Lee (1990)
- U_3 is asymptotically $\chi^2_{(B)}$

Outline

- Motivation
- Two new tests
- Example data / Simulations
- Asymptotics
- Comparison to other tests
- Summary

Comparison to other tests

- Yao, Wei and Hogan (1998, Biometrika)
  - Shift model
  - Allow for informative censoring (horizontal)
  - Do not make use of covariance to improve on efficiency
  - Variance estimates depend on estimated shift parameter

Comparison to other tests

- Functional ANOVA
  - Does not rely on parametric assumptions
  - Modeling longitudinal data using splines

  - Is not invariant to monotone transformations of outcome or time
  - Does not easily accommodate censoring of outcome values

Outline

- Motivation
- Two new tests
- Example data / Simulations
- Asymptotics
- Comparison to other tests
- Summary

Summary

- Computationally intensive
- Conceptually easy
- Distributions of time points of obs do not have to be the same
- Non-parametric
- Invariant to monotone transformations of data

Summary – cont.

- Prob of missing obs can depend on outcome value if same in both groups
- Censoring (e.g. of RNA values) can be accommodated easily

- Variation: score within bins and across neighboring bins
- Inverting
- Regression

May S. and DeGruttola V. (2007) “Nonparametric Tests for Two Group Comparisons of Dependent Observations Obtained at Varying Time Points”, *Biometrics,* 63, 194 - 200*.*

Funded by: U01 AI043638 and R01 AI 51164

Thanks to

- Lin Liu
- Ron Thomas
- LJ Wei
- Ed Gehan

Thank you!