Optimizing Diagnosis of HIV Treatment Failure with Selective Use of Gold Standard Test

Tao Liu, PhD

Brown University

NESS @ Harvard
April 26, 2014

Background

HIV in the developing world

- Most ($> 95\%$) of HIV-infected people live in resource limited settings (RLS).
- To control the spread of the virus, it is critical to provide proper treatment and care to these people.
- Key decision points in monitoring disease progress
  - When to initiate ART
  - Deciding when 1st-line ART is failing

Focus of this talk:

Diagnosis of 1st-line ART failure

1st-line ART failure = detectable viral load

Background

Diagnosis and decision making in RLS

Correct diagnosis of 1st-line ART failure is important...

- Failure to diagnose leads to:
  - Prolonged viral replication; extra clinical cost associated with OI; risk of transmitting resistant strains of virus
- Incorrect diagnosis leads to
  - More expensive 2nd-line therapy; accelerated progression toward 2nd-line resistance
- Europe / N. America
  - CD4 and VL measures about every 3 months
  - Genotyping to determine resistance to therapy
- Developing world (Africa, SE Asia, S America)
  - Monitoring HIV treatment in RLS is usually done in substandard conditions, compared with Europe & N America.

Diagnosing 1st line ART failure

Current guidelines for RLS

- In RLS, diagnoses and treatment decision usually based on limited amount of information
  - CD4 count (current, most recent, etc)
  - WHO stage
  - Clinical impressions

WHO recommendations for diagnosing 1st-line failure

Use CD4 count where available

Diagnose 'failure' if :

- CD4 < 200; recently CD4 < 350.
- More than 25% relative decrease in CD4 over 6 months.

Diagnosing 1st-line ART failure

Diagnostic accuracy of WHO approach

- Consensus has emerged about inadequacy of WHO criteria
  - Bisson et al. (2006, 2008); Deeks et al. (2000, 2002); Moore et al. (2005); Schecter and Tuboi (2006); Tuboi et al. (2007) Reynolds et al. (2009); Mee et al. (2008) among others.
- Example: 40% classified as treatment failures incorrectly diagnosed (Kantor et al., 2009)
- Research on improving diagnosis tries to optimize use of non-VL markers
  - Badri and Wood (2003); Bagchi et al. (2007); Bedell et al. (2003); Kantor et al. (2009); Mahajan et al. (2004); Foulkes et al. (2010)

Selective use of VL testing

Key idea

- Viral load (VL) can be regarded as the gold standard test.
- Is becoming available to HIV patients although on a limited basis (e.g. Kenya)
- We assume that
  - VL is available on a limited basis; that is, can obtain VL for fixed percentage of visits
  - VL will give correct diagnosis when applied

Question:

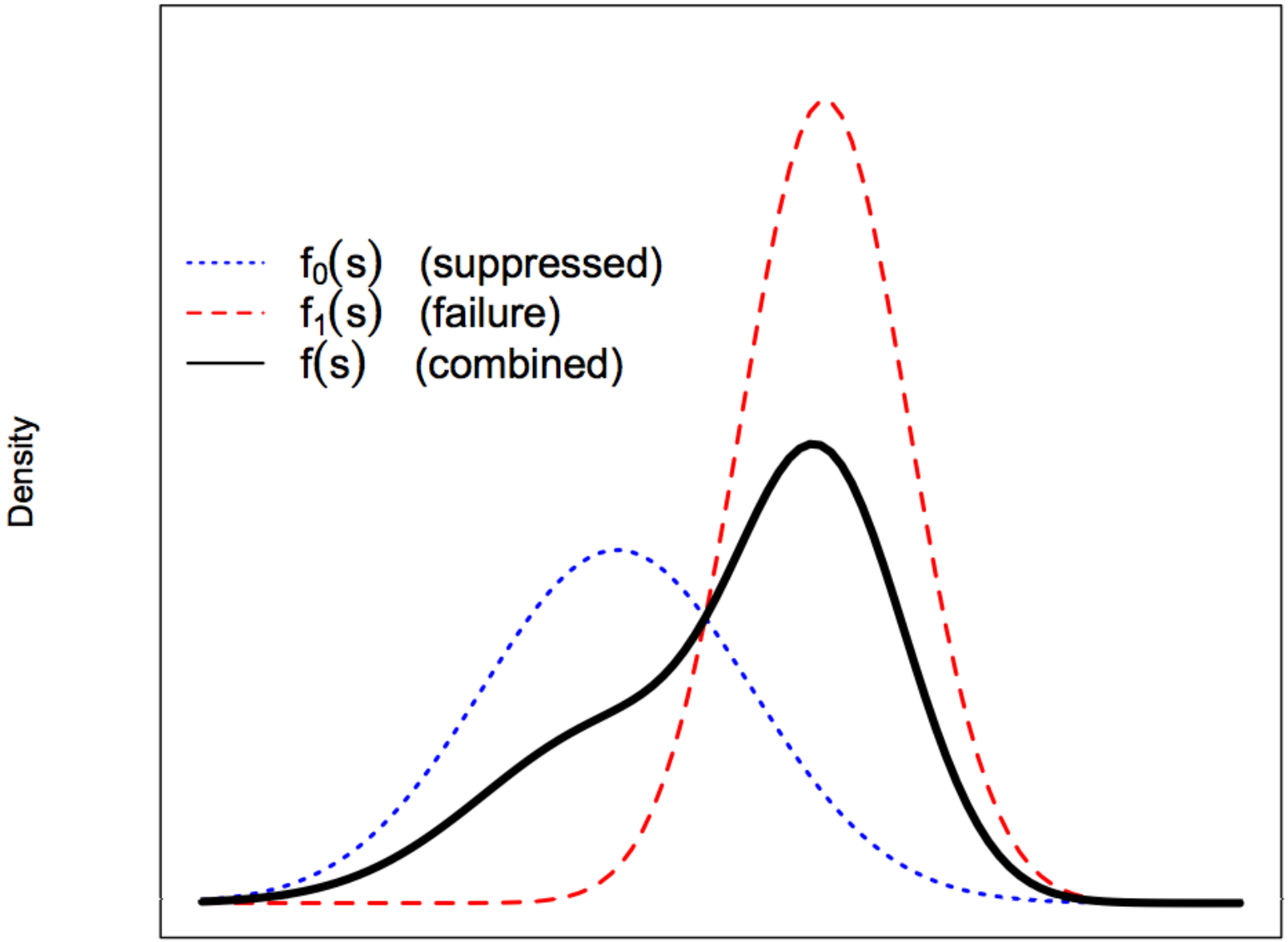
How to use the limited VL to optimize diagnostic accuracy?

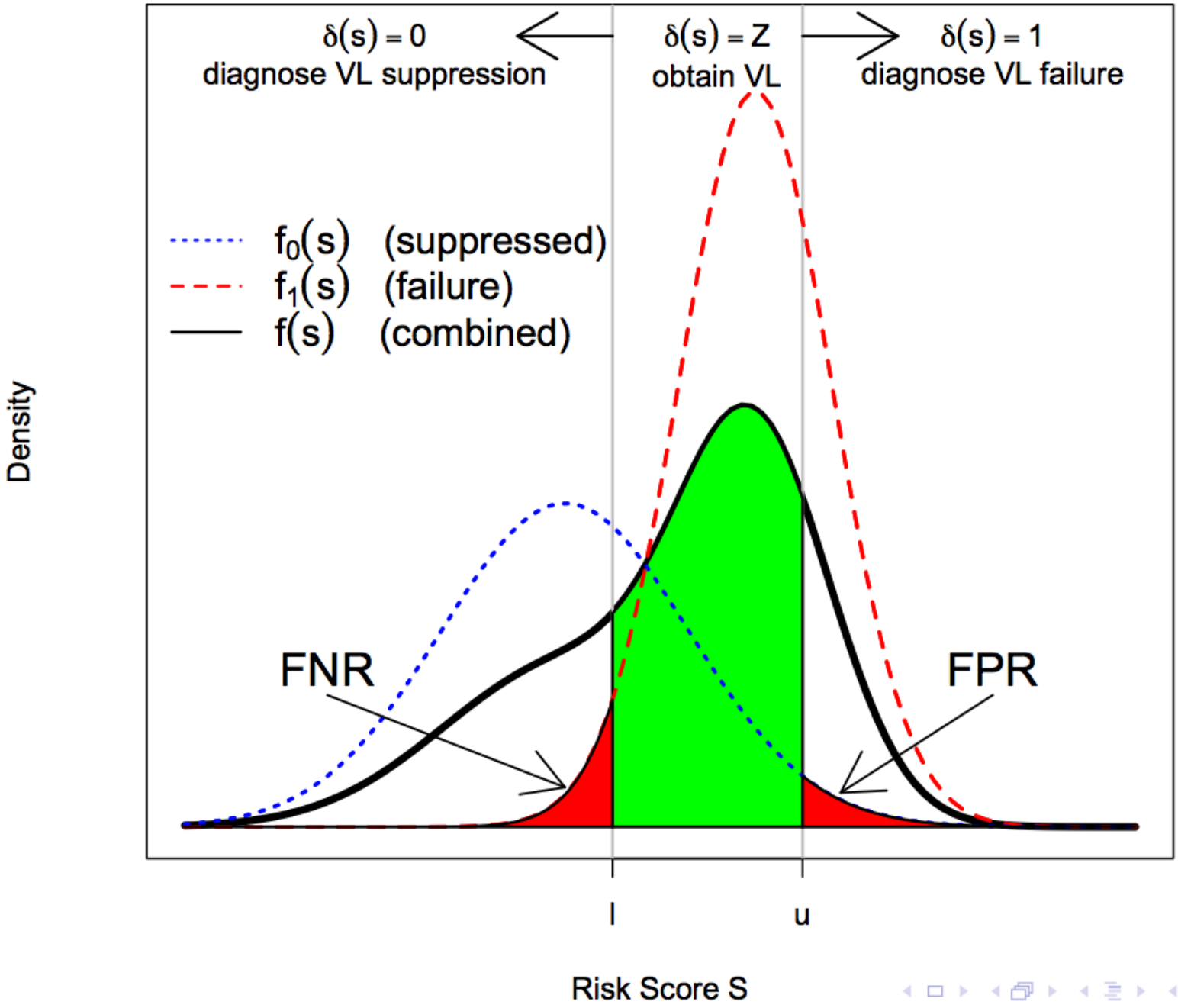
Selective use of VL testing

Approach

Selective use of VL in conjunction with available clinical tests

- Procedure:
  1. Use clinical marker(s) to obtain risk core S
  2. Obtain VL *only* for a subset with $S \in [l, u]$
     - $S < l \Rightarrow$ diagnose as failing
     - $S > u \Rightarrow$ diagnose as not failing
- Misclassifications will only occur for diagnoses based on S alone
- Statistical Question: *If percent of VL is fixed at ϕ, how to determine $[l, u]$ to minimize diagnostic error rates?*


$f_0(s)$ (suppressed)
$f_1(s)$ (failure)
$f(s)$ (combined)
Density
Risk Score S


δ(s) = 0
diagnose VL suppression
δ(s) = Z
obtain VL
δ(s) = 1
diagnose VL failure
f0(s) (suppressed)
f1(s) (failure)
f(s) (combined)
Density
FNR
FPR
l
u
Risk Score S

Formalizing the question

Notations

- $\mathbf{X}$ = vector of clinical markers = {CD4, CD4%, WHO stage, etc.}
- $S = S(\mathbf{X})$ is a scalar risk score
- Δ = true underlying VL status

$$\Delta = \begin{cases} 1 & \textit{if VL detectable} \\ 0 & \textit{if not} \end{cases}$$

- $p = \Pr(\Delta = 1)$ is prevalence of ART failure
- $F_1(s)$, $F_0(s)$ are risk score distribution for $\Delta = 1, 0$. So the population distribution of S is

$$F = pF_1 + (1-p)F_0.$$

- We consider two assumptions:
  1. $F_1 \succ F_0$.
  2. $f_1(s) \propto \exp(\beta_0 + \beta_1 s) f_0(s)$ (exponential tilt model)

Rule The classification rule maps a risk score S to a an action

$$R(s) : s \in \mathbb{R} \mapsto a,$$

where $a = 1$ means diagnose failure; $a = 0$ non-failure.

Tripartite rule Our rule is

$$R_{l,u}(s) = \begin{cases} 0 & \text{if } s \leq l \\ \Delta & \text{if } l < s \leq u \\ 1 & \text{if } s > u \end{cases}$$

Constraint The cut-offs l, u on S is subject to a constraint on the proportion ϕ of visits where a VL can be ordered

$$F(u) - F(l) \leq \phi.$$

ROC analysis of tripartite rules

Definition

Similar to conventional ROC analysis...

- Basic idea:
  - Compute all SE and SP for all rules $R_{l,u} \in \mathcal{R}_\phi$.
  - Make plot of SE vs $(1 - \text{SP})$.
- More formally, the ROC curve is

$$C_\phi(v) : t \in [0,1] \mapsto F_1 \circ H_\phi \circ F_0^{-1}(1-t),$$

  where $H_\phi(u) = \arg\inf_t \{F(u) - F(t) \leq \phi\}$.
- The function H_ϕ dictates that for each u and its associated FPR, the FNR is calculated based on the lower cutoff $l = H_\phi(u)$.

- *Nonparametrically*, ROC curve can be estimated by replacing F_0, F_1 and F by their empirical CDFs.
  - The resulting estimate of ROC curve is consistent.
  - The $\sqrt{n}$ estimate of ROC curve converges to a mixture of Brownian bridges.
- If the exponential tilt model (i.e. $f_1 = \exp(\beta_0 + \beta_1 s) f_0$) is assumed, we can use the approach of Qin (1999) to estimate F_0 and β, and then estimate ROC *semiparametrically*.

- Area under the ROC curve is a useful measure of diagnostic accuracy

$$AUC_\phi = \int C_\phi(v)\mathrm{d}v.$$

- Properties
  - (Interpretation) If $S \sim F_1$ and $S' \sim F_0$, then

$$AUC_\phi = \Pr\{S > H_\phi(S')\}.$$

  - (Bound) The AUC_ϕ is bounded

$$AUC_\phi \geq \frac{1}{2} + \phi + \frac{\phi}{2}.$$

Optimal rule selection

Lost and risk

Loss function captures misclassifications, e.g.

$$\begin{aligned} L_1(a, \Delta) &= \mathbf{1}(a \neq \Delta) \\ L_2(a, \Delta; \lambda) &= \lambda \mathbf{1}(a \neq \Delta = 1) + (1 - \lambda)\mathbf{1}(a \neq \Delta = 0), \end{aligned}$$

where $\lambda \in [0, 1]$ prioritizes weight to FP vs FN.

Risk function is expected loss associated with a rule R

$$risk(R) = \mathbb{E}_{(S,\Delta)}[L\{R(S), \Delta\}].$$

e.g.

$$\begin{aligned} \mathbb{E}\{L_1\} &= \textit{total misclassification rate}\,(TMR) \\ \mathbb{E}\{L_2\} &= \lambda pFNR + (1 - \lambda)(1 - p)FPR \end{aligned}$$

Optimal rule selection

- Selection the optimal tripartite rule can be framed as a *constrained optimization problem.*
  - Can be nonparametric with F_0 and F_1.
    - Determine $[l, u]$ via grid search.
  - Can be semiparametric, assuming exponential tilt model for F_0 and F_1.
    - Analytical solutions of l and u exist under L_1.
    - Optimal rules for different ϕ have a common center, $\frac{l+u}{2} = -\frac{\beta_0}{\beta_1}$.
- Inference
  - Evaluate the resulting rule using K-fold cross validation.

Application: Data from the Miriam Hospital Immunology Center, RI

Description of data

Table 2. Summary statistics for key variables ($n = 597$)

Marker	Mean	Median	IQR	Range
Virological marker				
VL at most recent visit (copies/mL)	11.8 K	75	(75, 400)	(12, >500 K)
Immunological markers				
CD4 count at most recent visit (cells/uL)	442	407	(254, 576)	(8, 1412)
6-month CD4 count change (%)	7.3	18	(−13, 33)	(−80, 736)
CD4 % at most recent visit	24	23	(17, 30)	(0.90, 59)
6-month CD4% change (%)	9.5	4.7	(−6.1, 16)	(−74, 209)

NOTE: K: thousand; IQR: interquartile range.

Application

Constructing risk scores

Two risk scores examined

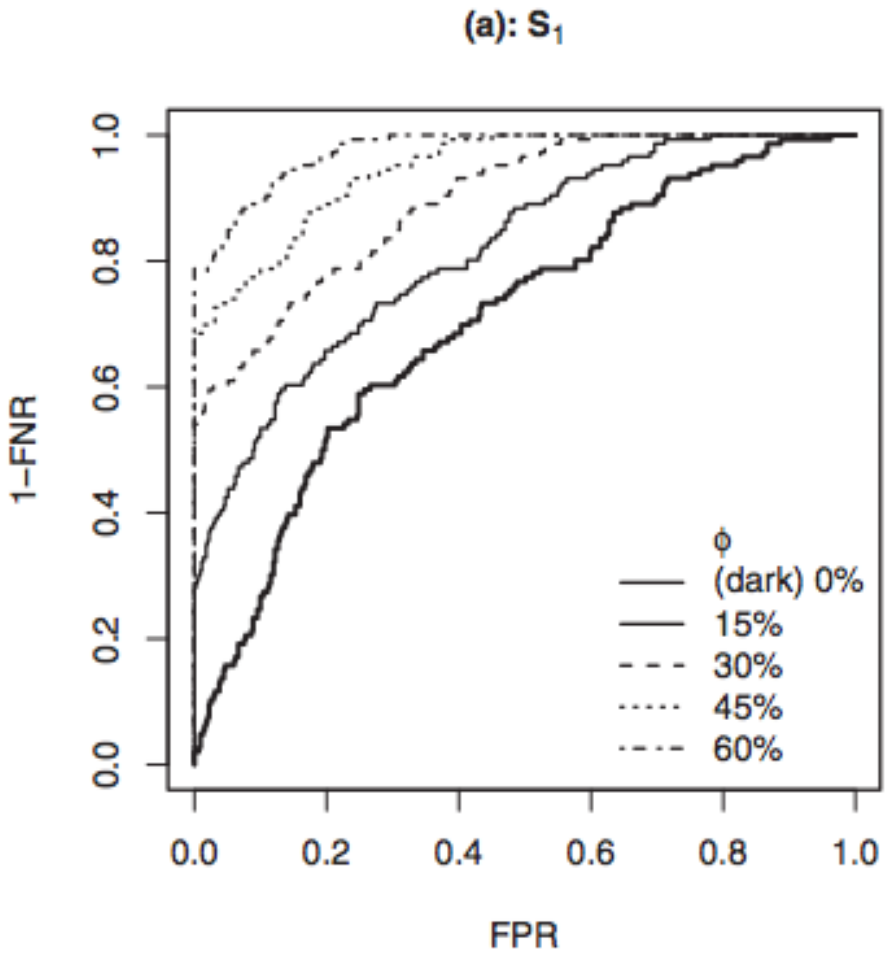
- $S_1 = -CD4$ count
- $S_2 = \mathbf{X}^\top \theta$, where

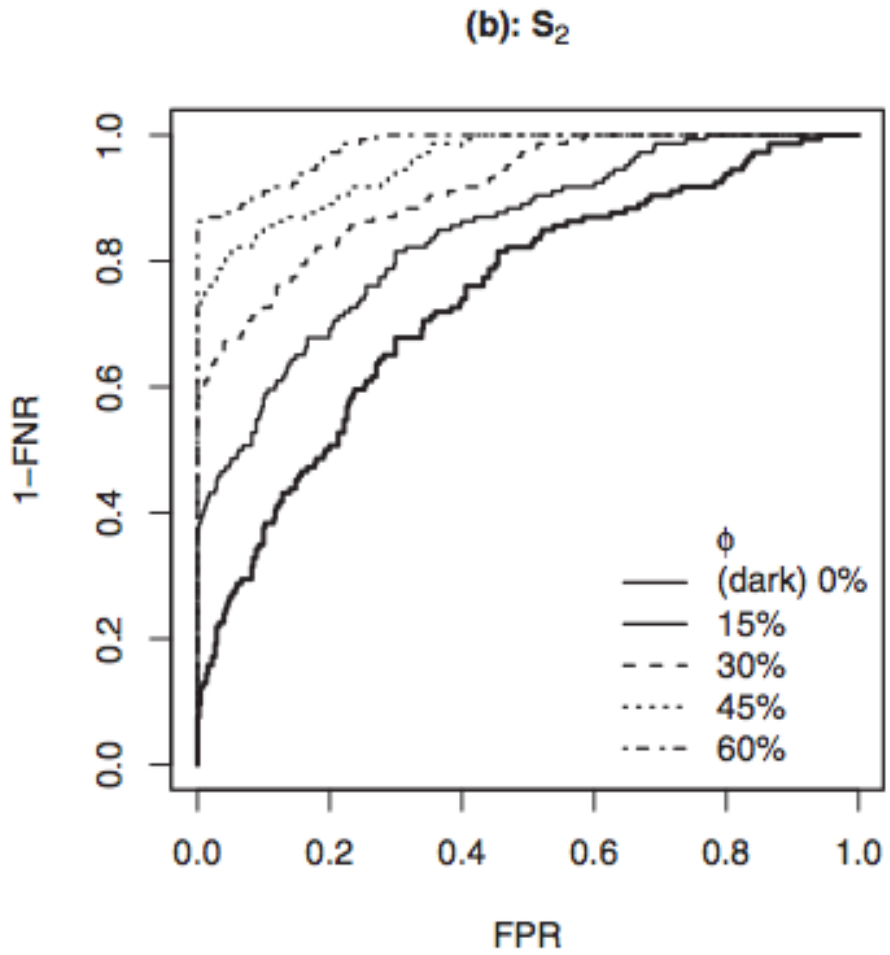
$$\mathbf{X} = \{CD4,\ CD4\%,\ \textit{chgs in } CD4 \textit{ and } CD4\%\}$$

and θ from the model

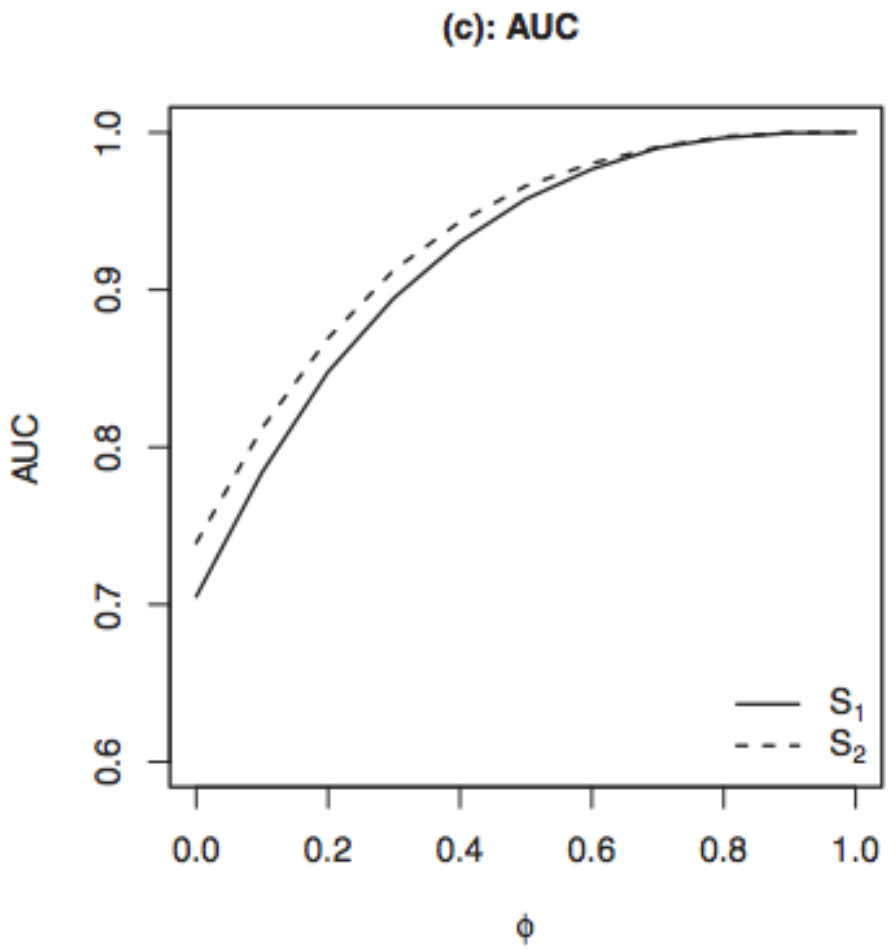
$$\text{logit}\{\Pr(\textit{failure} \mid \mathbf{X})\} = \mathbf{X}^\top \theta.$$

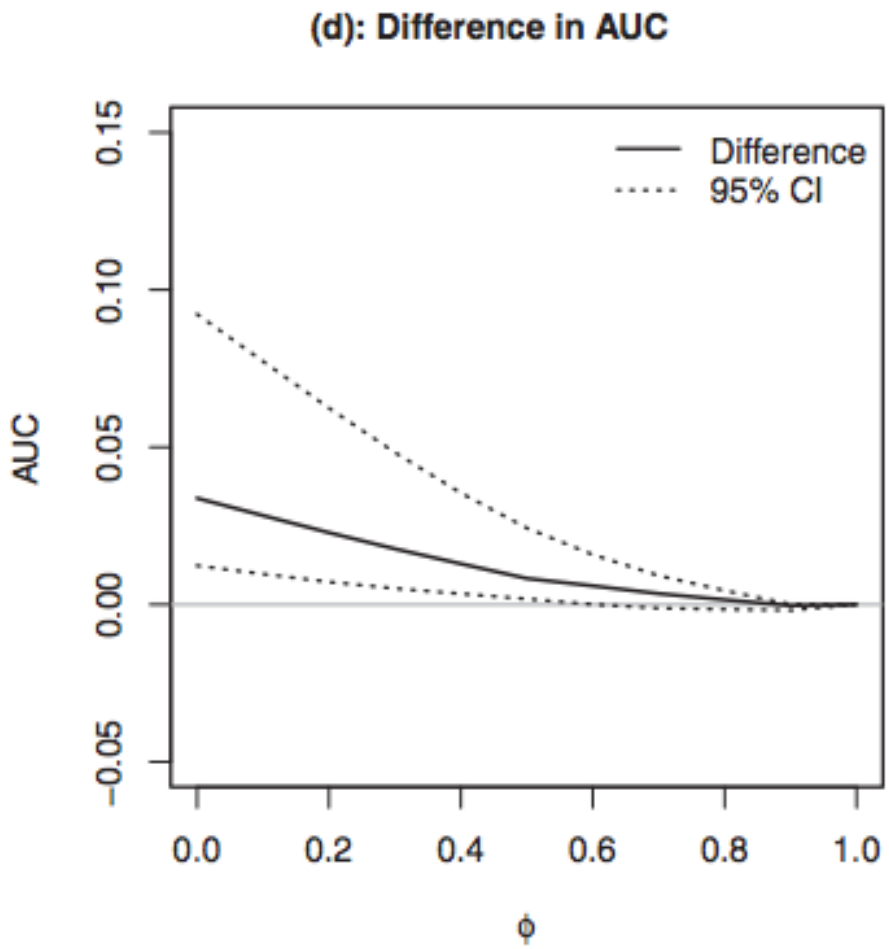
ROC analysis





ROC analysis





Optimal rule under $\mathbb{E}(L_2)$

λ	ϕ	Optimal cutoffs l	Optimal cutoffs u	FNR	FPR	TMR
1/4	0	17 (17)	17 (17)	.98 (.02)	.00 (.00)	.26 (.02)
	.15	17 (9)	201 (18)	.72 (.04)	.00 (.00)	.18 (.01)
	.30	13 (13)	284 (13)	.46 (.04)	.00 (.00)	.12 (.01)
1/2	0	120 (64)	120 (64)	.93 (.13)	.03 (.04)	.26 (.02)
	.15	90 (16)	216 (39)	.63 (.06)	.02 (.01)	.17 (.01)
	.30	17 (15)	284 (32)	.45 (.04)	.01 (.01)	.12 (.01)
3/4	0	302 (45)	302 (45)	.43 (.06)	.26 (.06)	.30 (.04)
	.15	216 (51)	317 (50)	.40 (.07)	.13 (.06)	.20 (.04)
	.30	226 (67)	417 (81)	.30 (.07)	.14 (.06)	.18 (.04)

WHO criteria: ART failure if CD4 <200. FNR = .70, FPR = .10, and TMR = .26.

Summary

- Price of VL has dropped (from ∼$100 to ∼$30 at AMPATH), but VL still a limited resource.
- Need methods for making most efficient use for monitoring ART.
- We formalized framework for using VL to augment clinical marker information.
- Substantial gains in diagnostic accuracy possible with small fraction of VL.
- Translation to RLS needed
  - Rule development requires VL to be available.
  - We are working on methods where VL only partially available.

Acknowledgment

- My colleagues:
  - Joseph Hogan, ScD (Brown)
  - Rami Kantor, MD (Brown)
  - Vivian Zhang, MS (Memorial Sloan-Kettering Cancer Center)
- Former student:
  - Lisa Wang (University of Tuebingen, German)
- Brown/Tuft/Lifespan CFAR (NIAID P30AI042853)

Journal of the American Statistical Association

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Optimal Allocation of Gold Standard Testing Under Constrained Availability: Application to Assessment of HIV Treatment Failure

Tao Liu [a] , Joseph W. Hogan [a] , Lisa Wang [a] , Shangxuan Zhang [b] & Rami Kantor [c]

[a] Department of Biostatistics, Center for Statistical Sciences , Brown University School of Public Health , Providence , RI , 02912