# The Analysis of Placement Values for Evaluating Discriminatory Measures

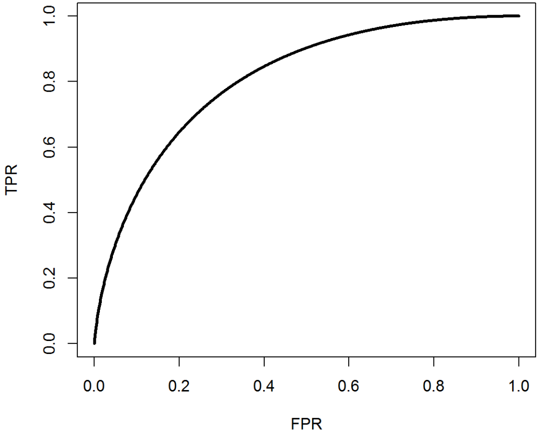
*Margaret Sullivan Pepe & Tianxi Cai*

Biometrics (2004)

Allison Meisner · April 22, 2014

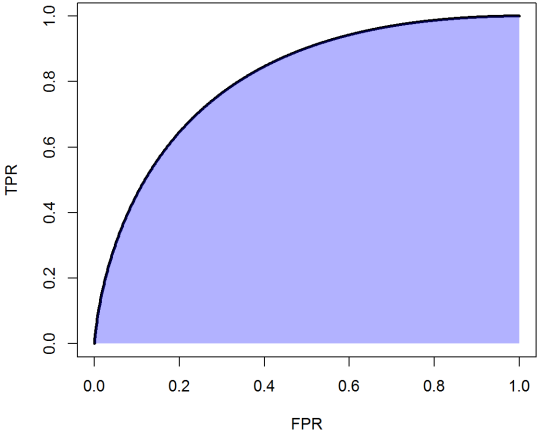
# Overview

This paper relates ROC curves and AUC to the distribution of placement values and uses these placement values to motivate a new approach for making inference about ROC curves and AUC.



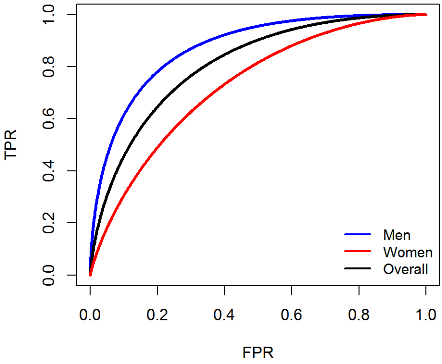
# Overview

This paper relates ROC curves and AUC to the distribution of placement values and uses these placement values to motivate a new approach for making inference about ROC curves and AUC.



# Scientific Motivation

- When building a risk prediction model, it may be the case that discriminative ability is affected by covariates.
- For example, we may believe that a given marker discriminates disease better in men than women. These differences are important to identify so that we know where to apply our risk prediction model.



# Statistical Motivation

As with Alonzo and Pepe (2002), the idea here is to fit the problem of covariate effects on discrimination into the regression framework.

- Alonzo and Pepe (2002) proposed an algorithm for fitting this regression model involving binary GLMs.
- After recognizing the connection between placement values (PVs) and the ROC curve (first identified by Hanley and Haijian-Tilaki (1997)), Pepe and Cai considered using PVs in a regression model.

# Enter Pepe & Cai...

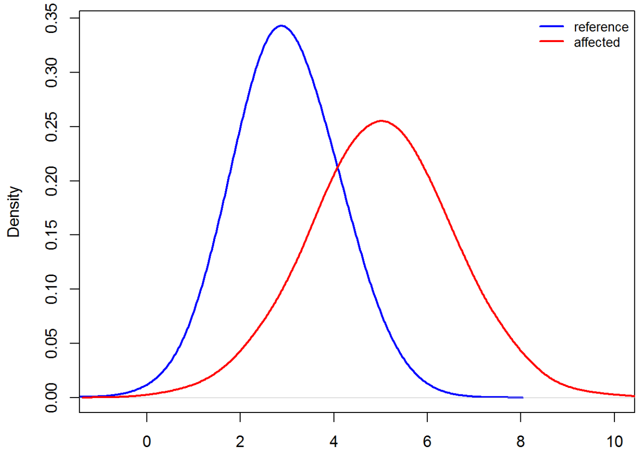
- Demonstrate the importance of placement values
  - For a continuous measure $Y$, the placement value of $Y$ is the proportion of the nondiseased (reference population) with values larger than $Y$:

$$U = 1 - F_{\overline{D}}(Y)$$

  - The distribution of placement values in the diseased (affected population) tells us about the discriminative ability of $Y$: if $U_D \sim \text{Uniform}(0, 1)$, then $Y$ has no discriminative ability.

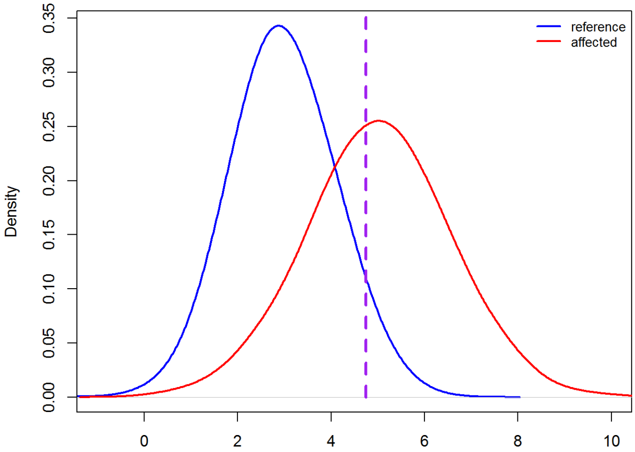
# Enter Pepe & Cai...

- Let's recall how these densities relate to **TPR**:



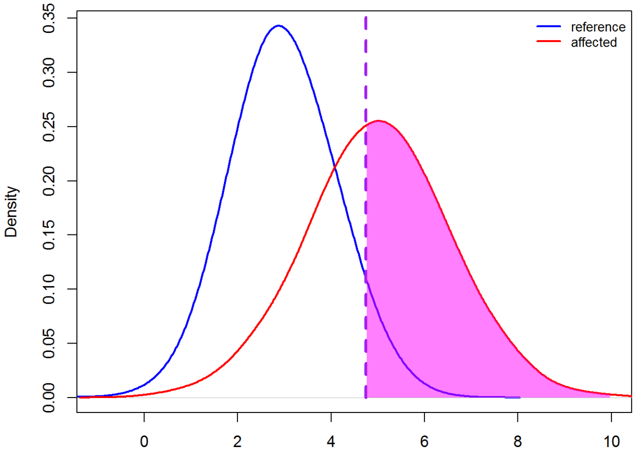
# Enter Pepe & Cai...

- Let's recall how these densities relate to **TPR**:



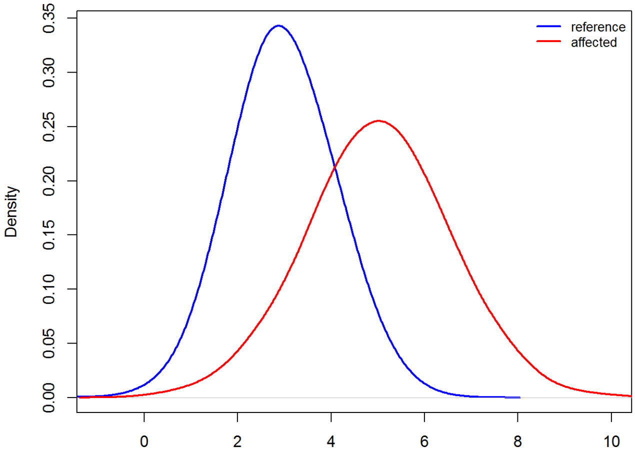
# Enter Pepe & Cai...

- Let's recall how these densities relate to **TPR**:



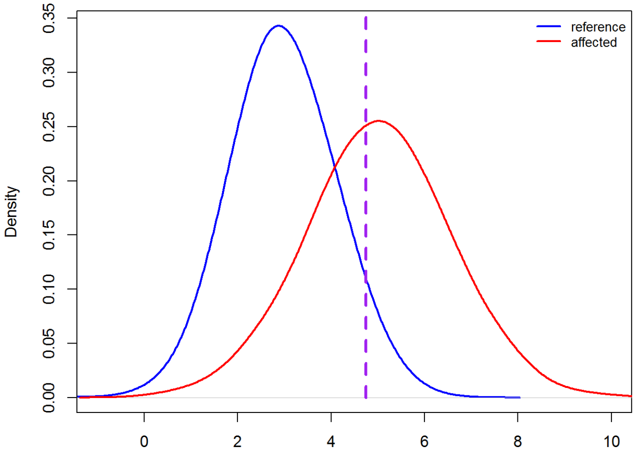
# Enter Pepe & Cai...

- And how these densities relate to **FPR**:



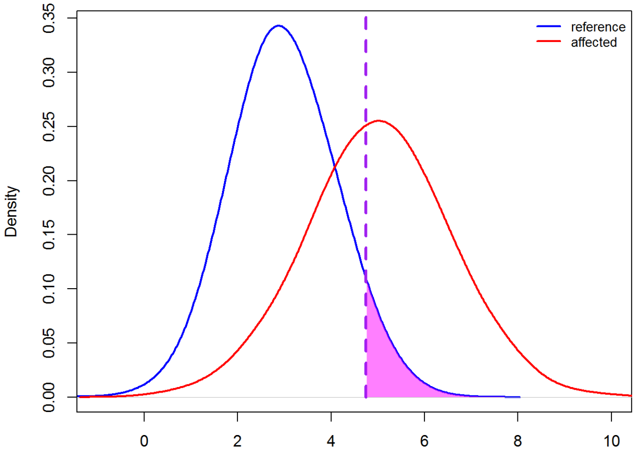
# Enter Pepe & Cai...

- And how these densities relate to **FPR**:



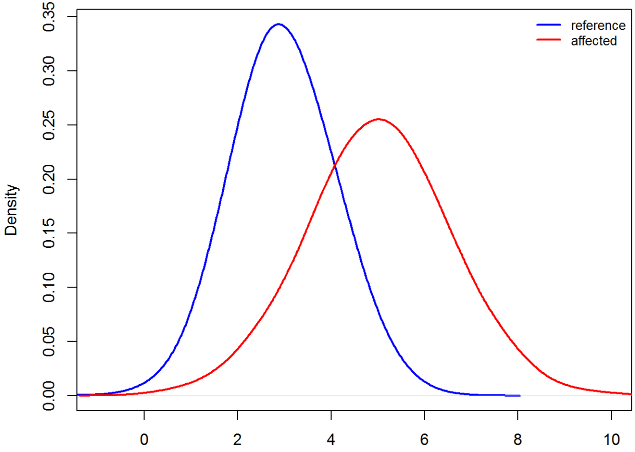
# Enter Pepe & Cai...

- And how these densities relate to **FPR**:



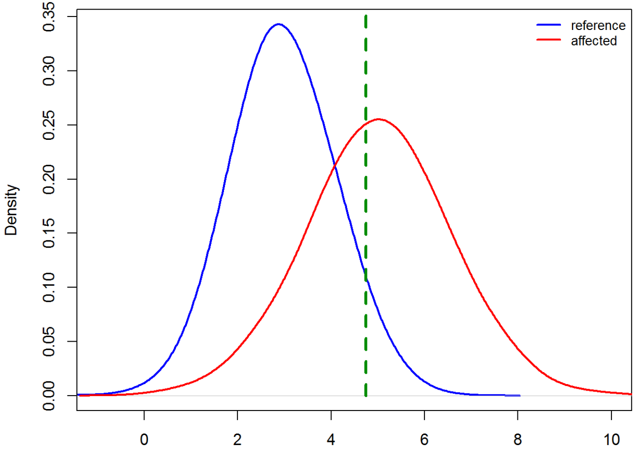
# Enter Pepe & Cai...

- Now, let's see how these densities relate to **PVs**:



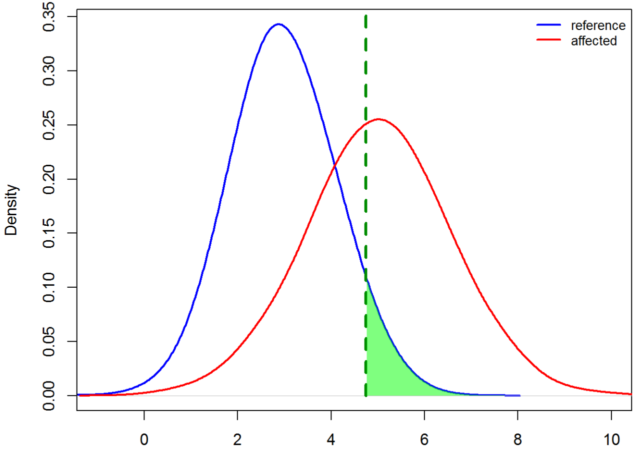
# Enter Pepe & Cai...

- Now, let's see how these densities relate to **PVs**:



# Enter Pepe & Cai...

- Now, let's see how these densities relate to **PVs**:



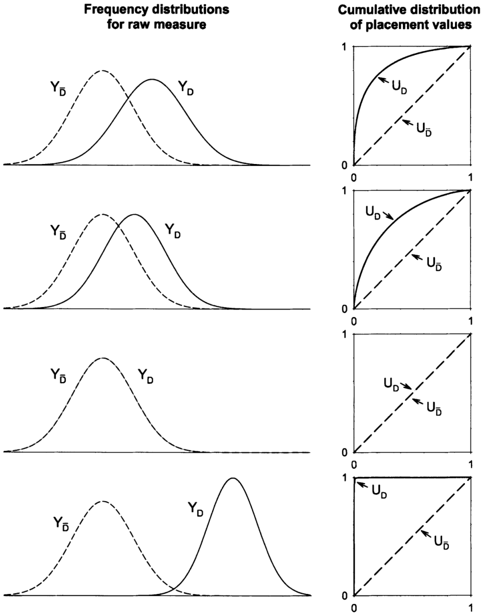
# Enter Pepe & Cai...

- Demonstrate the importance of placement values
  - We have

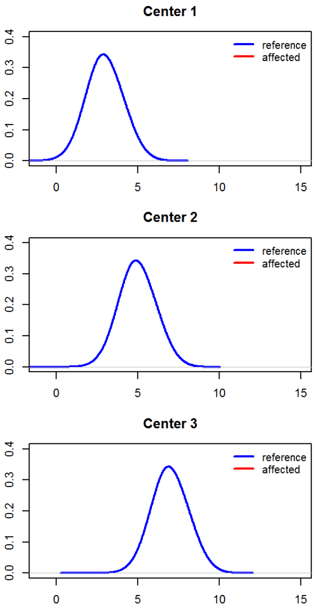
$$\begin{aligned} ROC(u) &= \text{TPR at FPR of } u \\ &= P(Y_D \geq F_{\overline{D}}^{-1}(1-u)) \\ &= P(1-u \leq F_{\overline{D}}(Y_D)) \\ &= P(1-F_{\overline{D}}(Y_D) \leq u) \\ &= P(U_D \leq u), \end{aligned}$$

    demonstrating the connection between the ROC and the CDF of the PVs.
  - Note also that since the expected value of a random variable is the area under its (1 – cdf), we have $AUC = E(1-U_D)$.

# Enter Pepe & Cai...



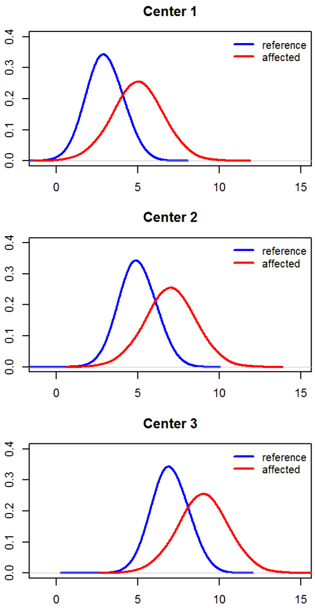
# Enter Pepe & Cai...




- Account for covariates
  - Covariate-specific PVs: covariates $\mathbf{Z}$ may affect the distribution of $Y$ in the reference population, which motivates covariate-specific PVs,

$$U = 1 - F_{\overline{D},\mathbf{Z}}(Y)$$

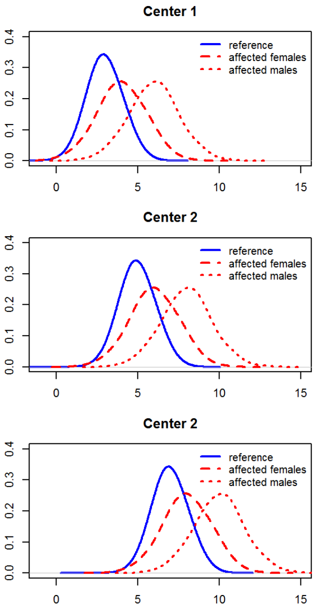
# Enter Pepe & Cai...




- Account for covariates
  - Covariate-specific PVs: covariates $\mathbf{Z}$ may affect the distribution of $Y$ in the reference population, which motivates covariate-specific PVs,

$$U = 1 - F_{\overline{D},\mathbf{Z}}(Y)$$

# Enter Pepe & Cai...




- Account for covariates
  - Covariate effects on discrimination: the distribution of $U$ may depend on covariates $\mathbf{Z}_D$ if discrimination is better in certain settings $\Rightarrow$ regression models for $U_D$ can quantify such covariate effects:

$$H_{\boldsymbol{\alpha}}(U_D) = -\boldsymbol{\beta}^T \mathbf{Z}_D + \epsilon, \ \ \epsilon \sim g(\cdot).$$

# Enter Pepe & Cai...

- Account for covariates
  - Connection with ROC regression: can show* that the model

$$H_{\boldsymbol{\alpha}}(U_D) = -\boldsymbol{\beta}^T \mathbf{Z}_D + \epsilon$$

    is equivalent to the class of ROC regression models expressed as (Pepe (1997))

$$ROC_{\mathbf{Z},\mathbf{Z}_D}(u) = g\{\boldsymbol{\beta}^T \mathbf{Z}_D + H_{\boldsymbol{\alpha}}(u)\}.$$

    This can be interpreted as the separation between subjects in the reference population with covariates $\mathbf{Z}$ to those in the affected population with covariates $(\mathbf{Z}, \mathbf{Z}_D)$.
  - Connection with AUC regression: can interpret AUC regression models as models for mean PVs

# Enter Pepe & Cai...

- **Proposal**
  - These results imply that **ROC analysis $\leftrightarrow$ analysis of PVs**.
  - Idea: leverage this finding to improve efficiency of parameter estimation.
    - This is accomplished by a pseudo-likelihood function and semi-parametric estimation of $F_{\overline{D},\mathbf{Z}}$ (details next time).
  - Additionally, a method for fitting AUC regression models based on GLM IRWLS is proposed based on the connection between AUC regression and mean PVs.

# Enter Pepe & Cai...

- Used simulations to
  - assess the properties of the pseudo-likelihood approach to ROC regression
  - compare the pseudo-likelihood approach to the binary regression approach of Alonzo and Pepe (2002) for ROC regression
- Analyzed FEV data using the proposed ROC and AUC regression methods

# Research Landscape

- Placement values have been commonly used in some areas of medicine for decades; for example, a child's height/weight is typically reported as a percentile relative to some healthy population.
- Previous papers recognized the connection between the ROC curve and PVs (Hanley and Haijian-Tilaki, 1997).
- ROC regression was proposed by Pepe (1997) and further discussed by Alonzo and Pepe (2002), who proposed fitting these models using binary GLM.
- AUC regression and similar methods were proposed by several groups, including Thompson and Zucchini (1989), Dorfman, Berbaum and Metz (1992) and Dodd and Pepe (2003).

# Impact of Pepe & Cai

Pepe & Cai accomplished several things:

1. Highlighted the relationship between PVs and ROC/AUC
2. Leveraged this relationship to develop a more efficient method for ROC regression
3. Used the connection between PVs and AUC to propose a method for AUC regression based on mean PVs