A Graphical Model Approach to Eyewitness Identification

Amanda Luby

Department of Statistics, Carnegie Mellon University

Advisor: Stephen Fienberg, CMU Statistics

September 29, 2016

Introduction

- Eyewitness Testimony versus Eyewitness Identification
- False identifications were a factor in 70% of DNA exonerations [1]
- Identified as a leading factor of wrongful convictions in 1932 [2]
- Active research area in psychology for the last three decades

[1] Innocence Project

[2] *Convicting the Innocent* - Edwin Borchard

Variables associated with Lineups

I. Simultaneous vs Sequential
II. Instructions
   - Biased - 'Choose the person you saw at the scene of the crime'
   - Unbiased -'The perpetrator may or may not be present in the lineup'
III. Retention Interval
   - How long between crime and lineup
IV. Lineup Size
V. Filler Similarity
VI. Type of crime
VII. Type of interaction with perpetrator
VIII. Lighting at time of crime
IX. Distance from scene of the crime
X. Race/Gender/Age differences

csafe
Center for Statistics and Applications in Forensic Evidence

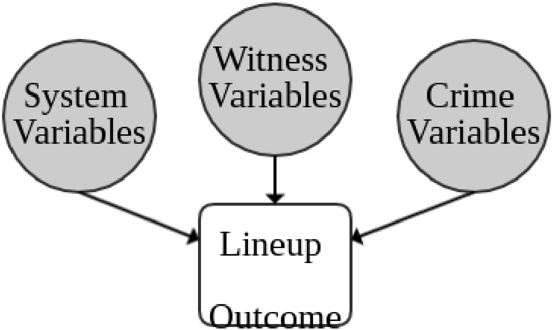
Things (almost) everybody agrees upon

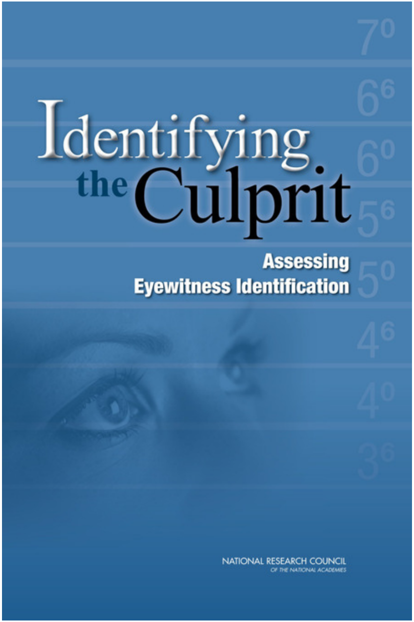
- Blind Administration
- Lineup Composition
- Instructions
- Confidence Statements
- Recording

What we're still arguing about

- Simultaneous versus Sequential presentation
- How to choose fillers?
- What's the role of non-system variables? (Race, gender, age, type of crime, lighting, etc...)

An Intuitive Model



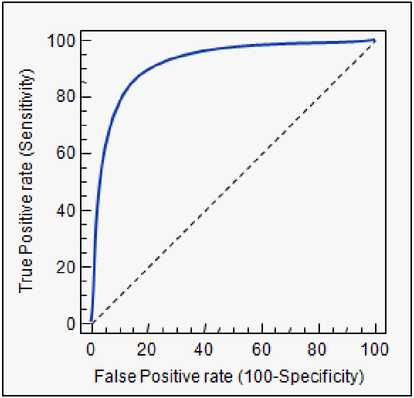


> "The committee recommends a broad exploration of the merits of different statistical tools for use in the evaluation of eyewitness performance"

How do we measure effects of lineup conditions?

Diagnosticity (likelihood) Ratio

$$\mathrm{DR} = \frac{P(\text{Suspect IDed}|\text{Guilty})}{P(\text{Suspect IDed}|\text{Innocent})}$$



Diagnosticity ratio may over-simplify

A Toy Example:

I. Lineup Procedure 1
   - 50 TA Lineups, 50 TP Lineups
   - 25 True positives
   - 15 False positives
   - DR $= \frac{25/50}{15/50} = 1.67$

II. Lineup Procedure 2
   - 50 TA Lineups, 50 TP Lineups
   - 6 True positives
   - 2 False positives
   - DR $= \frac{6/50}{2/50} = 3$

ROC analysis ignore inherent structure of lineups

I. Lineups as a (coerced) 2×2 Classification

	Make ID	Reject Lineup
Target Present	True Positive	False Negative
Target Absent	False Positive	True Negative

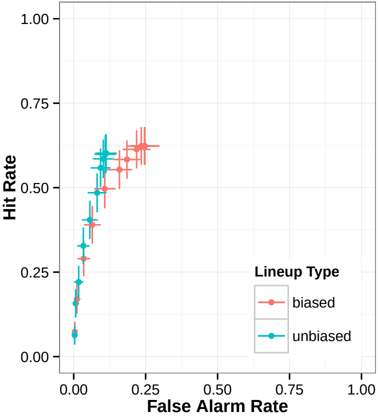
II. Lineups as a 2×3 Classification

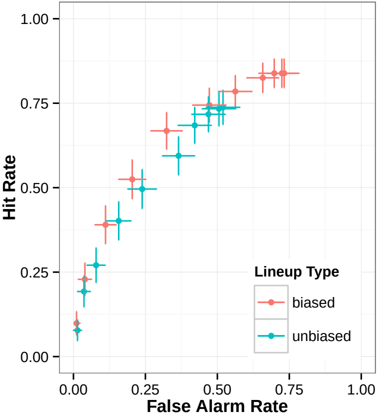
	ID Suspect	ID Filler	Reject Lineup
Target Present	True Positive	Filler ID (TP)	False Negatives
Target Absent	Innocent ID	Filler ID (TA)	True Negatives

Conclusions dependent on ambiguous definitions

TP	F-ID (TP)	FN
I-ID	F-ID (TA)	TN

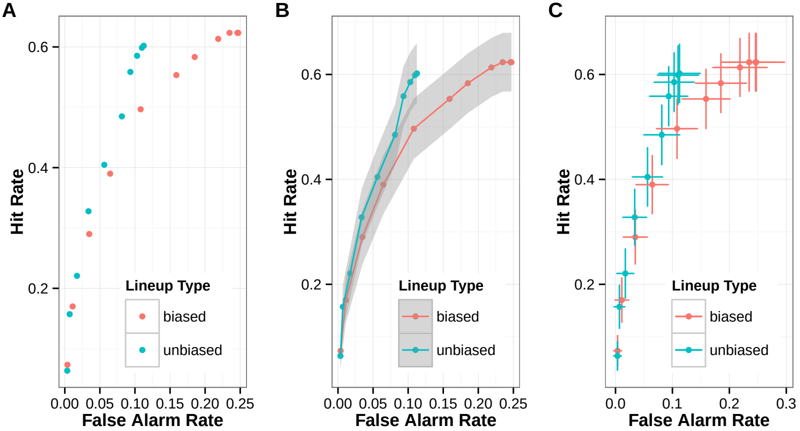
TP	**F-ID (TP)**	FN
I-ID	**F-ID (TA)**	TN



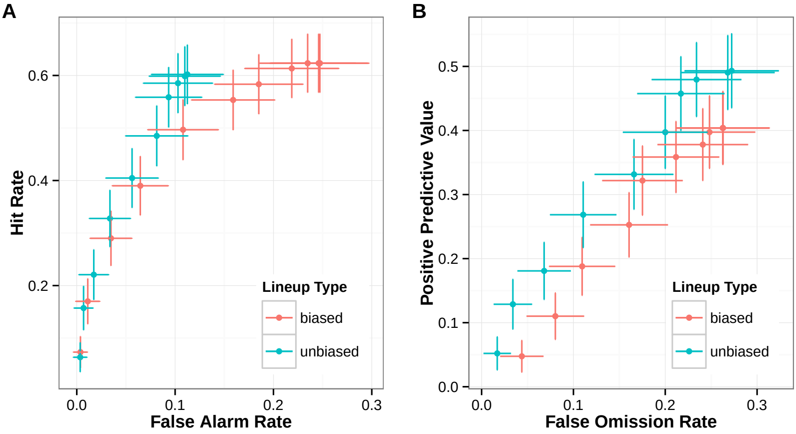


Uncertainty quantification in ROC analysis is unclear

Even optimistic assumptions lead to overlapping curves



Restricts comparisons to two quantities



Log Linear Analysis provides rigor and flexbility

I. Treat data as as $2 \times 3 \times 11 \times 2$ contingency table
   - TA/TP Lineup $\times$ Witness Choice $\times$ Confidence $\times$ Instructions
   - Can add more dimensions (variables) without changing theory

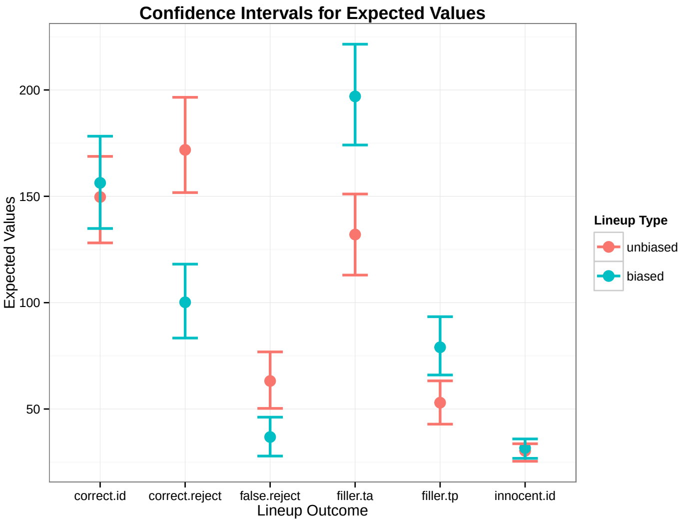
II. Hierarchical Model Search on training set(two-way interaction)
   i Unconditional edge exclusion test
   ii Conditional edge inclusion test

III. Use iterative proportional fitting algorithm to obtain expected values for each cell

IV. Evaluate model on testing set using goodness of fit tests
   - $G^2 = 2 \sum x_{ijkl} \log(\frac{x_{ijkl}}{\hat{m}_{ijkl}})$

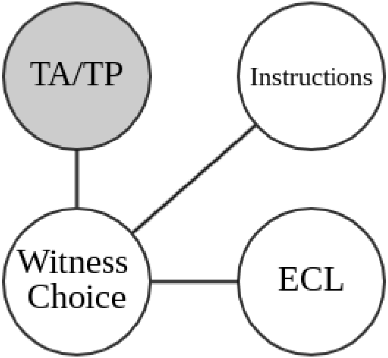
V. Make conclusions based on variables and edges included in the model and corresponding parameters

Expected values from the model make intuitive sense



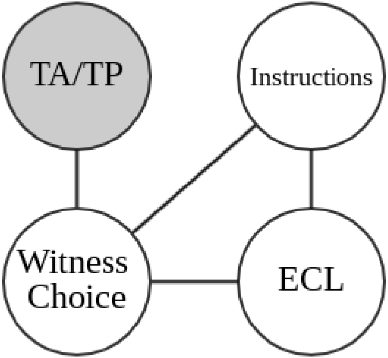
Resulting model and dependence graph

$$\log m_{\mathrm{wc,t,i,c}} = \alpha_{\mathrm{wc}} + \alpha_{\mathrm{t}} + \alpha_{\mathrm{i}} + \alpha_{\mathrm{c}} + \beta_{\mathrm{wc,t}} + \beta_{\mathrm{wc,i}} + \beta_{\mathrm{wc,c}}$$



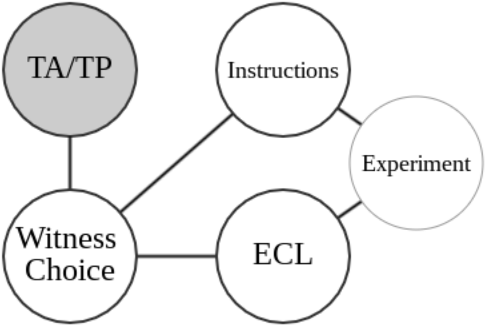
What about for a different experiment?

$$\log m_{\mathrm{wc,t,i,c}} = \alpha_{\mathrm{wc}} + \alpha_{\mathrm{t}} + \alpha_{\mathrm{i}} + \alpha_{\mathrm{c}} + \beta_{\mathrm{wc,t}} + \beta_{\mathrm{wc,i}} + \beta_{\mathrm{wc,c}} + \beta_{i,c}$$



We can reconcile with a single graph

$$\log m_{\text{wc,t,i,c}} = \alpha_{\text{wc}} + \alpha_{\text{t}} + \alpha_{\text{i}} + \alpha_{\text{c}} + \alpha_{\text{e}} + \beta_{\text{wc,t}} + \beta_{\text{wc,i}} + \beta_{\text{wc,c}} + \beta_{\text{e,i}} + \beta_{\text{c,i}}$$



Inconsistencies in Experimental Design

I. Ideal Setting

	ID Suspect	ID filler	Reject Lineup
Target Present	True Positive	Filler ID (TP)	False Negatives
Target Absent	Innocent ID	Filler ID (TA)	True Negatives

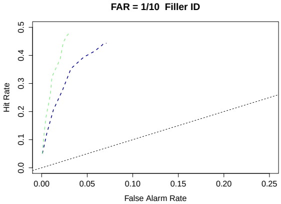
II. Commonly implemented in practice

	ID Suspect	ID Filler	Reject Lineup
TP	True Positive	Filler ID (TP)	False Negatives
TA	X	Filler ID (TA)	True Negatives

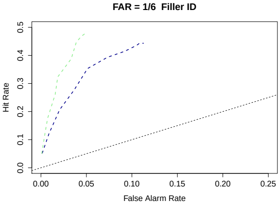
III. How this experimental design in analyzed

	ID Suspect	ID Filler	Reject Lineup
TP	True Positive	Filler ID (TP)	False Negatives
TA	$\frac{1}{6}$Filler ID (TA)	$\frac{5}{6}$Filler ID (TA)	True Negatives

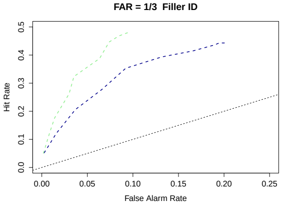
ROC Dependent on this inconsistency, L-L Model not



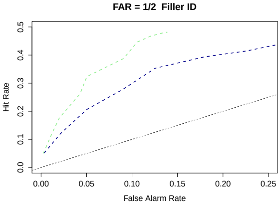
(a) FAR=$\frac{1}{10}$×Filler ID, $G^2 = 66.32$



(b) FAR=$\frac{1}{6}$×Filler ID, $G^2 = 64.38$



(c) FAR=$\frac{1}{3}$×Filler ID, $G^2 = 64.33$



(d) FAR=$\frac{1}{2}$×Filler ID, $G^2 = 60.55$

Model can handle variation in confidence

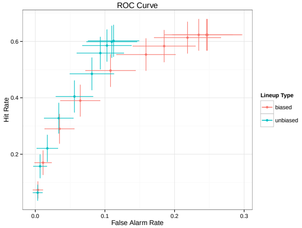


Figure: ROC Curve: Each point represents cumulative (FAR, HR) pair at a certain Expressed Confidence Level (ECL)

I. Assume confidence statements follow some probability distribution
II. Simulate new confidence statements for each observation
III. Plot new ROC curve
IV. Fit original log-linear model and calculate G^2
   - $G^2 = 2\sum x_{ijkl}\log(\frac{x_{ijkl}}{\hat{m}_{ijkl}})$

If ECL has small variance

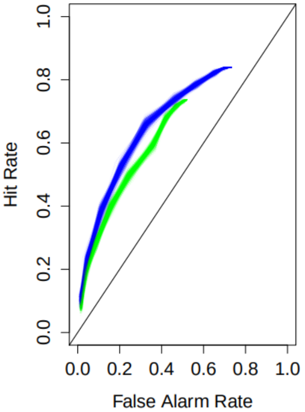


Figure: P(C-10) = .10, P(C) = .80, P(C+10)=.10
Out of 1000 simulations, the model never failed to fit.

If ECL has high(er) variance

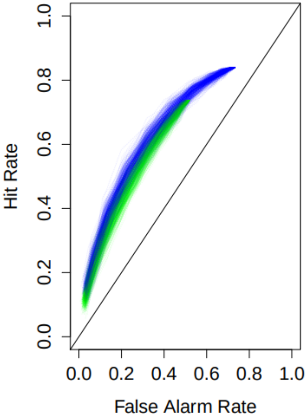
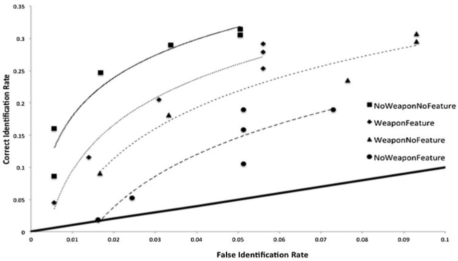
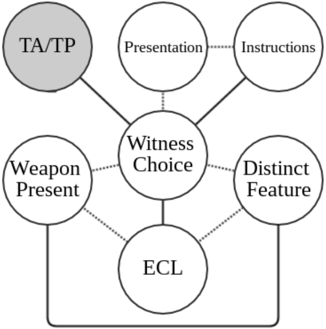


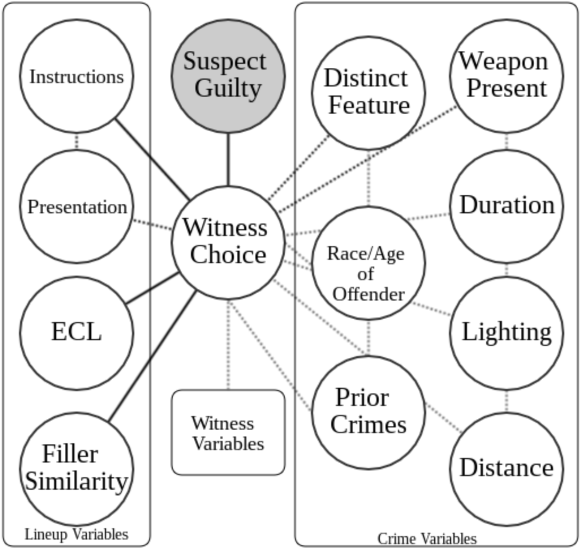
Figure: P(-20, -10, 0 , +10, +20) = (.2,.2,.2,.2,.2)
Out of 1000 simulations, the model failed to fit twice.

As complexity increases, graph interpretation stays the same

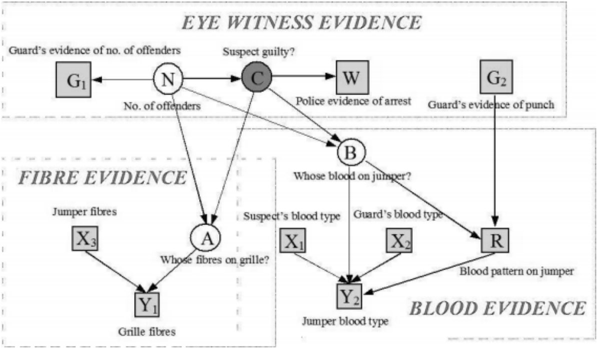




What about the other variables?



How could this fit in with other forensic analysis

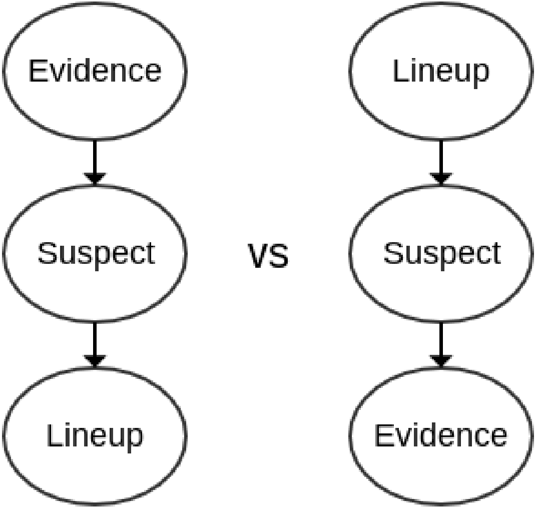


[3]

[3] Mortera, Dawid

Investigative tool versus confirmation evidence

Example: Some crime with a witness



Current Direction - Experimental Design

With Cleotilde Gonzalez and Stephen Fienberg

- Support from Arnold Foundation
- Fully documented
  - Which fillers are used?
  - How similar are the fillers?
- Reproducible
  - FERET Database
  - Experiment Materials
- Interactions-focused
- Repeated Measurements

Current Direction - Statistical Methodology

With Stephen Fienberg and Anjali Mazumder

- Combining with other evidence
- Bayesian methods
- Hierarchical Modeling
- Open-source tool
  - Multiple analysis methods
  - Data visualization

Conclusions

I. Our Contributions
- Evaluation of current statistical analyses
- Log-Linear framework for lineups
- First steps of experimental design

II. Next Steps
- Implementation
- Dissemination