

Blinding Project

Robin Mejia, CMU
July 18, 2019



www.forensicstats.org

Blinding has a rich history



But it goes back even further

- 1784 – French Academy of Sciences investigates claims of animal magnetism (Franklin and Laviosier)
- 1817 – Comparison of the sound of a Stradivarius violin to a new design.
- 1907 – Double-blind study of caffeine (Rivers and Webber)
- As these reports suggest, different parts can of a process can be blinded.

So many ways to blind

- Subjects of a study can be blinded to brands or other attributes of products when choosing their preference (wine taste tests, etc)
- Subjects of a study can be blind to whether they are receiving a treatment or placebo.
- Researchers can be blind to whether they are administering a treatment or placebo.
- Employees can be blind to whether their performance is being evaluated.

Similar types of blinding can (but often don't) happen in forensics labs



Why and when to blind?

- Forensic exams can be framed as a scientific study:
 - How similar are two items?
 - Is a known drug present in a sample?
 - Examiners can be blinded to extraneous (task-irrelevant) information
- Verification
 - When there is no ground truth, verification can be blind
- **Proficiency testing vs casework***
 - Do analysts perform differently when they know they're being tested vs not? Analysts can be blinded to whether a case is real or a test.

*Our focus today

Are tests different than work?



Research agrees

- A 1973 study of toxicology labs compared performance on open vs blind tests. Both false positives and false negative rates were higher in the blind tests. (Lamotte et al, 1977)
- Survey of supervisors at 156 PA hospitals showed a high prevalence of special practices in proficiency tests compared to lab work. (Cembrowski & Vanderline, 1988)
- Comparison of blind and open proficiency tests on blood lead levels concluded approximately 60% of clinical labs appeared to make special efforts on open tests. (Parsons et al, 2001)
- In forensics, some labs have implemented blind proficiency testing, but it is still rare (Koehler, 2013)

There have been calls in forensics

- In a 1992 report on forensic DNA testing in the United States, the National Academy of Sciences urged laboratories routinely engage in blind proficiency testing. (As did others)
- In 2016, The National Commission on Forensic Science Recommendation to the Attorney General (2016) recommends that the DOJ “require all DOJ FSSPs [forensic science service providers] to seek proficiency testing programs that provide sufficiently rigorous samples that are representative of the challenges of forensic casework.

<http://dels.nas.edu/Report/Technology-Forensic-Science/1866>

<https://www.justice.gov/archives/ncfs/page/file/905566/download>

What does blind forensic proficiency testing look like?

- We are discussing a type of blinding where:
 - Tests mimic actual casework.
 - Analysts do not know whether they are analyzing a real case or a blind test.
- This means that tests assess the entire case processing pipeline from submission to conclusion.
 - Evaluates entire quality management system, not just examiner performance on one piece of the process.
- *Blind proficiency testing can be implemented even when blinding has not been implemented in case work.*

What blinding is not

- An error rate study under controlled conditions.
 - Most techniques are likely to have some underlying uncertainty as well as a possibility for human variation and/or error.
 - There have been recent calls for more error rate studies, especially for subjective techniques (PCAST 2016)
 - Proficiency tests are not designed to assess the underlying validity of a technique or provide a formal error rate, though they may provide indications of support or areas to investigate.
 - A fully blind test may turn up issues in evidence handling, etc, not related specifically to analyst performance.

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

November 1-2, 2018

- Workshop hosted at Allegheny County Office of the Medical Examiner

Attendees	
Laboratory Directors	4
Quality Managers	8
Other Forensic Professionals	2
Professors	9
Graduate Students	2
Postdoctoral Fellow	1

- Labs represented ranged from <50 to >200 employees, including multi-lab systems

November 1-2, 2018

- Biggest take-away: Laboratories want to (and are starting to) implement blind proficiency testing, and they want support to make it happen
- Houston is a decade ahead of the curve. Lots to learn from the Houston experience, but it's not directly replicable.
- There are significant logistical, technical, and cultural obstacles to overcome.



Logistics

- Creating realistic samples
 - *Most labs do not have blinding in casework, so to blind a proficiency test, they must build an entire case. (more on this coming)*
- Paying for them
 - Will augment existing open testing, not replace it. Existing proficiency tests meet accreditation requirements
 - Houston estimated blind testing program costs about as much as open testing.
- Submission
 - Requires an outside partner
- Reporting
 - Local level – not releasing the case
 - Dealing with CODIS, AFIS, etc.
 - Reporting at a community level (will come back to this)

Technical Issues



STAR  LIMS
FORENSICS LIMS

The Crime Fighter BEAST



Integrated Evidence Tracking Systems for Police Departments and Crime Laboratories.

Laboratory Information Systems (LIMS)

- **“None of them can handle sequential unmasking”**
- “The thing you learn as a lab director is that the LIMS is more powerful than you are”
- “They all suck”
- “The Porter Lee table structure is SO BAD. There’s 2 developers who understand the table structure”
- Houston selected JusticeTrax 5 because it was the only one that could create a field called “proficiency” and blind it to everyone but quality.

Survey

- Just launched survey of LIMS capabilities and information practices. The first 14 responses have come in and...
 - 43% said that if an analyst has access to a case they can see the entire case file.
 - 29% said they could limit the information the analyst sees.
 - 79% said analysts confer on cases across disciplines.
 - *64% said analysts have access to the police/prosecution theory of a case through the documentation submitted. 50% said analysts confer with law enforcement.*
- If you want to weigh in:
<http://www.stat.cmu.edu/limesurvey/index.php/156378>

LIMS – take away

- In order to move to blinding in casework, it must be possible to limit which information analysts can access.
- If casework is not blinded, QA must create an entire fictitious case to implement blind proficiency tests, and work with outside partners.
- (There's a lot more to the survey, though. Watch this space.)

Samples

- Toxicology samples can be obtained from places like RTI.
 - Joint orders could lower costs
- For many types of evidence, HFSC shops at Walmart.
- Details matter
 - The handwriting's too neat. There's no way a cop submitted that.
 - No one would hold a crow bar that way (orientation of prints)
 - We never see that drug in that neighborhood; that officer only does big busts, etc...
 - Even tox samples will need labeling and submission

Who needs to know

- Most labs will need to work with a partner agency to submit.
 - Someone has to know the case isn't real, but should be a very limited circle
- CODIS/AFIS/etc
 - Will need someone to know internally so that case isn't uploaded.
- Ensuring reports aren't sent out
 - Should someone at police/DA's office to know so that if something does go out?
- Analysts
 - Should they know blind testing is happening?

Other issues

- What metrics do you want to review
 - Correct answer at the end of the process?
 - Correct answer by the first analyst to see a sample?
 - How does verification work?
 - Other issues to test? Turn around? Success of policies on communication, etc?
- Related: ensuring test cases are flagged so are not included in performance metrics?

Culture and History

- Most laboratories operate within a law enforcement chain of command and/or have close relationships to law enforcement.
 - Manly labs report into a police agency.
- Forensic laboratories operate in an adversarial context
 - “Results are discoverable. There’s nothing like an airline simulator in our community... We’ve had almost the opposite: Oops, we’ve gotta take Jeff out back and shoot him.”
- Consensus that implementing blind testing requires active support from senior management.
 - Culture change is possible. HFSC now publishes its results on its website.

Scale

- Implementing a significant level of blind testing require time, expertise, and money
- Most labs are smaller than Houston and do not have the resources to build a program from scratch.
 - “You have a quality team. I have a quality me.”
- 2009 Census of Publicly Funded Crime Labs: 411 laboratories, 397 respondents
 - 12% had fewer than 10 employees
 - 34% had fewer than 50 employees
- In our survey, labs reported between 3 and 180 examiners.

The Good News

- Forensic laboratory directors and quality managers are incredibly resourceful and dedicated to their jobs.
- We know of at least three labs in addition to Houston that are implementing blind testing programs.
- At the meeting, attendees expressed a strong desire to collaborate. Want to sharing of both SOPs and actual blind tests.
- Ways CSAFE can help:
 - Inter-lab studies and aggregation of result (with care not to claim this is a substitute for error rate studies)
 - Repository for best practices, trading of actual blind tests, etc.
 - Publications and norm-setting

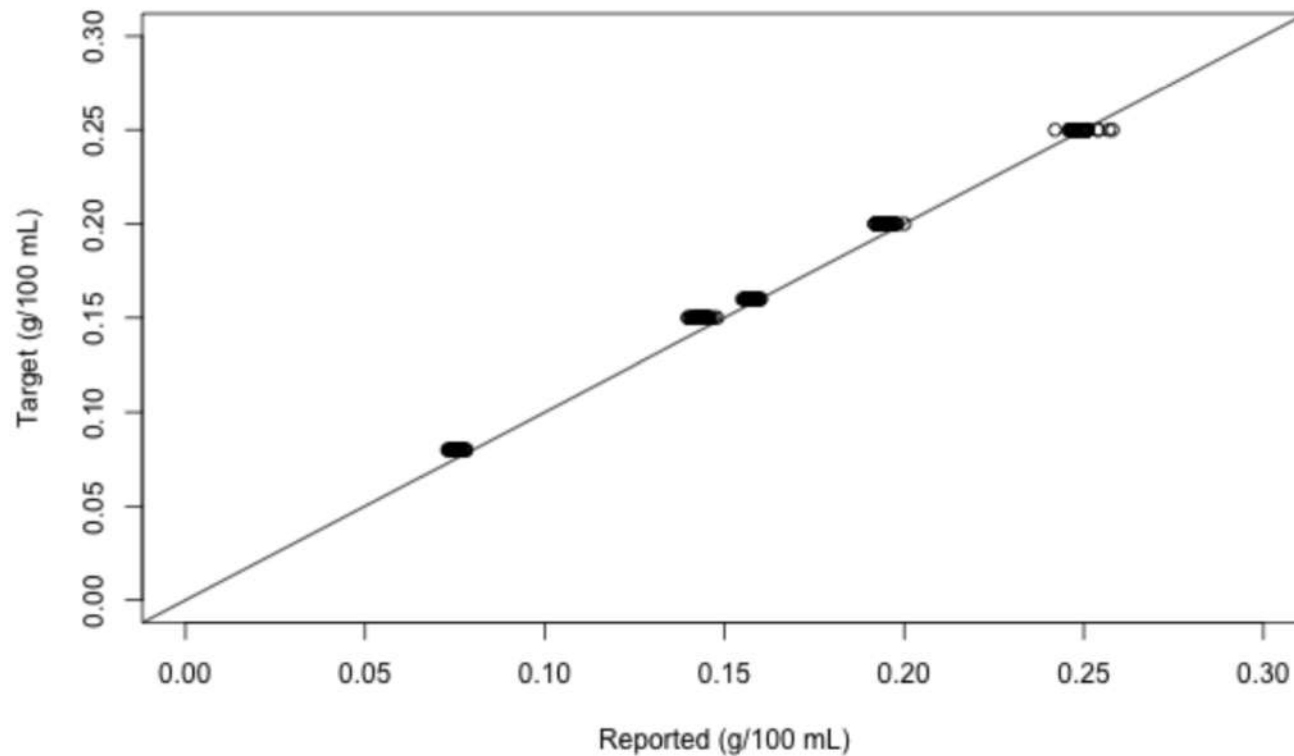
Houston Forensic Science Center

Type	Blind tests per month
Toxicology	14
Seized Drugs	30
Biology	4
Firearm BQC	1
Firearms blind verification	1
Latent Print Processing	3
Latent Print Comparison	10
Digital Forensics	1
Forensic Multimedia	1

More than 800 total blind tests have been completed to date.

Reported vs Manufacturer Spec (HFSC)

Figure 1. Target vs. reported blood alcohol concentrations



The Good News (quote version)

“Give people the tools they need to get those results and get that data into a central repository. We need to make sure that the data that we provide has value. We don’t have time for anything that has no value in the long run to help the community. This is not an industry that hoards information. We share processes and information. This is a very unique community in that we have not problems sharing what we do in order to help the community. Everybody has the same type of problems and time is one of them.”



Thank You!

- Robin Mejia: rmejia@andrew.cmu.edu
- Maria Cuellar: mcuellar@sas.upenn.edu
- Sharon Kelley (case processing/verification): smk8n@virginia.edu
- **The survey:**
<http://www.stat.cmu.edu/limesurvey/index.php/156378>