Measured Error Rates in Cartridge Case Comparisons

Dr. David P. Baldwin
Director, Midwest Forensics Resource Center

NIST Forensic Science Error Symposium
July 23, 2015

THE Ames Laboratory
Creating Materials & Energy Solutions
U.S. DEPARTMENT OF ENERGY

MFRC

Disclaimer

- This work was supported by Defense Biometrics and Forensics Office through the U.S. Department of Energy under Contract No. DE-AC02-07CH11358. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Biometrics and Forensics Office, Defense Forensic Science Center, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Acknowledgements

- Participants and their agencies
- This work was supported by Defense Forensic Office through the U.S. Department of Energy under Contract No. DE-AC02-07CH11358
- NIST for the opportunity to present at this conference
- Department of Energy , Office of Basic Energy Sciences, Science Undergraduate Laboratory Internship program for supporting 3D data acquisition
- Dr. Stanley Bajic, Professor Max Morris, Daniel Zamzow, Melinda Schlosser, and Stacie Johnson, Mackenzie Beyer, Ryan Spotts
- Defense Forensic and Biometric Agency including Dr. Jeffrey Salyards, Rick Tontarski, Henry Maynard, Garold Warner, and the staff of the firearms section at USACIL.
- Forensic Initiative at WVU provided the firearms, ammunition, and support in firing and collecting samples for the study including Dr. Keith Morris, Mike Bell, and several student staff members.
- Forensic Research Committee of ASCLD (Jay Henry)
- AFTE provided the largest group of participants (Jay Stuart).
- Story County Iowa Sheriff's office made the pilot study possible, particularly Sheriff Fitzgerald, Sergeant Backous, and Detective Rhoads.
- SWGGUN Chair Andy Smith provided valuable guidance in designing the study and aided in identifying the proper selection of ammunition.
- Firearms & Toolmarks Unit at the FBI Laboratory provided input on design
- Ames Laboratory Shipping and Receiving Department for coordinating and handling shipment of all of the materials for the study, and in particular Vicki Sieve.

Goals of the Study

- Measure false positive and false negative error rates by practicing firearms examiners for comparisons of fired cartridge cases
- Determine uncertainties in the measured rates

Important Design Criteria

- Sets must incorporate multiple independent comparisons
- Multiple groups of examiners must be examining independent sets of samples (to obtain a measure of uncertainty)
- Measure examiner rates, not agency rates (no review)
- Use accepted standard range of conclusions
  - AFTE range of conclusions
- Incorporate a measure of sample quality
- Simulate realistic sample presentation

Experimental Design

- Sets of 3 Knowns + 1 Questioned
  - Mimics a questioned case and a handgun in evidence with multiple test firings
- 15 Sets provided to each participant
  - No overlap or repeats between sets (avoid biasing effects of repeats)
  - No comparisons between sets (15 independent comparisons)
- Asked each participant to look at knowns first and identify how many were suitable for comparison
  - Internal measure of rate of good pattern production
- "Spoiler": each kit contained 5 same-source and 10 different-source sets (not announced)
- With 25 guns we randomly assigned each examiner to 1 of 5 groups
- Groups A through E (see Table)

Sample Set Design

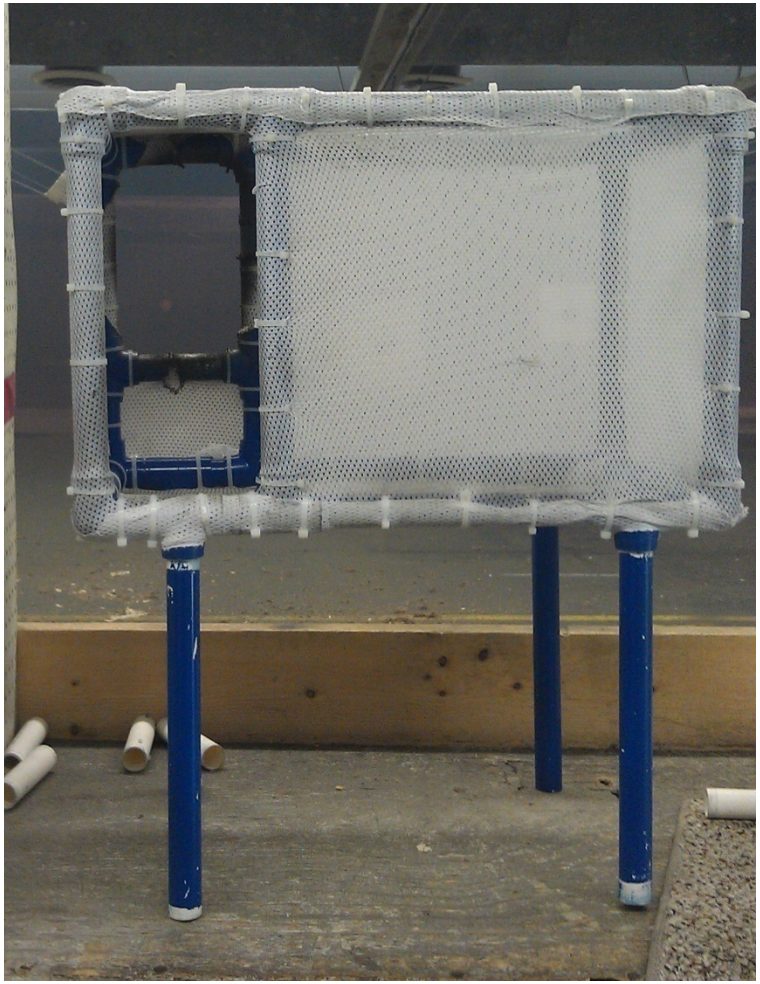
A	B	C	D	E
A1-A1	B1-B1	C1-C1	D1-D1	E1-E1
A2-A2	B2-B2	C2-C2	D2-D2	E2-E2
A3-A3	B3-B3	C3-C3	D3-D3	E3-E3
A4-A4	B4-B4	C4-C4	D4-D4	E4-E4
A5-A5	B5-B5	C5-C5	D5-D5	E5-E5
B v D: 1v2, 2v3, 3v4, 4v5, 5v1 and other skip permutations	C v E	D v A	E v B	A v C
C v E	D v A	E v B	A v C	B v D

Materials Used

- 25 new Ruger SR-9 semiautomatic 9-mm handguns
  - Moderate price, new model replaced P95
- 20,000 fired rounds of Remington L9MM3 FMJ
  - 2 lots
  - 3 days on the range
- Materials obtained and samples collected at WVU
- Each weapon fired 200 times before collection
- 800 rounds collected from each
- Order known to within 100 rounds (collected 100 from catcher at a time)

Ruger SR-9

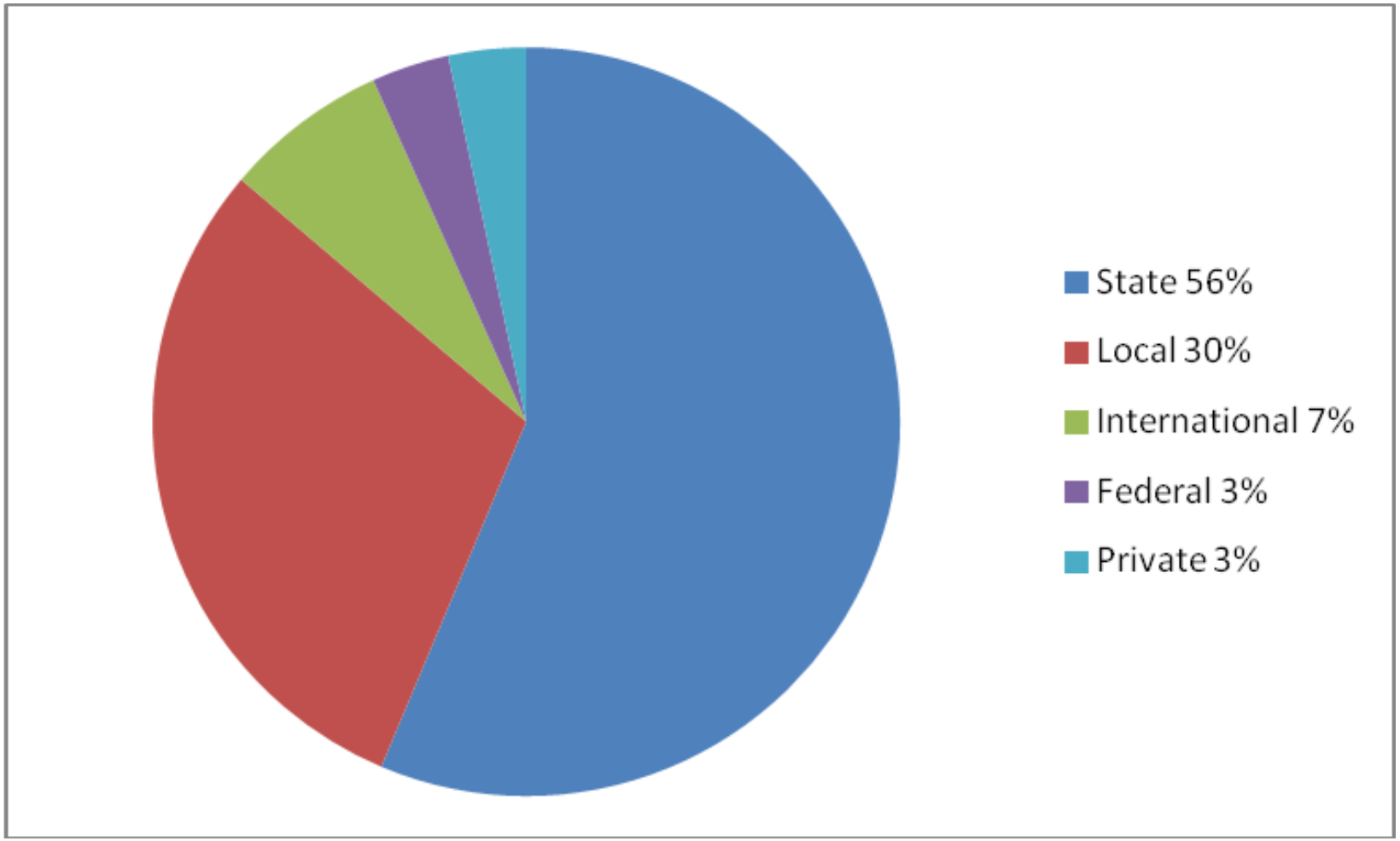
Brass Catcher



Participants

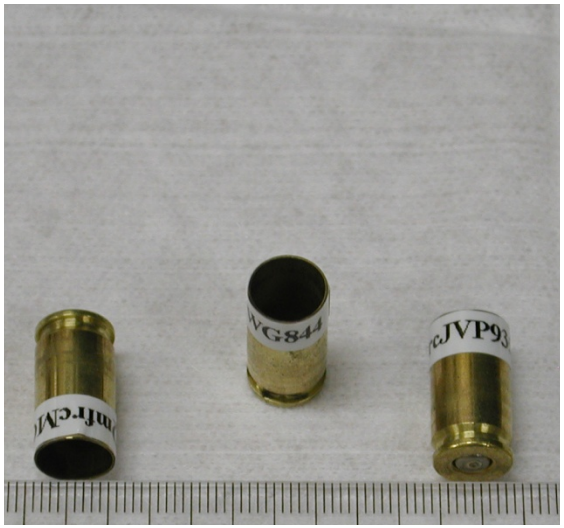
- Use of Human Subjects in federally funded project required review of design by Institutional Review Boards at Iowa State University and at DoD
- Mitigate risk to participants by making responses anonymous
- Informed Consent from Voluntary participants
- Solicited from AFTE membership and ASCLD participating agencies
- Active examiners only (low rates mean little confidence in rates for small numbers in any subgroups)
- Attempt to recruit 200 to 300
- 284 enrolled, 218 responses

Makeup of Participants



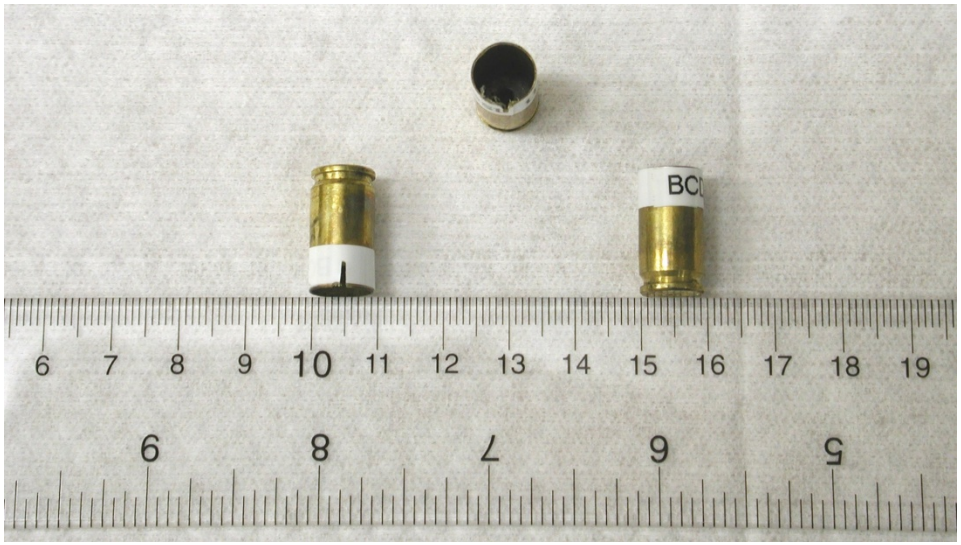
Labelling

- “Kmfrcxxxyyy” or “Qmfrcxxxyyy”
- Random alpha numeric coding
- Knowns and Questioned



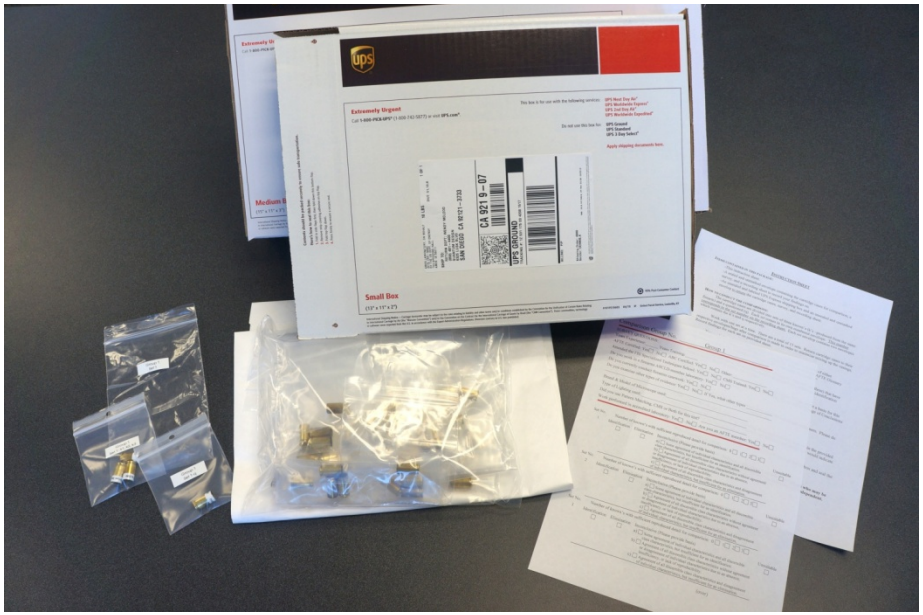
International Participants

- U.S. arms control regulations required damaging cases to prevent reloading
- Cut with a handheld rotary tool with a cutoff wheel



Packaging

- Packaged in 15 sets of 3 k + 1 q.
- Instructions
- Answer sheet
- Blank return envelope
- Prepaid return shipping package



Survey and Answers

Comparison Group No. _____

SURVEY QUESTIONS:

Years Experience:_____ Years Training:_____

AFTE Certified: Yes☐ No☐ ABC Certified: Yes☐ No☐ Other: _____

Attended the FBI Specialized Techniques School: Yes☐ No☐ CMS Trained: Yes☐ No☐

Do you work in a firearms ASCLD-member laboratory: Yes☐ No☐

Do you currently conduct firearms casework: Yes☐ No☐

Do you examine other types of evidence: Yes☐ No☐ If Yes, what other types _____

_____

Brand & Model of Microscope used:_____

Type of Lighting used:_____

Did you use Pattern Matching, CMS or Both for this test? _____

Work performed in accredited laboratory: Yes☐ No☐ Are you an AFTE member: Yes☐ No☐

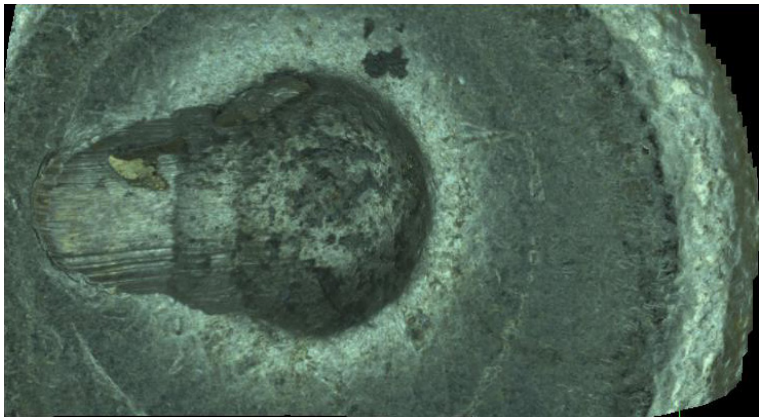
Set No. Number of known's with sufficient reproduced detail for comparison: 0☐ 1☐ 2☐ 3☐

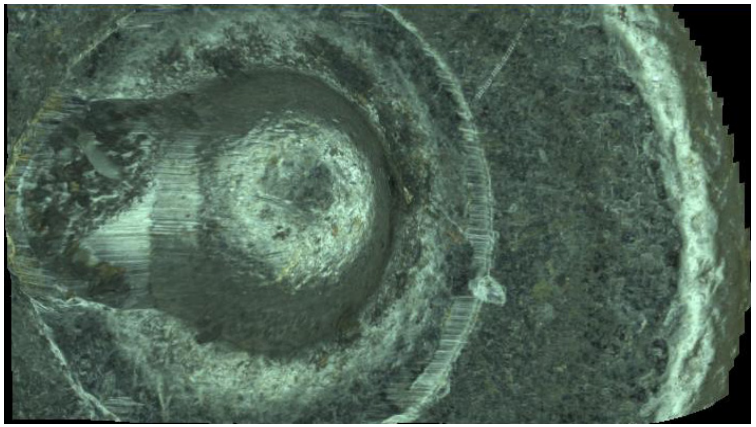
	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
1	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

By the Numbers

- Not everyone answered every question or supplied a response for every comparison
  - Non responsive answers not included in totals
- 5 (known same-source) x 218 (examiners) = 1090
- 10 (known different source) x 218 (examiners) = 2180 (but only 2178 responses)
- Suitability of knowns: 3 (knowns) x 15 (sets) x 218 (examiners) = 9,810 (but only 9702 responses)

Knowns with Insufficient Detail



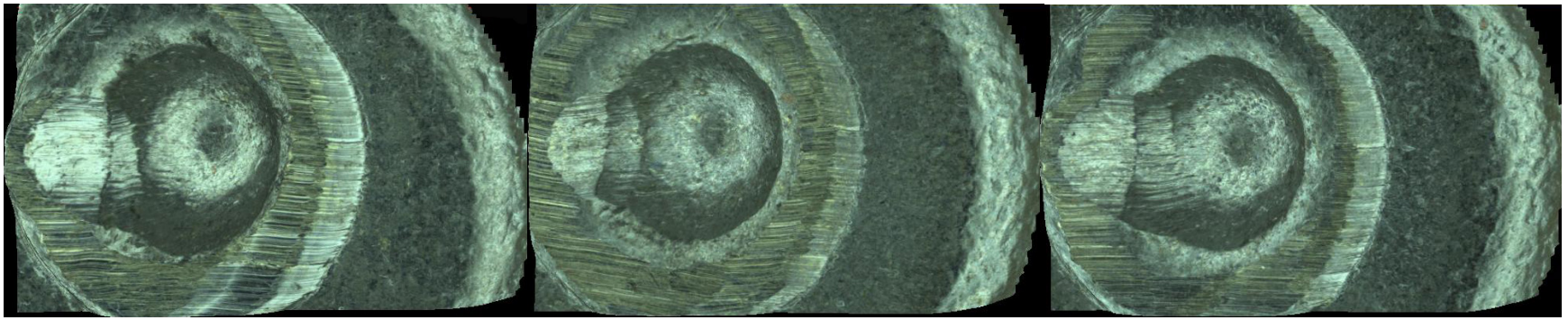


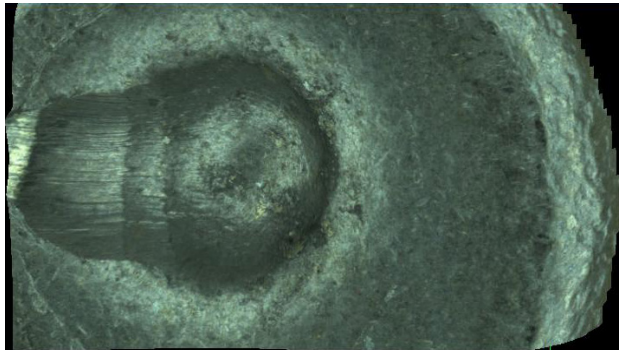
- Tend to have little or no breech-face marks reproduced or lots of obscuring deposits
- 2D projection of 3D image data from Alicona variable focus microscope (full surface shown "in focus" here)

Results for Known Same-Source Comparisons

- False negatives: 4/1090 = 0.3670%
  - 95% CI (Clopper-Pearson): 0.1001% to 0.9369%
- Include 11 Inconclusives (not errors): 15/1090 = 1.376%
  - 95% CI: 0.7722%, 2.260%
- Rate of unsuitable mark production: 225/9702 = 2.319%
  - 95% CI: 2.174% to 2.827%
- Conclusion: the rate of poor mark production may be entirely producing or obscuring the rate of examiner error (false-neg.)

False Negative



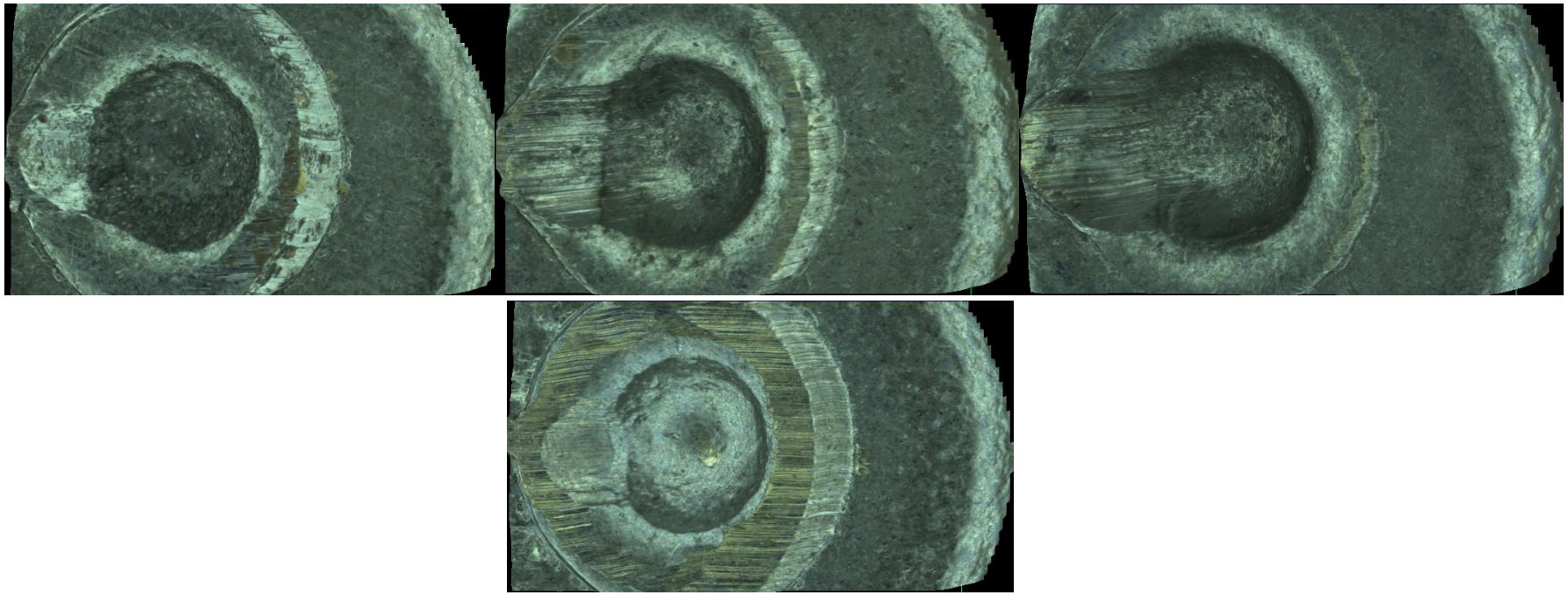


- Every false negative error was made by examiners who used only elimination and never used inconclusive (agency policy?)

Results for Known Different Source Comparisons

- Identifications from known different-source cases: 22/2178 = 1.010%
- However, 20 of 22 errors by 5 participants
- Indicates a highly heterogeneous distribution of error rates
- Statistical analysis based on this type of distribution of rates in a beta-binomial model
- Maximum Likelihood Estimator 0.939%
  - 95% CI: 0.360% to 2.261%
- Conclusion: false-positive error rates vary widely between different examiners

False Positive



- Some were difficult (like this), some should not have been too difficult
- Regardless, they are errors

Use of Inconclusive

- 96 examiners (44%) did not use Inconclusive (used Elimination for samples without sufficient corresponding detail for an identification)
- 45 (21%) used only Inconclusive to denote insufficient corresponding detail
- 77 (35%) used a mixture of inconclusive and elimination
- Given same model of ammunition and firearms throughout, what does inconclusive mean to this third group?

Policy Implications

1) False positives may be caught by peer review (not included in this study), but robust peer review and proficiency testing is required to systemically minimize risk of errors in reported results.

Policy Implications (cont'd)

2) Current policies regarding range of conclusions may be causing eliminations to be reported when insufficient detail exists for a conclusion – inconclusive would be a more appropriate response but is not available to some

Policy Implications (cont'd)

3) Use of Inconclusive:

- Only 11 true same-source comparisons reported as inconclusive – probably all with insufficient detail
- 735 different source reported as inconclusive (1421 reported as elimination) – statistically ~98% of these with sufficient detail for comparison
- These inconclusives don't reflect uncertainty in the comparisons, but policy that limits the use of elimination
- With sufficient detail present, a different conclusion term would be both more informative and not imply uncertainty in the examination

Proposed Future Work

- Given the relative size of false negative and poor mark reproduction rates: Study the variation in poor reproduction rates
  - Firearm model, between multiple guns of same model, with different make and material of cartridges, between and within lots, with age of firearm, etc.
  - Are there true false negatives and should QA systems be designed to catch them?
- Study effectiveness of QA systems in catching the types and rate of false positives seen
  - Include evaluation of possible confirmation bias in study
- How well would examiners perform with 3D data for the same sample set?
  - Develop tools for 3D manipulation and lighting

Thank You

- For your attention
- For your participation and support