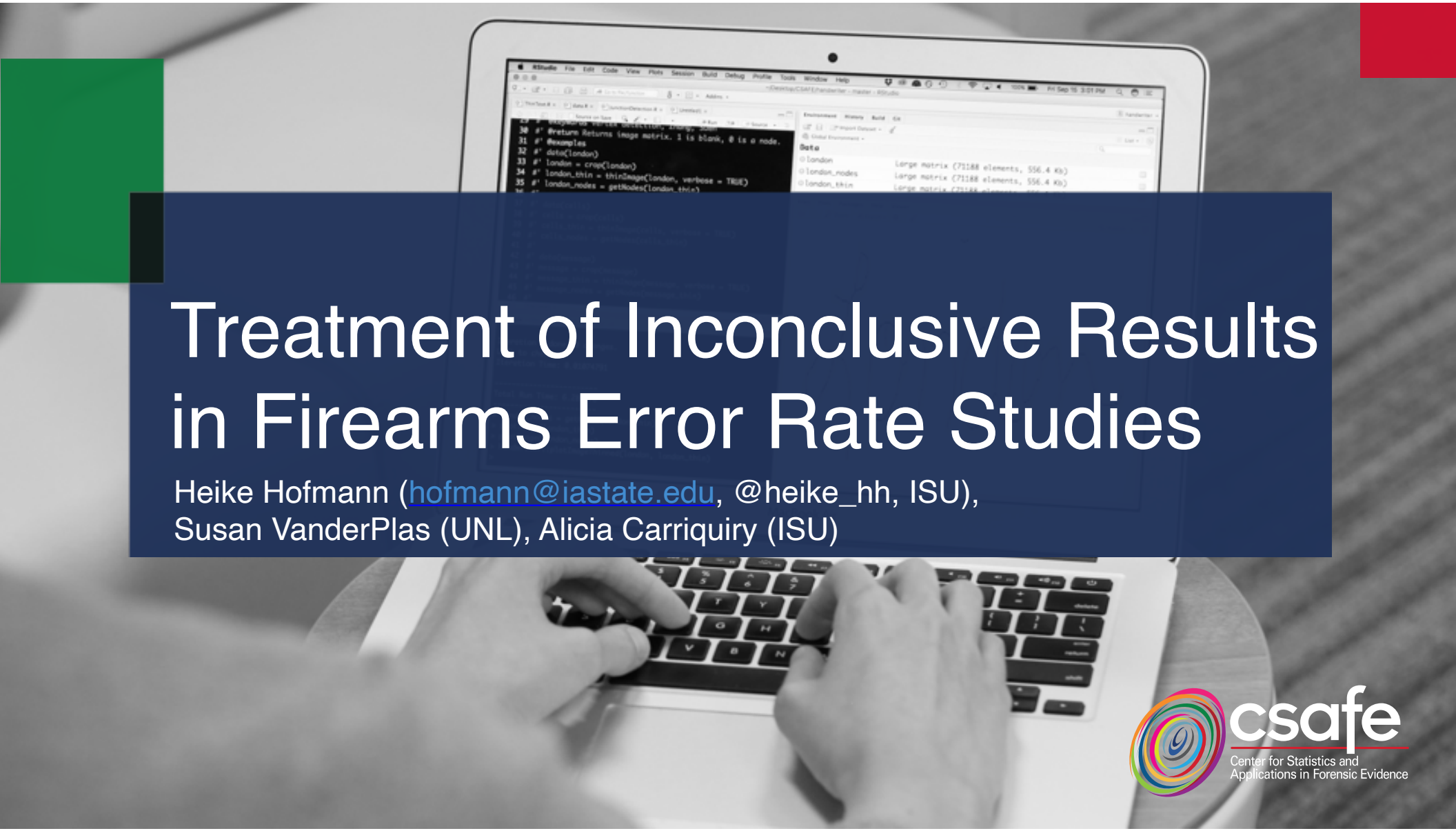


Treatment of Inconclusive Results in Firearms Error Rate Studies

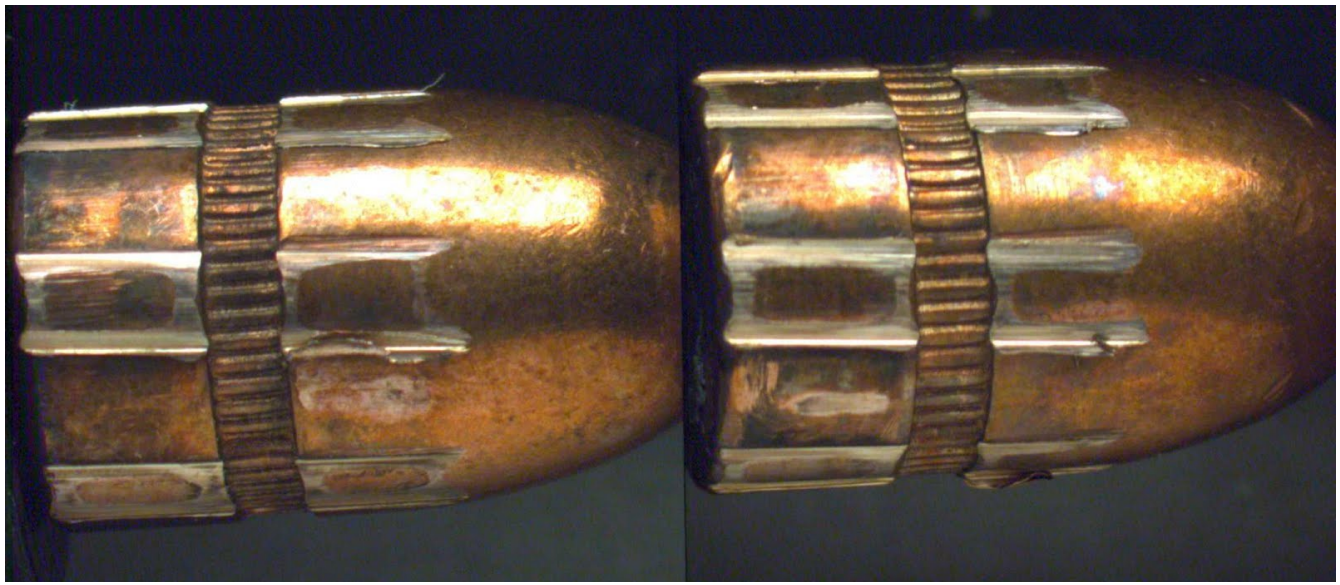
Heike Hofmann (hofmann@iastate.edu, @heike_hh, ISU),
Susan VanderPlas (UNL), Alicia Carriquiry (ISU)

csafe
Center for Statistics and Applications in Forensic Evidence

CSAFE webinar Feb 2021

Outline

- ★ Defining error rates for firearms evidence
- ★ Impact of inconclusive decisions on error rates
- ★ Process error and inconclusives



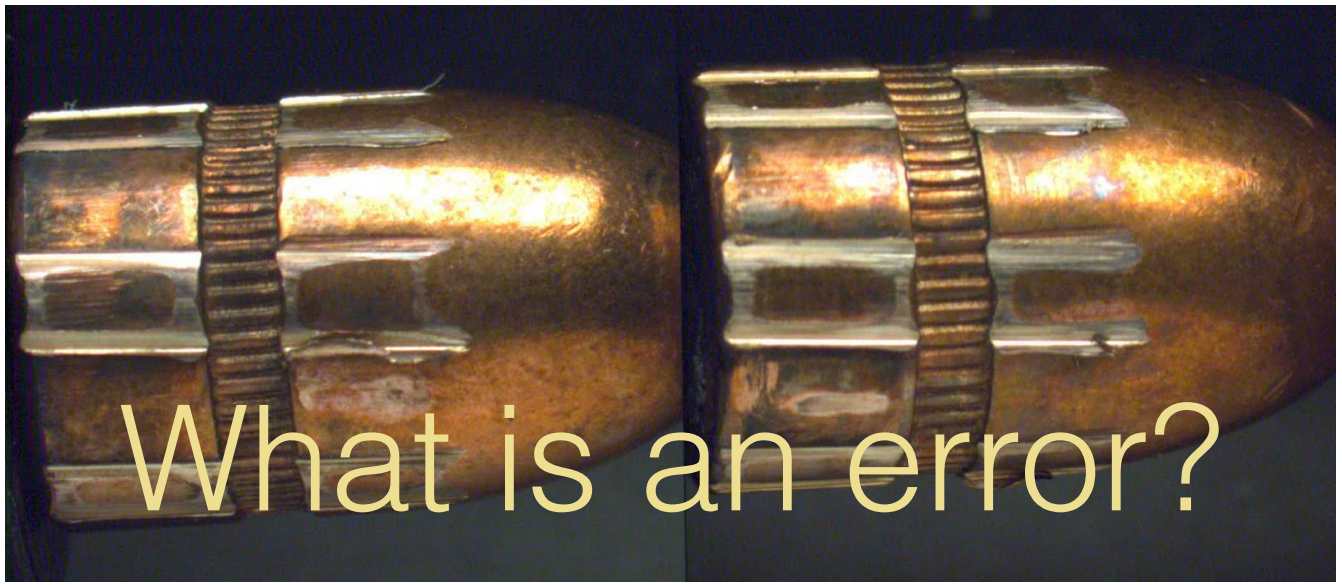
Over-arching Objective

★ **Same Source Problem**: do two pieces of firearms evidence come from the same source?

★ **Currently**: Firearms and Toolmarks Examiner use visual inspection under a comparison microscope: *subject bias, error rates?*

"much forensic evidence – including, for example, bite marks and firearm and toolmark identification is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing." (National Research Council 2009)

★ **Goals**: (1) determine *score* as objective measure for the match, (2) **establish error rates**



Quantifying Errors

★ Ground truth needed to establish error rates: need **case studies**
i.e. casework does not allow for assessing errors

★ Case studies:

- ★ premise: the participant (firearms examiner) does not know ground truth
- ★ premise: the participant should not be able to infer a conclusion based on anything but the comparison
- ★ Gold standard: (blind testing) the participant does not know they are being tested
  https://www.houstonforensicscience.org/event/5ae08c1brWanqy%202017.pdf
- ★ Reality: participant compares a number of questioned items to a number of reference items - conclusions according to AFTE Theory of identifications

AFTE Range of Conclusions

https://afte.org/about-us/what-is-afte/afte-range-of-conclusions

AFTE Criteria for Identification Committee. Theory of identification, range striae comparison reports and modified glossary definitions. AFTE Journal, 24(3):336–340, 1992.

★ **Identification**

Agreement of a combination of individual characteristics and all discernible class characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool.

★ **Inconclusive**

(a) Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.

(b) Agreement dual characteristics due to an absence,

Inconclusive results are a perfectly valid result of a comparison under AFTE rules

(c) Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

★ **Elimination**

Significant disagreement of discernible class characteristics and/or individual characteristics.

★ **Unsuitable**

Unsuitable for examination.

csafe
Center for Statistics and Applications in Forensic Evidence

What makes an Error?

Dror IE, Scurich N. (2020)
(Mis)use of scientific measurements in forensic science.
Forensic Sci Int. 6 (2), p. 333-338.

Errors according to AFTE rules

PARTICIPANT'S DECISION	EVIDENCE: Same-source	EVIDENCE: Different-source
Identification	Correct	Error
Exclusion	Error	Correct
Inconclusive	Correct	Correct

Inconclusives as errors

★ Treatment of inconclusive results hugely impacts error rates

★ AFTE rules measure an examiner's error

PARTICIPANT'S DECISION		
Identification	Correct	Error
Exclusion	Error	Correct
Inconclusive	Never considered as potentially incorrect	

★**Process error:** process does not result in identification or elimination

Dror & Scurich (2020) proposal:

PARTICIPANT'S DECISION	EVIDENCE: Same-source	EVIDENCE: Different-source
Identification	Correct	Error
Exclusion	Error	Correct
Inconclusive	Error	Error

Case Studies

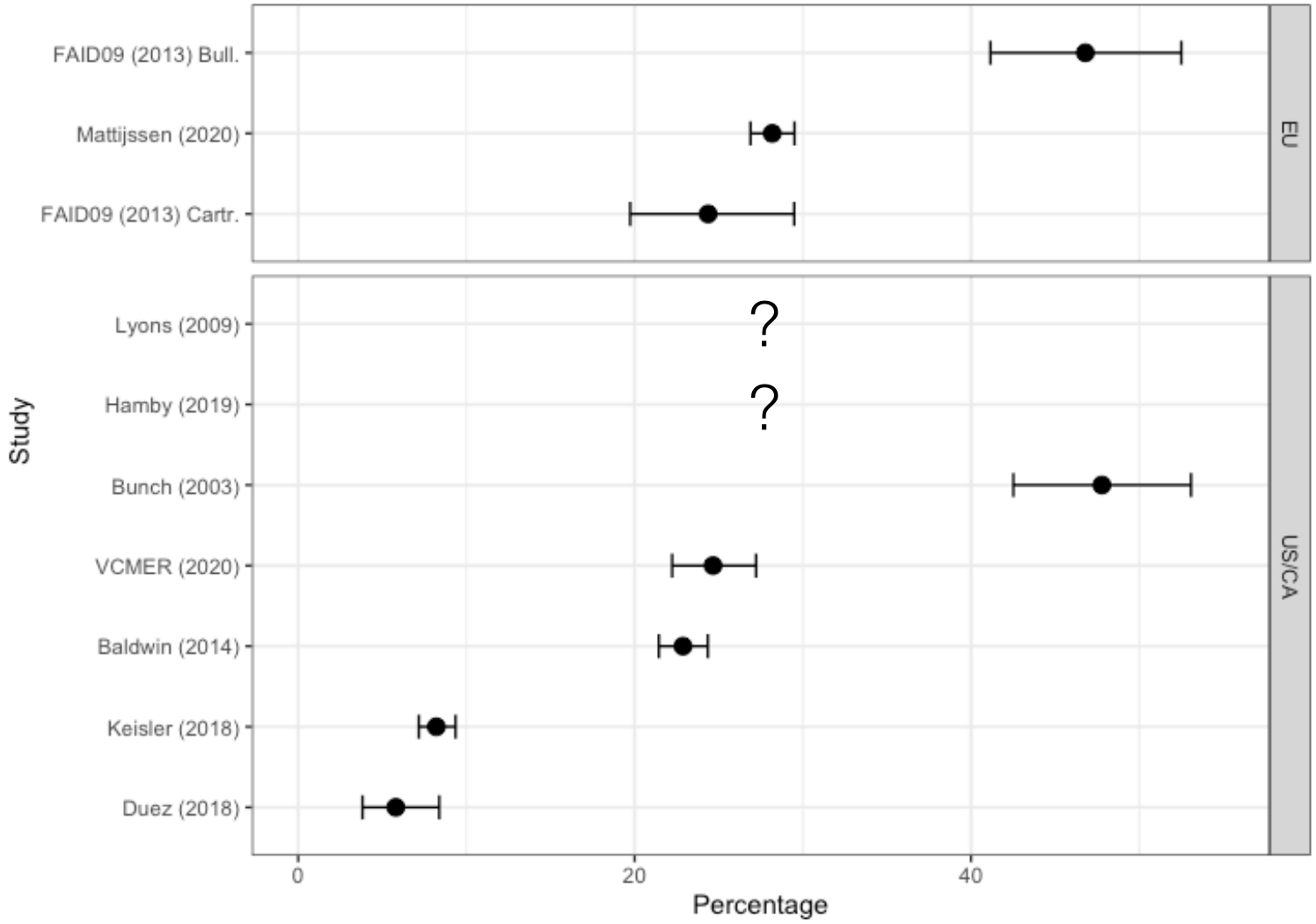
Ten studies on firearms evidence

- **Breech faces** [Baldwin et al., 2014, Keisler et al., 2018, Bunch and Murphy, 2003, Fadul Jr. et al., 2012, *Duez et al., 2018, VCMER Chapnick et al., 2020*, FAID Pauw-Vugts et al., 2013]
- **Bullet lands** [Hamby et al., 2019, FAID Pauw-Vugts et al., 2013]
- **Extractor marks** [Lyons, 2009]
- **Firing pin aperture marks** [Mattijssen et al., 2020]

studies based on virtual microscopy in italics

csafe
Center for Statistics and Applications in Forensic Evidence

Percentage of Inconclusive Results



★ 95% exact confidence intervals (Clopper-Pearson)

csafe
Center for Statistics and Applications in Forensic Evidence

Inconclusives as exclusions

- Idea: do not count inconclusive decisions as final
- Only distinguish between Identifications and no Identification

Inconclusive results counted as exclusion/ elimination

PARTICIPANT'S DECISION	EVIDENCE: Same-source	EVIDENCE: Different-source
Identification	Correct	Error
No Identification	Error	Correct
No Identification	Error	Correct

Types of Errors

★ AFTE

PARTICIPANT'S DECISION	EVIDENCE: Same-source	EVIDENCE: Different-source
Identification	Correct	Error
Exclusion	Error	Correct
Inconclusive	Correct	Correct

★ Process Error

PARTICIPANT'S DECISION		
Identification	Correct	Error
Exclusion	Error	Correct
Inconclusive	Error	Error

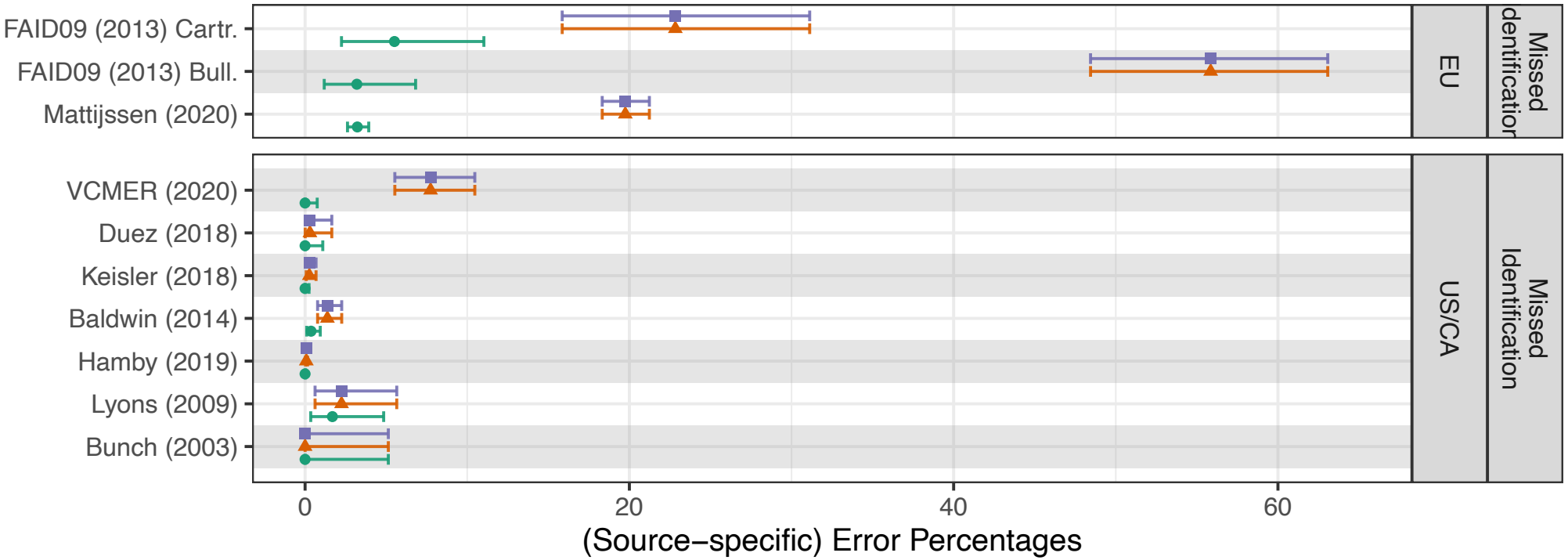
★ Inconclusives and eliminations are no identifications

PARTICIPANT'S DECISION		
Identification	Correct	Error
No Identification	Error	Correct
No Identification	Error	Correct

Missed Identifications

For AFTE trained examiners no huge difference for different errors (except for virtual microscopy study)

Percentage of same-source comparisons that were not identified



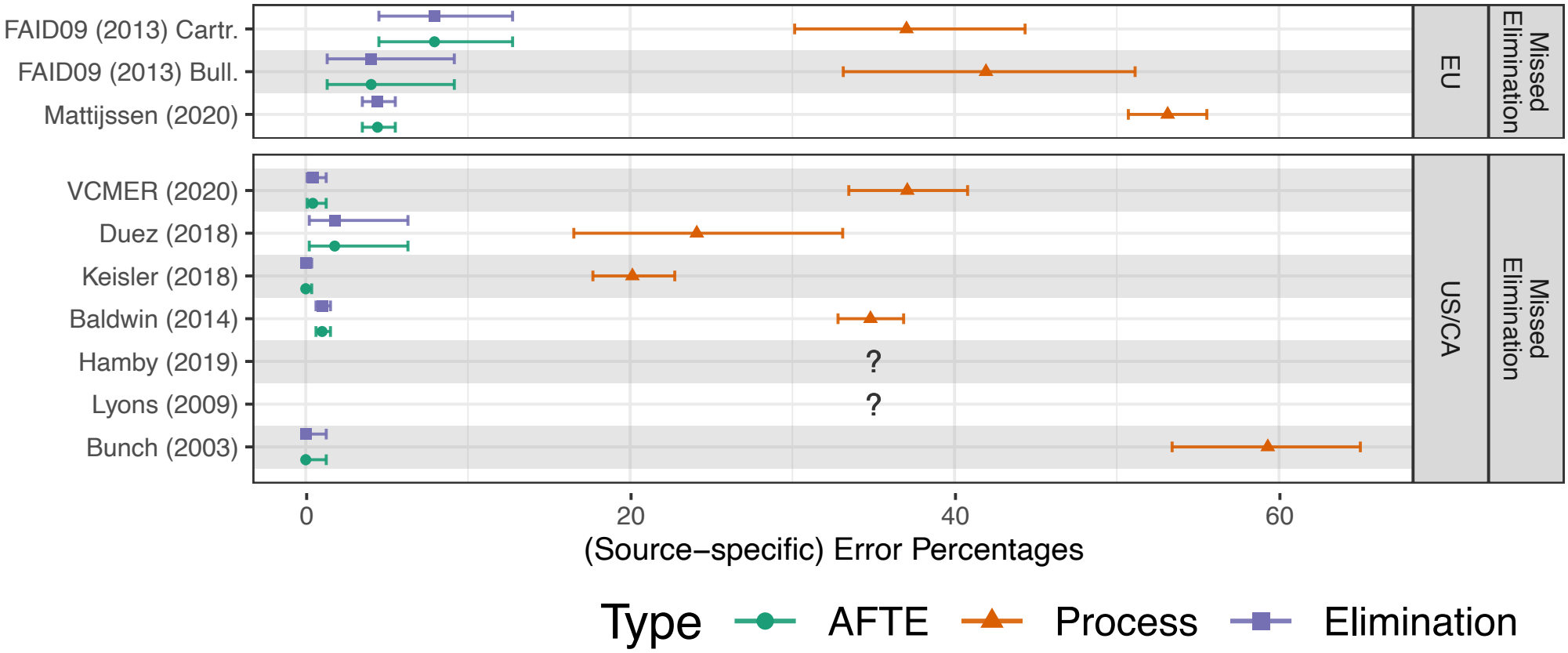
Fadul study did not allow an error analysis

Type — AFTE — Process — Elimination

Missed eliminations

Process errors huge - due to large percentage of inconclusive results for comparisons

Percentage of different-source comparisons that were not excluded



Types of Errors

- ★ AFTE error: inconclusive results are not errors
- ★ Process error: inconclusive results are always errors
- ★ Trade-off: inconclusive results are not identifications, i.e. only distinguish between identification and no identification

- ★ Case Studies: trade-off is more principled than AFTE, and error rates are only slightly increased

Looking at some numbers …

http://bit.ly/FTE-error-rate-worksheet

Baldwin

Experiment Count Data

	Identification	Inconclusive	Elimination	Source Total
Same Source	1075	11	4	1090
Different source	22	735+2[a]	1421	2180
Conclusion Total	1097	748	1425	3270

★ **Missed identification**

P (Elimination | same source) = 4/1090 = 0.0037

P (inconclusive or elimination | same source) = (11 + 4)/1090 = 0.0138

★ **Missed elimination**

P (identification | different source) = 22/2180 = 0.0101.

P (inconclusive or identification | different source) =

= (22 + 737)/2180 = 0.3482.

★ Probability for failing to eliminate MUCH higher than failing to identify

csafe
Center for Statistics and Applications in Forensic Evidence

Conclusions

- AFTE rules do not count inconclusive decisions as errors by examiners

  *Bigger picture needs to consider if the **process** results in the correct conclusion*

- Higher error rate for eliminations/exclusions than for identifications
  - Some labs do not allow exclusions based on individual characteristics
  - Making exclusions might be a cognitively harder task - difference in training?

Hofmann, VanderPlas, Carriquiry, *Treatment of Inconclusives in the AFTE Range of Conclusions*, Law, Probability & Risk, accepted.

csafe
Center for Statistics and Applications in Forensic Evidence

Thank You!

Questions?

csafe
Center for Statistics and Applications in Forensic Evidence

Heike Hofmann (hofmann@iastate.edu, @heike_hh)
Susan VanderPlas (UNL), Alicia Carriquiry (ISU)