

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Publications, Department of Statistics

Statistics, Department of

---

2021

## Treatment of Inconclusive Results in Firearms Error Rate Studies

Heike Hofmann

*Iowa State University*, [hofmann@iastate.edu](mailto:hofmann@iastate.edu)

Susan VanderPlas

*University of Nebraska - Lincoln*

Alicia L. Carriquiry

*Iowa State University*, [alicia@iastate.edu](mailto:alicia@iastate.edu)

Follow this and additional works at: <https://digitalcommons.unl.edu/statisticsfacpub>



Part of the [Other Statistics and Probability Commons](#)

---

Hofmann, Heike; VanderPlas, Susan; and Carriquiry, Alicia L., "Treatment of Inconclusive Results in Firearms Error Rate Studies" (2021). *Faculty Publications, Department of Statistics*. 130.  
<https://digitalcommons.unl.edu/statisticsfacpub/130>

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

1-2021

## Treatment of Inconclusive Results in Firearms Error Rate Studies

Heike Hofmann

*Iowa State University*, [hofmann@iastate.edu](mailto:hofmann@iastate.edu)

Susan VanderPlas

*University of Nebraska - Lincoln*

Alicia L. Carriquiry

*Iowa State University*, [alicia@iastate.edu](mailto:alicia@iastate.edu)

Follow this and additional works at: [https://lib.dr.iastate.edu/csafe\\_conf](https://lib.dr.iastate.edu/csafe_conf)



Part of the [Forensic Science and Technology Commons](#)

---

### Recommended Citation

Hofmann, Heike; VanderPlas, Susan; and Carriquiry, Alicia L., "Treatment of Inconclusive Results in Firearms Error Rate Studies" (2021). *CSAFE Presentations and Proceedings*. 72.

[https://lib.dr.iastate.edu/csafe\\_conf/72](https://lib.dr.iastate.edu/csafe_conf/72)

This Presentation is brought to you for free and open access by the Center for Statistics and Applications in Forensic Evidence at Iowa State University Digital Repository. It has been accepted for inclusion in CSAFE Presentations and Proceedings by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Treatment of Inconclusive Results in Firearms Error Rate Studies

## Disciplines

Forensic Science and Technology



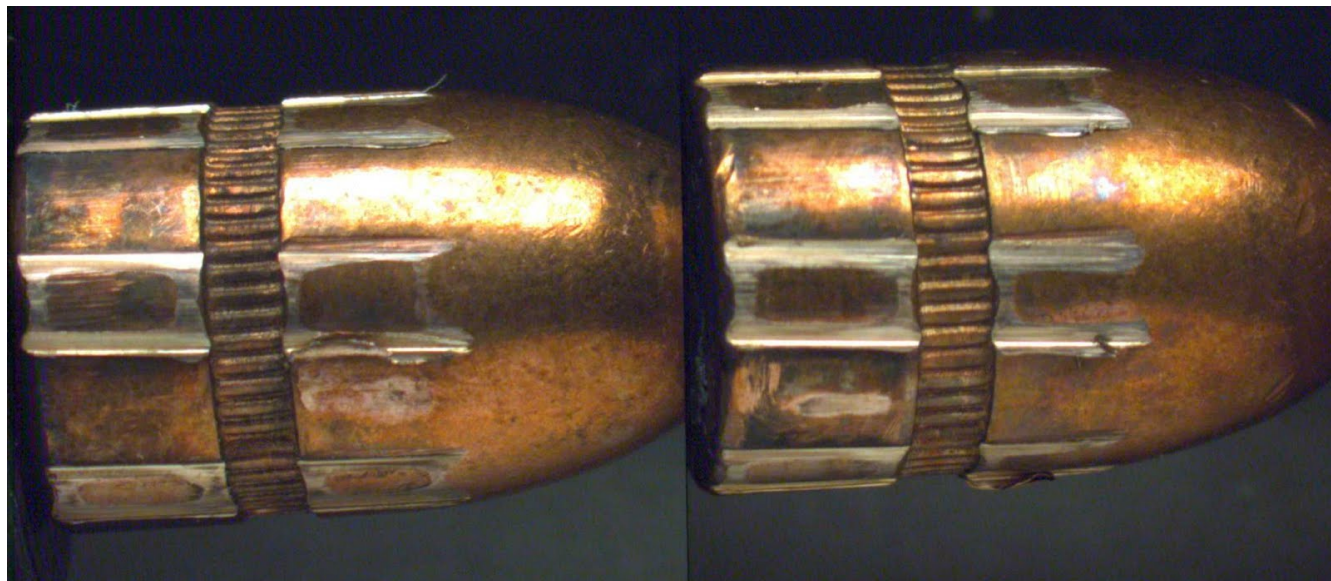
# Treatment of Inconclusive Results in Firearms Error Rate Studies

Heike Hofmann ([hofmann@iastate.edu](mailto:hofmann@iastate.edu), @heike\_hh, ISU),  
Susan VanderPlas (UNL), Alicia Carriquiry (ISU)



# Outline

- ★ Defining error rates for firearms evidence
- ★ Impact of inconclusive decisions on error rates
- ★ Predictive probabilities and errors



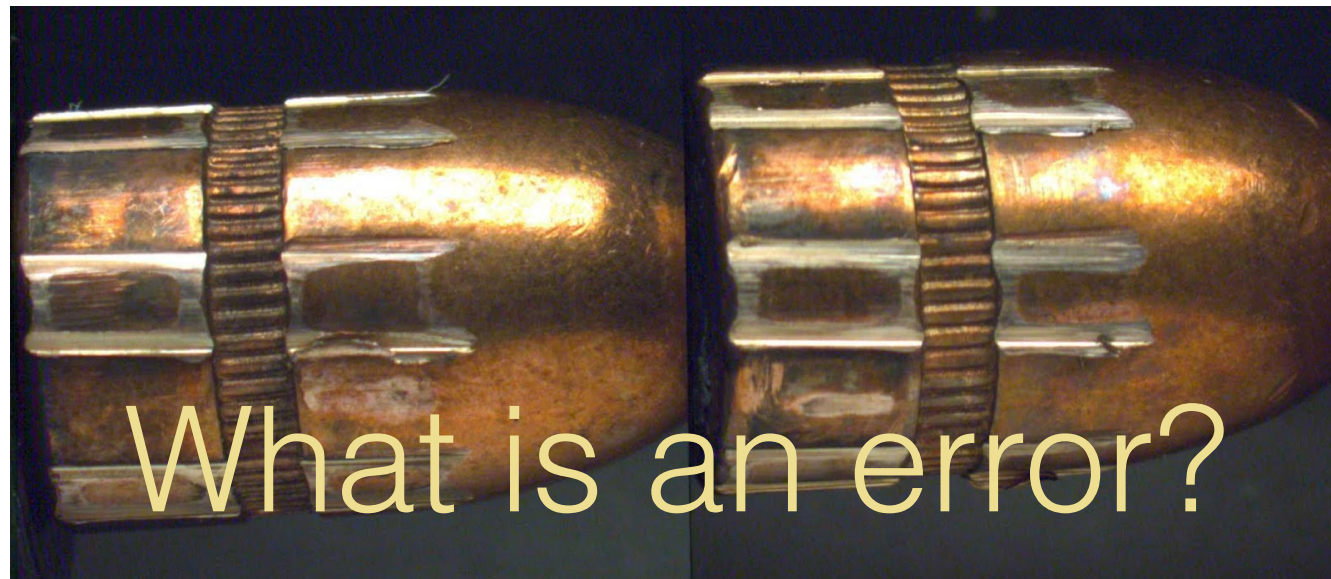
# Over-arching Objective

★ **Same Source Problem:** were two bullets fired through the same gun barrel?

★ **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**



What is an error?

# Quantifying Errors

- ★ Ground truth needed to establish error rates,  
*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 study: 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

## 1. 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.

## 2. Inconclusive

- (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.

## 3. Elimination

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

## 4. Unsuitable

Unsuitable for examination.



# 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.

	<u>EVIDENCE</u>	
	Same-source	Different-source
<u>PARTICIPANT'S DECISION</u>		
Identification	Correct	Error
Exclusion	Error	Correct
Inconclusive	Never considered as potentially incorrect	

# Studies

- ★ **Baldwin** [Baldwin et al., 2014]: Ruger SR9 cartridge cases, open-set study.  
15 test sets (3 references, 1 questioned)
- ★ **Keisler** [Keisler et al., 2018] cartridge cases, open set study.  
20 test sets (1 reference, 1 questioned)  
126 participants.
- ★ **Brundage-Hamby** [Hamby et al., 2019] 10 consecutively manufactured Ruger P-95 barrels, closed-set study.  
~507 test sets (10 pairs references, 15 questioned)
- ★ **Lyons** [Lyons, 2009] 10 consecutively manufactured Colt 1911A1 extractors, closed-set study.  
Test set (10 pairs of references, 12 questioned).

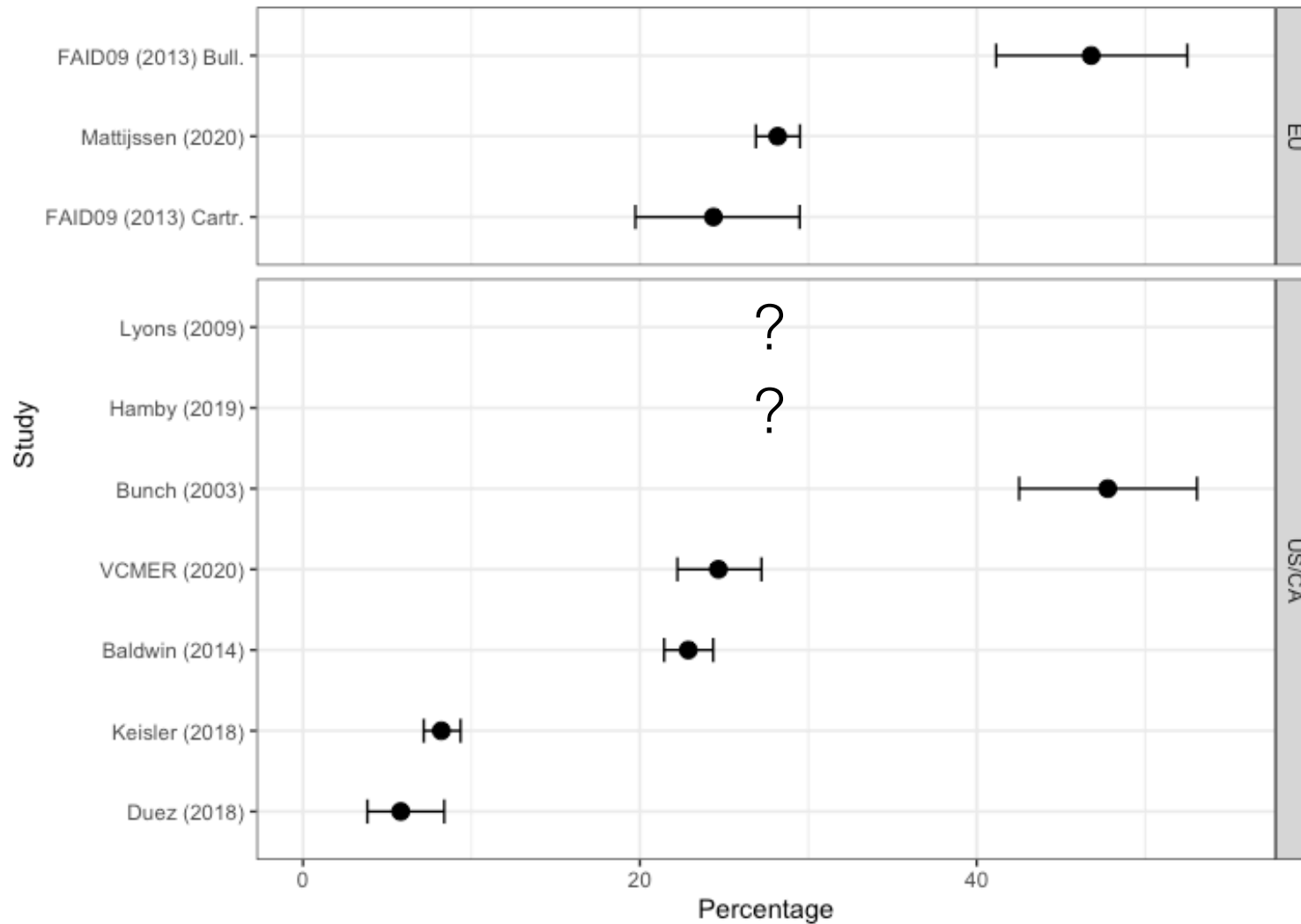
# Studies (2)

- ★ **Bunch** [Bunch and Murphy, 2003] consecutively manufactured Glock breech faces, open-set study.  
Pairwise comparison of 10 cartridge cases.  
8 participants.
- ★ **Fadul** [Fadul Jr. et al., 2012] 10 consecutively manufactured slides, closed-set study.  
Test set1(0 pairs of references, 15 questioned).
- ★ **Duez** [Duez et al., 2018] virtual microscopy, breech face comparisons, open-set study  
two test sets (3 references, 4 questioned)  
56 participants.
- ★ **VCMER** [Chapnick et al., 2020] virtual microscopy, breech face comparisons, open-set study  
16 tests sets (2 references, 1 questioned)  
76 participants.

# Studies (3)

- ★ **Mattijssen** [Mattijssen et al., 2020] firing pin aperture marks from 200 Glock pistols, open-set study.  
Test set: 60 pairwise comparisons  
77 participants.
- ★ **FAID-2009** [Pauw-Vugts et al., 2013] 5 bullets, 5 cartridge cases, open-set study  
10 tests sets (3 references, 1 questioned)  
64 participants.

# Percentage of Inconclusives



★ 95% exact confidence intervals  
(Clopper-Pearson)

# Inconclusives as errors

★ Treatment of inconclusive results hugely impacts error rates

★ AFTE rules measure an examiner's error

<u>PARTICIPANT'S DECISION</u>	
Identification	Correct
Exclusion	Correct
Inconclusive	Never considered as potentially incorrect

## ★ Process error:

process does not result in identification or elimination

Dror & Scurich (2020) proposal:

<u>PARTICIPANT'S DECISION</u>	<u>EVIDENCE</u>		
	Same-source	Different-source	Inconclusive
Identification	Correct	Error	Error
Exclusion	Error	Correct	Error
Inconclusive	Error	Error	Correct

# Inconclusives as exclusions

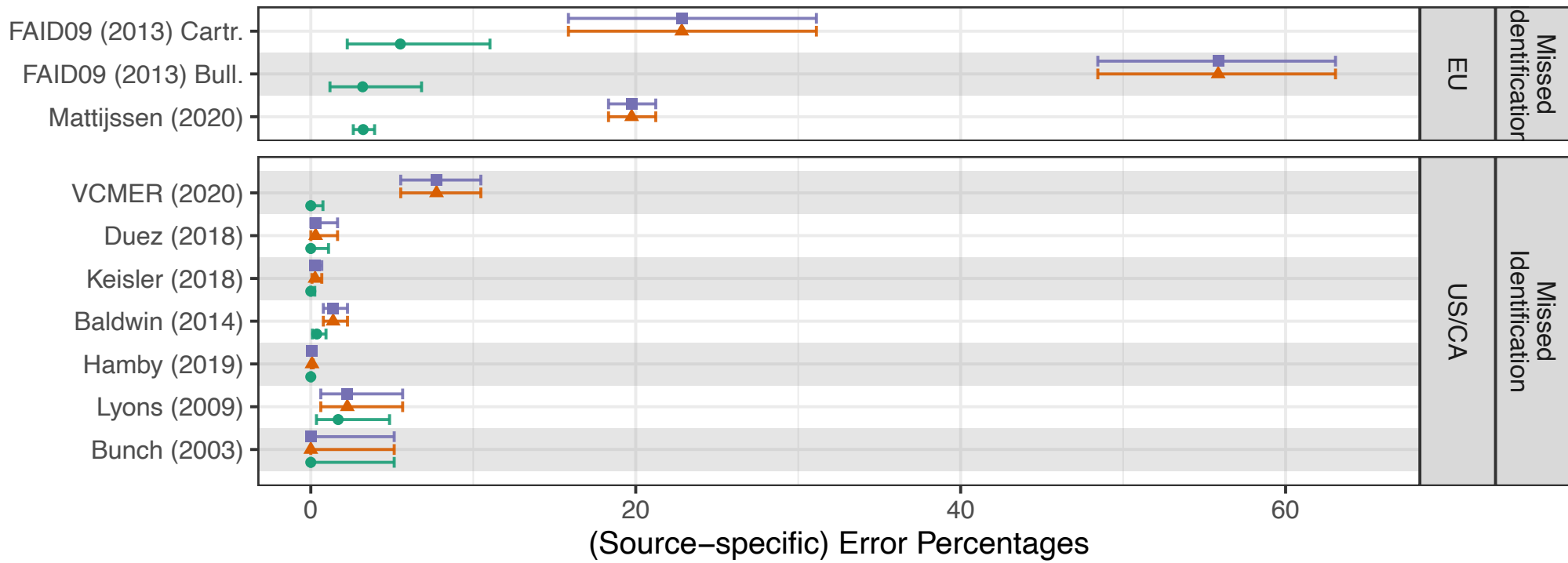
- ★ Idea: do not count inconclusive decisions as final
- ★ Only distinguish between Identifications and no Identification

		<u>EVIDENCE</u>	
		Same-source	Different-source
<u>PARTICIPANT'S DECISION</u>	Identification	Correct	Error
	No Identification	Error	Correct
	No Identification	Error	Correct

# Types of Errors

## ★ Missed identifications

Percentage of same-source comparisons that were not identified



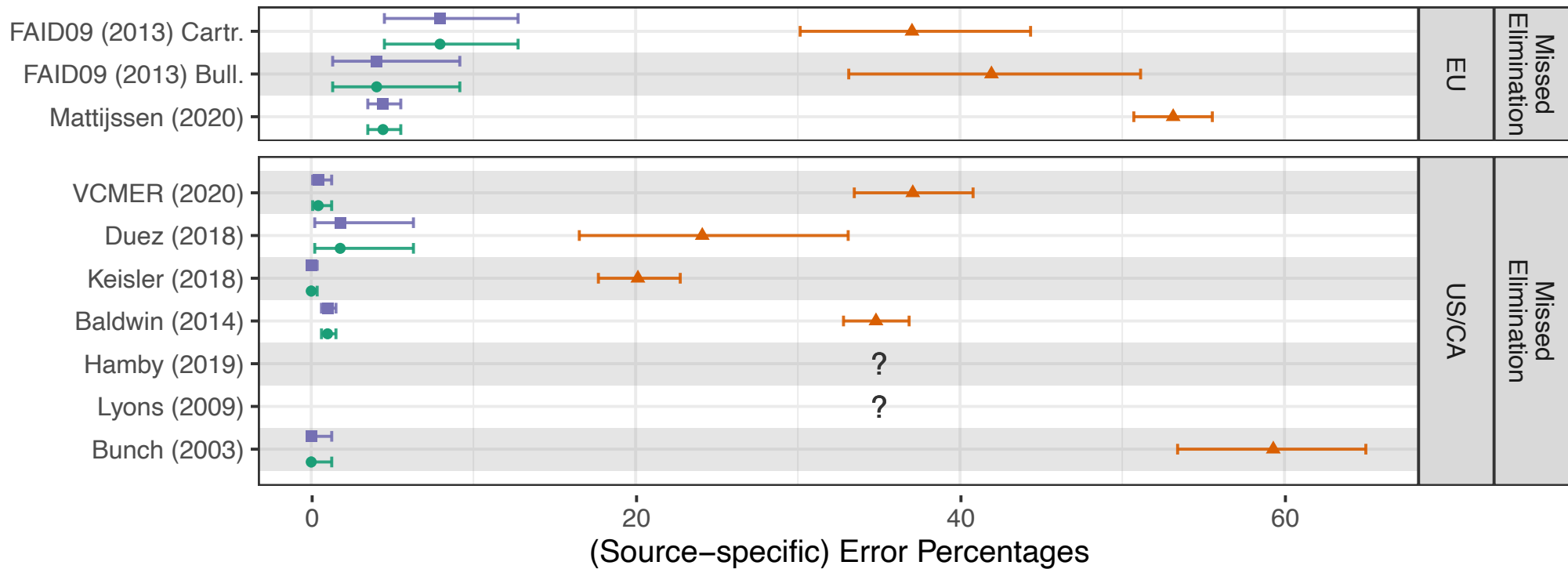
Type ● AFTE ▲ Process ■ Elimination



# Types of Errors

## ★ Missed eliminations

Percentage of different-source comparisons that were not excluded



Type ● AFTE ▲ Process ■ Elimination

# 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

# Predictive Probabilities

# Situational Error Rates

- ★ None of the previous error rates are actually of interest in a legal situation
- ★ What we want: what does an examiner's testimony mean about source of evidence?

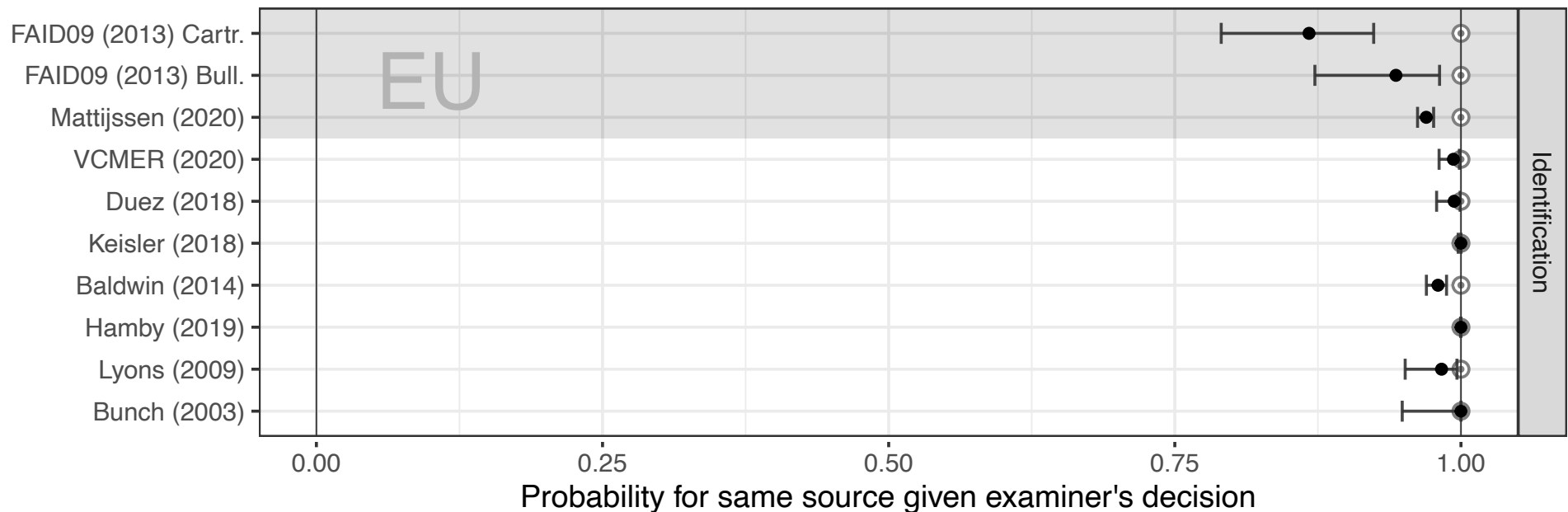
## ★ Predictive Probabilities

$P(\text{same source} \mid \text{examiner's conclusion}) = ?$

$P(\text{different source} \mid \text{examiner's conclusion}) = ?$

# Examiner: identification was made

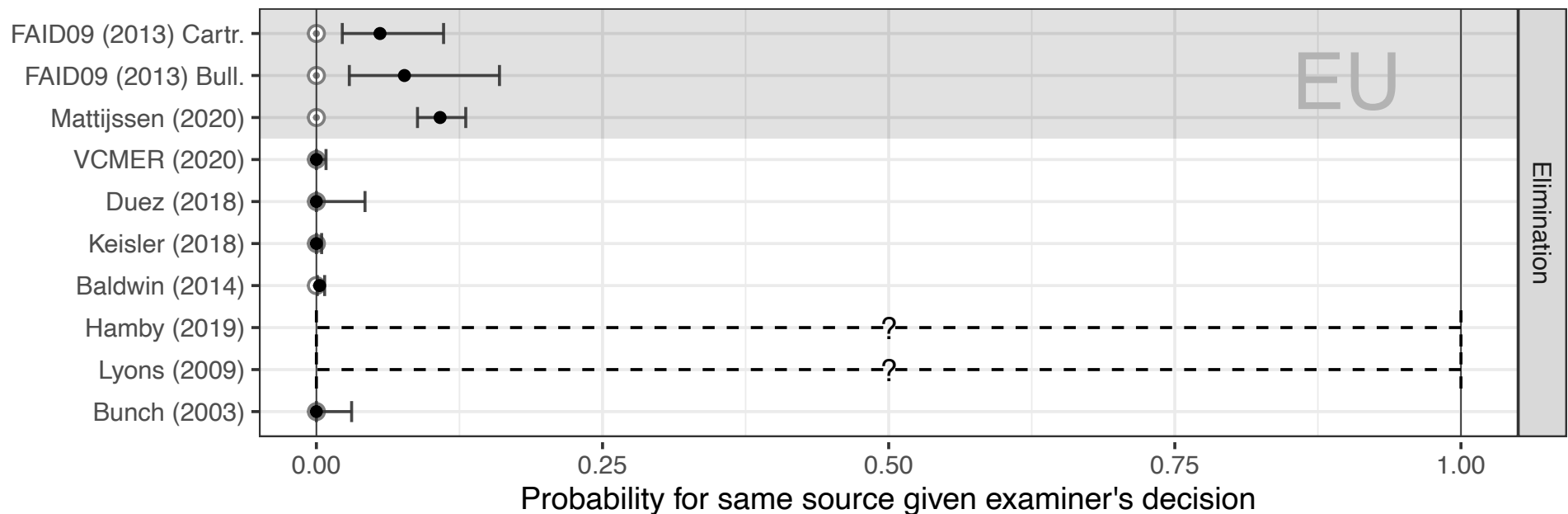
- ★ AFTE trained examiners: probability for same source given identification was made well above 95%



- ★ 95% exact confidence intervals (Clopper-Pearson)

# Examiner: exclusion was made

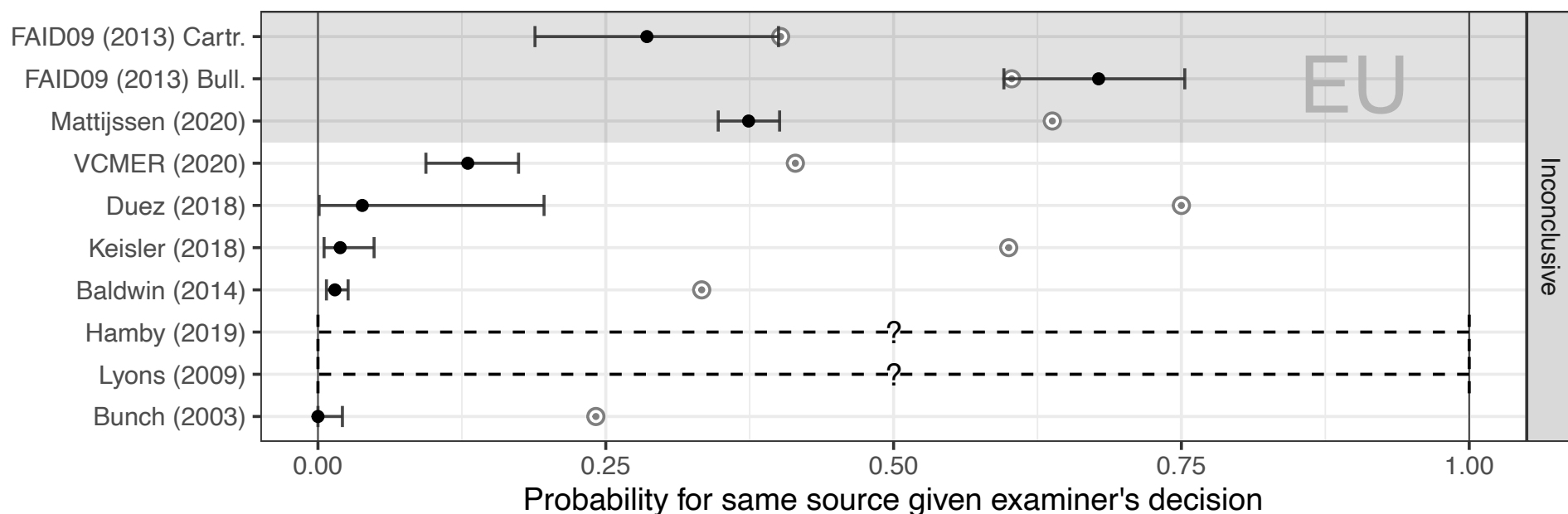
- ★ AFTE trained examiners: probability for same source given exclusion was made well below 1%



- ★ 95% exact confidence intervals (Clopper-Pearson)

# Examiner: inconclusive result

- ★ AFTE trained examiners: probability for same source given inconclusive should be independent of source
- ★ Expected Values based on number SS/DS comparisons



# Looking at some numbers ...

<b>Baldwin</b>	Experiment Count Data			Source Total
	Identification	Inconclusive	Elimination	
Same Source	1075	11	4	1090
Different source	22	735+2 <sup>a</sup>	1421	2180
Conclusion Total	1097	748	1425	3270

## ★ Missed identification

$$P(\text{same source} \mid \text{elimination}) = 4/1425 = 0.0028$$

$$P(\text{same source} \mid \text{inconclusive or elimination}) = (11 + 4)/(748 + 1425) = 0.0069$$

## ★ Missed elimination

$$P(\text{different source} \mid \text{identification}) = 22/1097 = 0.0201.$$

$$P(\text{different source} \mid \text{inconclusive or identification}) = \\ = (22 + 737)/(748 + 1097) = 0.4114.$$

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



# 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*
- ★ For legal situations predictive probabilities are more informative:  
*What does a specific testimony mean, and what is the error rate of that?*
- ★ 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?



Thank You!

Questions?



Heike Hofmann ([hofmann@iastate.edu](mailto:hofmann@iastate.edu), @heike\_hh)  
Susan VanderPlas (UNL), Alicia Carriquiry (ISU)

# CSAFE

- Center for Statistics and Applications in Forensics Evidence
- <https://forensicstats.org/about/>