

CTS Error Rates, 1992 – 2005

Firearms/Toolmarks



Doug Murphy, FBI Laboratory Firearms/Toolmarks Unit

Errors

True Error

Conclusion that is factually wrong, False Positive (mis-identification, Type I error) or False Negative (false elimination, Type II error)

“Unexpected Response”

No conclusion in an exam that most qualified examiners would be able to identify or eliminate

Issues with Inconclusives

Human Performance

Failure to apply an appropriate ID standard

Variation in Test Materials

Hundreds of test bullets, cartridge cases and toolmarks cannot all be reproduced precisely, therefore some inconclusives in Proficiency or Validity tests may be legitimate and justified.

Analysis of Rates

False Positive Rate

- number of incorrect identification conclusions divided by the total number of exam results given on true exclusions (not divided by the total number of exams)
- this is a standard statistical definition and is described explicitly in the NAS Report

Analysis of Rates

False Exclusion Rate

- number of incorrect exclusion conclusions divided by the total number of exam results given on true identifications (not divided by the total number of exams)

Analysis of Rates

- Sensitivity
 - number of correct identification conclusions divided by the total number of exam results on true identifications
 - may vary considerably depending on tool

Analysis of Rates

- Specificity
 - number of correct exclusion conclusions divided by the total number of exams conducted on true exclusions

The Six Exam-Result Conditions

Each exam result consists of two parts, fact and opinion

There are two possible facts and three possible opinions (ID, INC, EX)

$2 \times 3 = 6$, so there are 6 Exam-Result Conditions

The Six Exam-Result Conditions

When Fact=ID (True ID), Three results are possible:

ID - A correct identification

FE - A false elimination

II - A no conclusion result for a true identification

The Six Exam-Result Conditions, continued

When Fact=Elim (True EX), three results are possible:

EX - Correct Exclusion

MI - Mis-Identification

IE - No Conclusion on a True Exclusion

Collaborative Testing Services

- Forensic Laboratory Proficiency Testing
- Supervised by ASCLD Proficiency Advisory Committee
- U.S. and Foreign Lab participation
- Anyone who buys test can participate

Collaborative Testing Services

“ Since it is the laboratory’s option how the samples are to be used (e.g. training exercise, known or blank proficiency testing, research and development of new techniques), the results compiled in the summary report are not intended to be an overview of the quality of work performed in the profession and cannot be interpreted as such.”

Collaborative Testing Services

March 30, 2010 Statement:

CTS Summary Reports should not be used to determine forensic science discipline error rates.

- Tests may be purchased by anyone
- Some non-forensic science organizations participate
- It is solely the responsibility of the participant or accrediting agency to determine the acceptability of an examiner's response
- Reported results do not reflect post examination scrutiny by reviews such as laboratory quality assurance measures

Issues Noted in CTS Results

- Report language varies considerably
 - “may have been fired in”
 - “was probably fired in”
 - “could not be identified as”
- Some reports appear to be non-native English

Value of CTS Results

- Monitor any trends in error rates
- Determine if certain types of exams are more prone to error, take corrective actions
- Defense of F/T against inaccurate error rate claims
- Despite limitations, useful as a supplement to properly designed validity tests

Classifying CTS Responses

- Each individual conclusion section report is analyzed to determine the total number of exam results and to categorize each as one of the six different exam result types
- If written conclusion is unclear or absent, table results are used

Classifying CTS Responses

- Results with qualifying words or phrases such as “was probably fired from” or “in the condition in which it was received” are treated as inconclusive
- If two tools are provided, Exclusions of a second tool due to an ID to the first tool are not tabulated. By extension, all other tools in the universe could be excluded this way.

CTS Results 1992 - 2005

- Firearms (bullets and cartridge cases)
- False Positive Rate $137/9111 = 1.5\%$
- False Negative Rate $31/6114 = 0.5\%$
- Sensitivity $5863/6114 = 95.9\%$
- Specificity $5203/9111 = 57.1\%$

CTS Results 1992 - 2005

- Bullets Only
- False Positive Rate $52/2072 = 2.5\%$
- False Negative Rate $21/2020 = 1.0\%$
- Sensitivity $1843/2020 = 91.2\%$
- Specificity $899/2072 = 43.4\%$

CTS Results 1992 - 2005

- Cartridge Cases Only
- False Positive Rate $59/4851 = 1.2\%$
- False Negative Rate $6/2406 = 0.2\%$
- Sensitivity $2365/2406 = 98.3\%$
- Specificity $2903/4851 = 59.8\%$

CTS Results 1992 - 2005

- Toolmarks Only
- False Positive Rate $84/4950 = 1.7\%$
- False Negative Rate $51/3388 = 1.6\%$
- Sensitivity $3070/3388 = 90.6\%$
- Specificity $2866/4950 = 57.9\%$

CTS Results 1992 - 2005

- General Comments
- Bullet false positive (fp) rate is higher than toolmark fp rate
- Firearms fp rate is 1.5%, same as P&M (see Bunch calculations)

CTS Results - 1992, 1993

- Test 92-4, 4 bullets, one Colt 1911 barrel
- Results indicate which responders are trainees
- Trainees Removed vs. Overall
- false id rate 0.00% vs. 0.83%
- false excl. rate 0.65% vs. 1.28%

CTS Results - 1992, 1993

- Test 92-11, a doorknob a a pair of slip-joint pliers
- Results indicate which responders are trainees
- Trainees Removed vs. Overall
- false id rate 0.00% vs. 0.00%
- false excl. rate 6.59% vs. 8.60%

Possible Sources of Error

- Mislabeling of evidence by examiner
- Mislabeling of evidence by CTS
- Mistake in report or notes
- Poor judgment during exam (inappropriate application of identification standard)
- Poor training resulting in inappropriate identification standard
- Microscopic similarity (extremely unlikely, but still theoretically possible)

Developing Statistical Goals

- Lower false positive and false negative rates as much as possible, but what are the side effects?
- Stricter identification standard will likely result in lower sensitivity, i.e., more “no conclusions”

Developing Statistical Goals

- What is the right combination of low false positives and high sensitivity?
- Any proposed or currently used exam procedure, identification standard or QA protocol should be evaluated (at least in part) on its effect on false positive, false negative, sensitivity and specificity rates.

Thank You / Questions