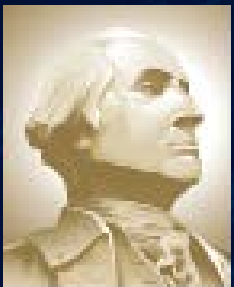


R&D and Maintenance in the making of Successful Databases of Trace Evidence Materials

EDWARD G BARTICK

Department of Forensic Sciences



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

ebartick@gwu.edu

Trace Evidence Data Workshop, NIST, Gaithersburg, MD July 19-20, 2016

Content

- Introduction
 - EGB – Interest & Previous work in databases
- References
 1. Forensic Science Research and Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature, eds. E G Bartick and McK Floyd, (GWU), **Posted on NIJ and NCJRS websites**
 2. Best practices in the collection and use of biometric and forensic Datasets, Austin Hicklin (Noblis), George Kiebuszinski (Noblis) and Melissa Taylor (NIST), **Draft for comment**
- Requirements for successful trace evidence databases for casework use

Introduction

- EGB – Strong Interest in R&D of databases
- EGB – Databases while with the FBI
 - Electrical Tapes by ATR-IR
 - Textile Fibers by Generic and Sub-generic Polymeric Composition by IR, Raman
 - Black copy toners by IR Microscopy & GC/MS, with SLM collaborator on MVA statistical analysis
 - Explosives by ATR- IR, Raman
 - Duct tapes – Physical Measurement, IR-ATR
- EGB & SLM - Textile fibers by multiple characterization with MVA statistical analysis – controlled data from known manufacturer's information – Round robin study for discrimination
- **Ultimate Goal** – Casework ready for full characterization and statistical probabilities

Forensic Science Research and Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature

eds. E G Bartick, McK Floyd, George Washington University, Report Booklet Posted on NIJ and NCJRS websites

This presentation is based in part on a workshop funded by a **National Science Foundation (NSF)** grant (SMA-1533843) from the Science of Science and Innovation Policy and Biological Anthropology Programs in the Directorate for Social, Behavioral and Economic Sciences (SBE), and the **National Institute of Justice (NIJ)**, Office of Justice Programs, U.S. Department of Justice (DOJ).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the Department of Justice.

The workshop was conducted at the **American Association for the Advancement of Science (AAAS)**, Washington, DC, May 26-27, 2015.

Workshop Sections

1. Experimental Design and Statistics
2. Interpretation and Assessment
3. Policy Implications

Requirements for Successful Trace Evidence Databases for Intelligence & Casework Use

1. Recognition and Expression of the needs S Koch
2. Strong understanding of the physical and chemical properties, plus manufacturing process of materials at hand
3. Sound research on analytical methods (1) SL Morgan, SLM
4. Representative sample data for the usage - to avoid biases (2)

Representative sample data: Bias (2)

- Does the database include samples within the designated geological region of the questioned sample?
- Under populated datasets?
- Consider possible within class variation by multiple analyses of samples in dataset and questioned sample.
- Was the questioned sample collected under weather conditions that could alter the composition from those of the dataset?

Separate Reference: Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med, 2(8): e124.

Requirements for Successful Trace Evidence Databases for Intelligence & Casework Use

- Quality control of data input (2)

Quality Control



Quality Control of Data Input (2)

- **The data collection process should be planned in advance to minimize the possibility of human error**
- Primary advantage of **controlled datasets** is that they allow for definitive **ground truth source attribution** — but only if reasonable quality assurance procedures are used to avoid administrative errors
- Effective methods of avoiding administrative errors are generally based on planning
- Design the collection process to be as foolproof as possible
- Limit the possibility of transcription mistakes by using preprinted labels or barcodes for each subject
- Avoid having humans writing long numbers

QC - Continued

- Have someone duplicate the analysis work to double check the results and check the data input.

Requirements for Successful Trace Evidence Databases for Intelligence & Casework Use

1. Recognition and Expression of the needs S Koch
2. Strong understanding of the physical and chemical properties of materials at hand
3. Sound research on analytical methods (1) SL Morgan, SLM
4. Representative sample data for the usage - to avoid biases (1,2)
5. Quality control of data input (2)
6. A means to apply statistics on the results (1) S Lund
 - Sampling repeatability – reliability same answer every time with different samplings of same exemplar
 - Materials population statistics – Probabilities of finding materials with the same properties within the environment

Requirements - Continued

7. Development of a sound, robust software
 - Easy usage
 - Well tested, accurate search results
 - Robust, accessible platform for website dissemination & operation
8. Well trained users on the materials analysis, use of the software and interpretation of the results
9. Data maintenance and software updates
10. \$\$ - Funding!! (1) EGB

Government's Role in Funding Scientific Research

- 2016 US Funding of Research – \$64 Billion
- NIH - \$32 billion
- NASA - \$19 billion
- NSF - \$6 billion overall research. Forensics - about \$2.5 Million Not counting the IU/CRC
- NIJ – The major funding agency for forensic research, about \$30 million This is about 0.1% of NIH
- NIST - \$7 million internal, \$4 million for Ctr. of Excell.

Reference: J. Mervis, Research Agencies Revel in Final 2016 Budget, *Science*, January 2016, Vol 351 Issue 6268, pp.10-11.

If Forensic Science is going to fulfill the need for the rigorous research recommended by the NRC 2009 report, it needs a great deal more funding!!!

Summary of Forensic Database Requirements:

TO BE USEFUL TO FORENSIC SCIENTISTS AND
LAW ENFORCEMENT FOR INVESTIGATIVE
INTELLIGENCE AND SUBSTANTIATING
EVIDENTIAL LINKS OR ASSOCIATIONS
BETWEEN CRIME SCENES AND SUSPECTS –

FORENSIC TRACE EVIDENCE DATABASES
REQUIRE CAREFUL AND THOROUGH ANALYSIS
OF SAMPLES FOR DATA INPUT AND A ROBUST
SOFTWARE PLATFORM.

Thank you for your interest
in trace evidence databases.

Appendix

A. Requirements

- 1-5

- 6-10

B. Research Design & Evaluation ... Pub. Table of Contents

- Section I. Research Design & Statistics

- Section II. Interpretation and assessment

- Section III. Policy Implications

C. Best Practices... Table of Contents.

- Sections: 1. Introduction, Section 2. Collection

- Sections: 3. Documentation & Dissemination, 4. Usage

A. Requirements for Successful Trace Evidence Databases for Intelligence & Casework Use

1. Recognition and Expression of the needs S Koch
2. Strong understanding of the physical and chemical properties, plus manufacturing process of materials at hand
3. Sound research on analytical methods (1) SLM, SLM
4. Representative sample data for the usage - to avoid biases (2)
5. Quality control of data input (2)
6. A means to apply statistical results on the results (1) JBK, S Lund
 - Sampling repeatability – reliability of same answer every time with different samplings of same exemplar
 - Materials population statistics – Probabilities of finding materials with the same properties within the environment

A. Requirements - Continued

7. Development of a sound, robust software
 - Easy usage
 - Well tested, accurate search results
 - Robust, accessible platform for website dissemination & operation
8. Well trained users on the materials analysis, use of the software and interpretation of the results
9. Data maintenance and software updates
10. \$\$ - Funding!! (1) EGB

B. Section I: Experimental Design and Statistics

1. **Plenary:** The State of Research in the Forensic Sciences
C Gatsonis, Brown U
2. Experimental Design in the Physical Sciences SL
Morgan, USC
3. Experiments in the Social Sciences D Scheufele, UWI
4. Non-Experimental Research in Anthropology
M London, GWU & UMD, Smithsonian Inst. and K.G.
Hatala, Met. Museum Nat. Hist.
5. Appropriate Statistics JB Kadane, U Pitt.

B. Section 2: Interpretation and Assessment

1. **Plenary II.** Impact of Problematic Literature J Lentini, Fire Anal.
2. Mistakes in Conducting and Reporting Research M Shermer, Skeptic Mag.
3. Peer Review, OM Smith, AAAS
4. Recognition and Mitigation of Cognitive Bias, I Dror, London College
5. Treatment of Error and Uncertainty in the Literature, T Vosk, Defense Atty.
6. Impact of Literature on the Admissibility Process, M Ambrosino, DOJ, Atty.

B. Section 3: Policy Implications

1. **Plenary III.** Policy Implications of Inadequate Literature R Kostoff, GA State
2. A Quality and Gap Analysis: In Forensic Science Literature D Runkle, AAAS
3. View from a member of the NCFS: A Perspective on Deliberations About Forensic Science and The Path Forward SJ Gates, Jr, U MD
4. How do We Gain Faith in the Scientific Literature? S Cole, UC - Irvine
5. Government's Role in Funding Scientific Research G LaPorte, NIJ, EG Bartick, GWU
6. The Future of Forensic Science Impacted by OSAC Standards M Stolorow, NIST

C. Best Practices in the Collection and Use of Biometric and Forensic Datasets

1 Introduction

1.1 *Appropriateness for use*

1.2 *Categorizing datasets by intended use*

1.3 *Reproducibility and generalization of results*

1.4 *What can go wrong: examples of problems*

2 Collection

2.1 *Types of collection: controlled, operational, manipulated, & synthetic data*

2.2 *Source attribution (ground truth)*

2.3 *Representativeness*

2.4 *Sampling bias*

2.5 *Quality control for controlled dataset collection*

C. Best Practices..... Continued

- 3 Documentation and dissemination
 - 3.1 *Documentation*
 - 3.2 *Data formatting*
 - 3.3 *Dissemination: NIST Catalog and Taxonomy*
 - 3.4 *Dissemination: Public vs sequestered datasets*
- 4 Usage
 - 4.1 *Making effective use of available data*