Automated Quality Metrics and Probabilistic Models in Friction Ridge Examination: A Practical Guide to Implementation

Henry Swofford, Ph.D., CLPE

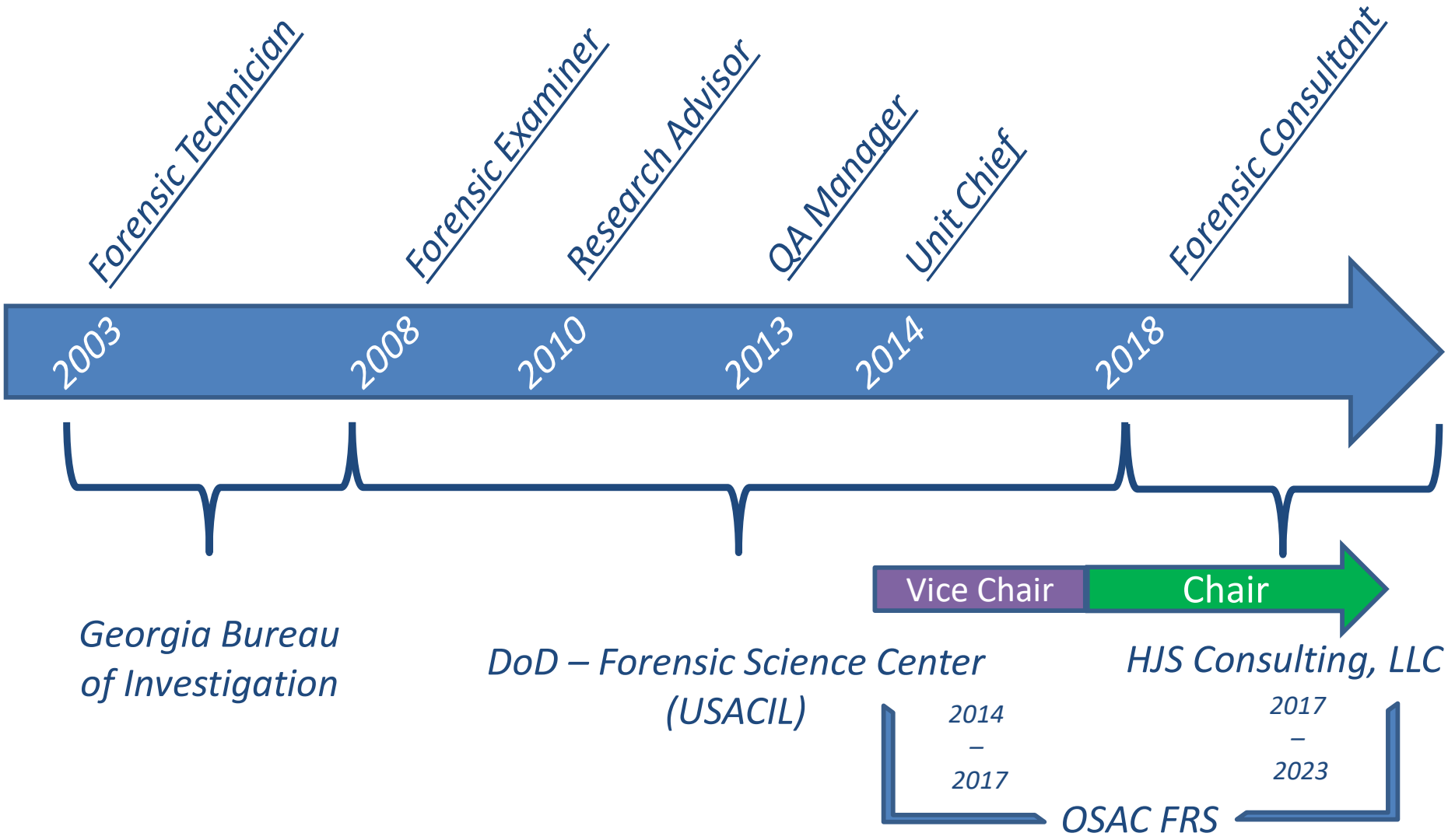
Principal Forensic Analyst
HJS Consulting, LLC
Washington, D.C., USA

International Association for Identification (IAI)
Omaha, NE
August 1, 2022

Disclaimer

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of any organization or entity (government or non-government) of which the author is affiliated.

Background & Experience



Background & Experience

Similarity
Quantification

Forensic Science International 287 (2018) 113–126

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation

H.J. Swofford[a,*], A.J. Koertner[a], F. Zemp[b], M. Ausdemore[c], A. Liu[d], M.J. Salyards[a]

[a] U.S. Army Criminal Investigation Laboratory, Defense Forensic Science Center, USA
[b] School of Criminal Justice, Forensic Science Institute, University of Lausanne, Switzerland
[c] Department of Mathematics and Statistics, University of South Dakota, USA
[d] Department of Statistics, University of Virginia, USA

ARTICLE INFO

Article history:
Received 3 December 2017
Received in revised form 8 March 2018
Accepted 27 March 2018
Available online 3 April 2018

Keywords:
Forensic science
Fingerprints
Strength of evidence
Weight of evidence
Likelihood ratio
Probability

ABSTRACT

The forensic fingerprint community has faced increasing amounts of criticism by scientific and legal commentators, challenging the validity and reliability of fingerprint evidence due to the lack of an empirically demonstrable basis to evaluate and report the strength of the evidence in a given case. This paper presents a method, developed as a stand-alone software application, *FRStat*, which provides a statistical assessment of the strength of fingerprint evidence. The performance was evaluated using a variety of mated and non-mated datasets. The results show strong performance characteristics, often with values supporting specificity rates greater than 99%. This method provides fingerprint experts the capability to demonstrate the validity and reliability of fingerprint evidence in a given case and report the findings in a more transparent and standardized fashion with clearly defined criteria for conclusions and known error rate information thereby responding to concerns raised by the scientific and legal communities.

Published by Elsevier B.V.

Forensic Science International 320 (2021) 110703

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

A method for measuring the quality of friction skin impression evidence: Method development and validation

H. Swofford[a,*], C. Champod[a], A. Koertner[b], H. Eldridge[a,c], M. Salyards[d]

[a] School of Criminal Justice, Forensic Science Institute, University of Lausanne, Switzerland
[b] U.S. Army Criminal Investigation Laboratory, Defense Forensic Science Center, USA
[c] RTI International, Inc., USA
[d] Compass Scientific, LLC, USA

ARTICLE INFO

Article history:
Received 25 September 2020
Received in revised form 13 January 2021
Accepted 19 January 2021
Available online 23 January 2021

Keywords:
Forensic science
Fingerprints
Quality metric
Probability

ABSTRACT

The forensic fingerprint community has faced increasing criticism by scientific and legal commentators, challenging the validity and reliability of fingerprint evidence due to the lack of an empirical basis to assess the quality of the friction ridge impressions. This paper presents a method, developed as a stand-alone software application, DFIQI ("Defense Fingerprint Image Quality Index"), which measures the clarity of friction ridge features (locally) and evaluates the quality of impressions (globally) across three different scales: value, complexity, and difficulty. Performance was evaluated using a variety of datasets, including datasets designed to simulate casework and a dataset derived directly from casework under operational conditions. The results show performance characteristics that are consistent with experts' subjective determinations. This method provides fingerprint experts: (1) a more rigorous approach by providing an empirical foundation to support their subjective determinations from the Analysis phase of the examination methodology, (2) a framework for organizations to establish transparent, measurable, and demonstrable criteria for Value determinations, (3) and a means of flagging impressions that are vulnerable to erroneous outcomes or inconsistency between experts (e.g., higher complexity and difficulty), and (4) a method for quantitatively summarizing the overall quality of impressions for ensuring representative distributions for samples used in research designs, proficiency testing and error rate testing, and other applications by forensic science stakeholders.

Quality Assessment

Background & Experience

DEPARTMENT OF THE ARMY
DEFENSE FORENSIC SCIENCE CENTER
4930 N 31ST STREET
FOREST PARK GA 30297-5205

CIFS-FSL-LP 09 March 2017

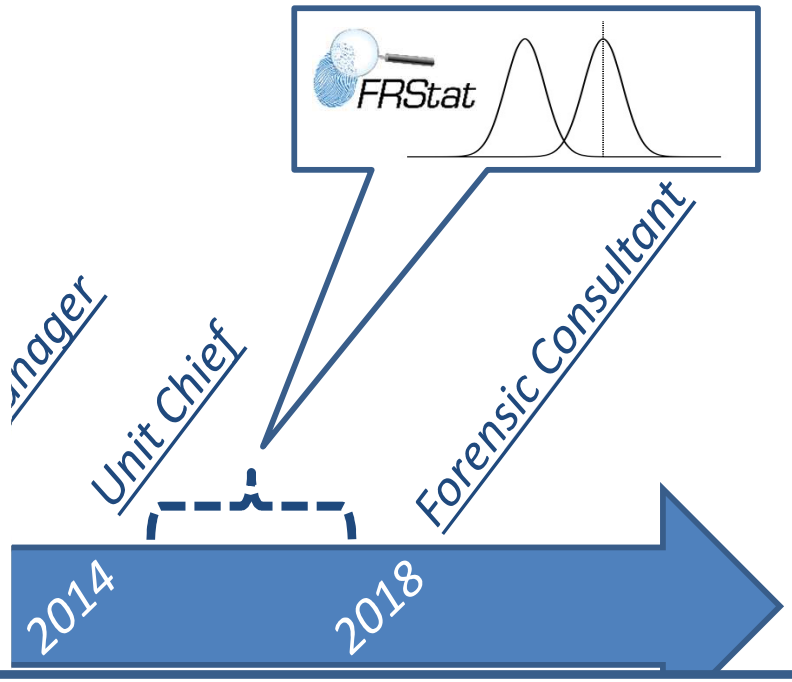
INFORMATION PAPER

SUBJECT: Modification of Latent Print Technical Reports to Include Statistical Calculations

1. Purpose: To provide information regarding the use of statistical calculations in latent print technical reports and guidance on how to interpret the significance of the results.

2. Background:

   a. Results of forensic latent print examinations are traditionally based on the interpretations and opinions of forensic experts and reported as categorical statements of inclusion or exclusion of a particular individual as the source of a latent print.



Reactions experienced against the use of probabilistic reporting and models:
"unnecessary", "unhelpful", "creates backlogs", "confusing to juries", "too weak", "imprecise", "inaccurate", "technology not ready", "don't understand", "not comfortable testifying to statistics", "not a panacea for error", "undermining to experts", "courts don't require it", "not generally accepted / admissible", "creates reasonable doubt", "threatening to careers" ...

Background & Experience

Then

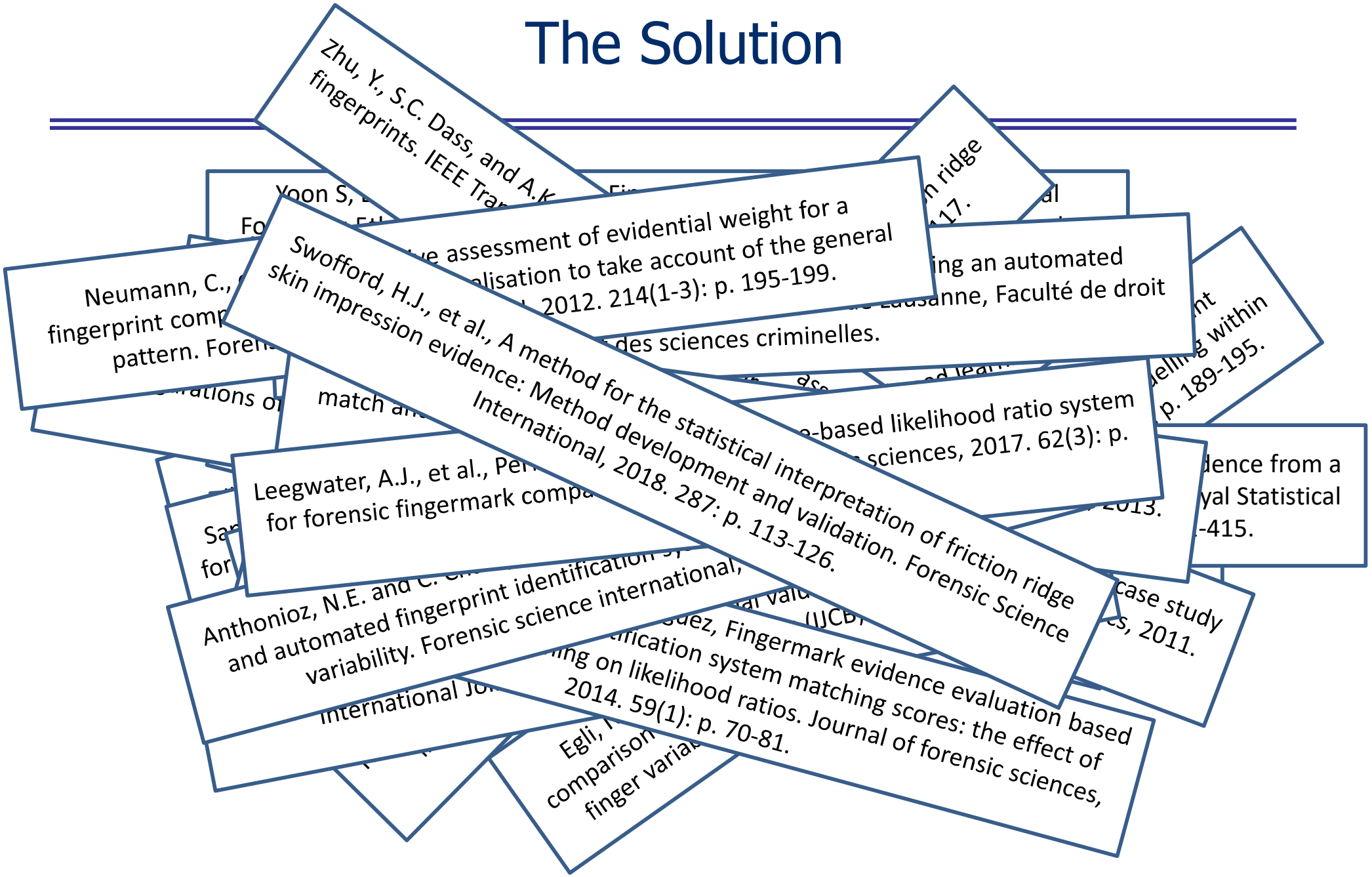
Need better research . . .

Now

Need better strategy . . .

Calls for Reform

The Solution



Research and technology have proliferated . . .

The Result

But . . . Current practice remains *the same . . .*

Challenges

Challenges

Law, Probability and Risk (2021) **00**, 1–37 doi:10.1093/lpr/mgab003

Mt. Everest—we are going to lose many: a survey of fingerprint examiners' attitudes towards probabilistic reporting

H. Swofford
School of Criminal Justice, Forensic Science Institute, Univ

S. Cole† and V. King
Department of Criminology, Law & Society, University o

[Received on 25 May 2020; revised on 11 February 2021;

Over the past decade, with increasing scientific scrutiny on f
have been several efforts to introduce statistical thinking and pr
practice. These efforts have been met with mixed reactions—a
downright hostility, towards this objective. For probabilistic re
practice, more than statistical knowledge will be necessary. So
critical to effectively understand the sources of concern and barr
reports the findings of a survey of forensic fingerprint examiner
the discipline and practitioners' attitudes and characterizations
despite its adoption by a small number of practitioners, commur
reporting in the friction ridge discipline faces challenges. We fou
rently report probabilistically. Perhaps more surprisingly, most
probabilistically, in fact, do not. Furthermore, we found that t
probabilistic reporting as 'inappropriate'—their most comm
attorneys would take advantage of uncertainty or that probabili
misunderstood by, other criminal justice system actors. If proba
much work is still needed to better educate practitioners on the in
tic reasoning in order to facilitate a path towards improved report

Keywords: reporting; testimony; fingerprint; categoric; probabili

Forensic Science International: Synergy 4 (2022) 100220

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: www.sciencedirect.com/journal/forensic-science-international-synergy

Probabilistic reporting and algorithms in forensic science: Stakeholder perspectives within the American criminal justice system

H. Swofford *, C. Champod

School of Criminal Justice, Forensic Science Institute, University of Lausanne, Switzerland

ARTICLE INFO

Keywords:
Forensic science
Pattern evidence
Probabilities
Statistics
Algorithms

ABSTRACT

In recent years, there have bee
rithms across several forensic sc
argue they promote greater sci
challenging to meaningfully scr
Consequently, the forensic cor
proposed approach has counte
dation for a path forward, this
perspectives of key criminal jus
judges, and other academic sch
computational algorithms in fo

1. Introduction

Forensic science has long been considered a cornerstone for advancing investigations and establishing facts in question to support criminal and civil litigation. Under the powerful aura of science, interpretations and conclusions made by forensic experts are often presented as tantamount to fact—the silent witness—that courts can rely on in their pursuit of justice. For decades on end, forensic evidence was broadly considered infallible and rarely questioned. In February 2009, however, that all changed with the release of the National Research Council's (NRC) report on the needs of the forensic science community, highlighting that "[t]he law's greatest dilemma in its heavy reliance on forensic evidence, however, concerns the question of whether—and to what extent—there is *science* in any given forensic science discipline"

Forensic Science International: Synergy 3 (2021) 100142

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: https://www.journals.elsevier.com/forensic-science-international-synergy/

Implementation of algorithms in pattern & impression evidence: A responsible and practical roadmap

H. Swofford*, C. Champod

School of Criminal Justice, Forensic Science Institute, University of Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 5 December 2020
Received in revised form
10 February 2021
Accepted 11 February 2021
Available online 18 February 2021

Keywords:
Forensic science
Pattern evidence
Algorithms
Statistics
Models
Automation

ABSTRACT

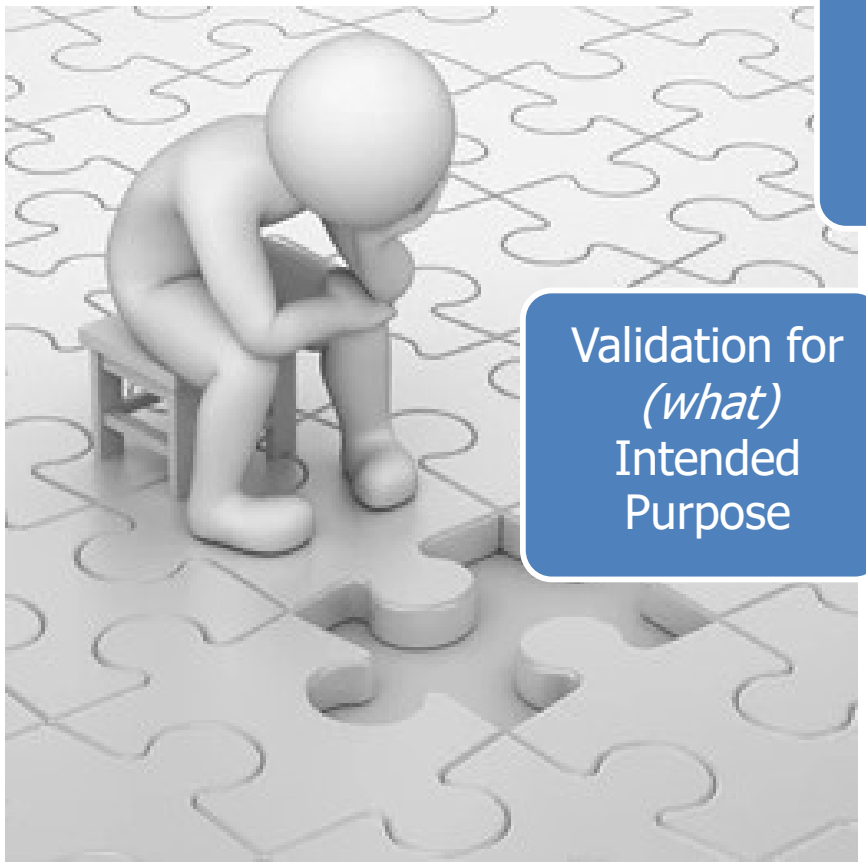
Over the years, scientific and legal scholars have called for the implementation of algorithms (e.g., statistical methods) in forensic science to provide an empirical foundation to experts' subjective conclusions. Despite the proliferation of numerous approaches, the practitioner community has been reluctant to apply them operationally. Reactions have ranged from passive skepticism to outright opposition, often in favor of traditional experience and expertise as a sufficient basis for conclusions. In this paper, we explore *why* practitioners are generally in opposition to algorithmic interventions and *how* their concerns might be overcome. We accomplish this by considering issues concerning human—algorithm interactions in both real world domains and laboratory studies as well as issues concerning the litigation of algorithms in the American legal system. Taking into account those issues, we propose a strategy for approaching the implementation of algorithms, and the different ways algorithms can be implemented, in a *responsible* and *practical* manner.

1. Introduction

Over the years, the forensic science community has faced increasing criticism from scientific and legal scholars, challenging the validity and reliability of many forensic examination methods that rely on subjective interpretations by forensic practitioners [1–9]. Of particular concern is the lack of an empirically demonstrable basis to substantiate conclusions from pattern and impression evidence, which has led to calls for reform through the its long-standing history and ubiquitous practice. In the friction ridge discipline in particular, there have been a number of notable efforts by researchers for which algorithms have been introduced to provide quantitative or statistical approaches to the analysis and evaluation of evidence [10–37]. Despite the proliferation of proposed methods, however, the practitioner community has been reluctant to apply them operationally. Reactions toward the intervention of statistical methods, even statistical *concepts*, have ranged from passive observation and skepticism to outright op-

The Challenge . . .

There's more to it than just performance . . .



"willing to use"

Psychological and Sociological Factors

"able to use"

Legal & Regulatory Requirements

Validation for *(what)* Intended Purpose

Strategy

"trusted to use"

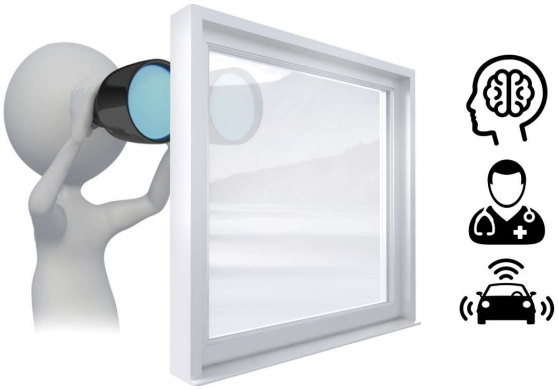
Quality Management System

Will people be *willing* to use it?

Q: What are the behavioral tendencies toward algorithms?

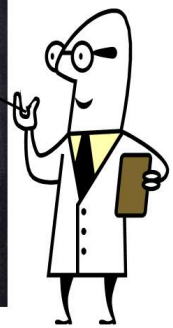
People tend to be averse to algorithms and prefer to rely on their own judgment—often despite knowledge that their own judgment is typically inferior to that of algorithms. This phenomenon is exacerbated when people:

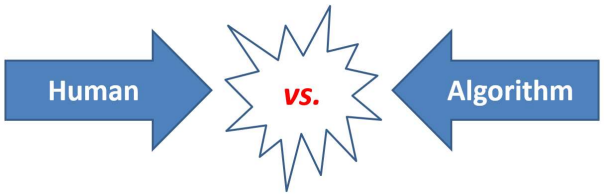
→ Possess domain expertise
→ Are faced with high stakes decisions
→ Presented with an imperfect algorithm



Possible Explanations:

- Sociopsychological factors (e.g.):
  - Dehumanizing
  - Unethical (for important decisions)
  - Desire for perfection
  - Self-worth
  - Poor education (on scientific issues) and lack of emphasis on scientific thinking
- Overconfidence bias
- Perceived inability of the algorithm to account for idiosyncratic factors





Will people be *able* to use it?

Q: Is the technology admissible in court?

Computational methods have potential to improve performance, but often at the cost of transparency and explainability – Courts must balance these competing values as it considers admissibility and upholds the constitutional rights of the defendant. Admissibility depends on:

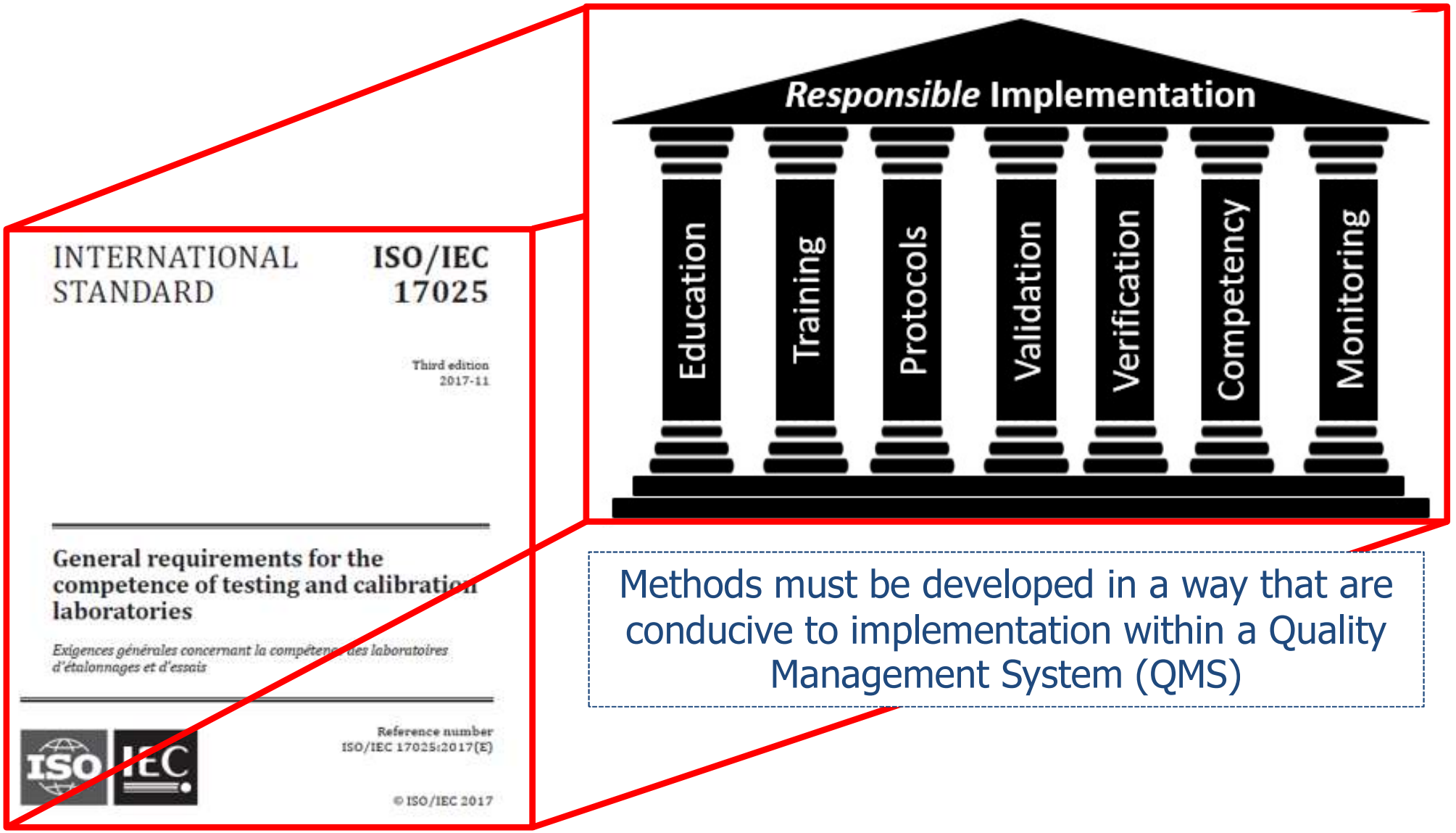
1. Do validation studies adequately address the circumstances in the present case?
   - ➢ Rules of Evidence: *Relevance, Reliability*

2. Can the defendant challenge the credibility of the algorithm (e.g., design, validation, operation)?
   - ➢ Constitutional Rights: *Due Process, Confrontation*

3. To what extent did the algorithm impact the ultimate conclusion?
   - ✓ Greater impact → Greater need for scrutiny, transparency, and explainability

Will people be *trusted* to use it?

How will it fit within a Quality Management System (QMS)?



Methods must be developed in a way that are conducive to implementation within a Quality Management System (QMS)

Issues to Reconcile

1. We must be clear about how human experts and computational tools can cohabitate (e.g., roles, responsibilities).

2. We must allow for flexibility in how the computational tools are used, and the extent to which they provide the basis for the evaluation of the evidence.

3. We must allow for the computational tools to be implemented in a way that preserves existing practices and does not introduce risk to the admissibility of the evidence overall.

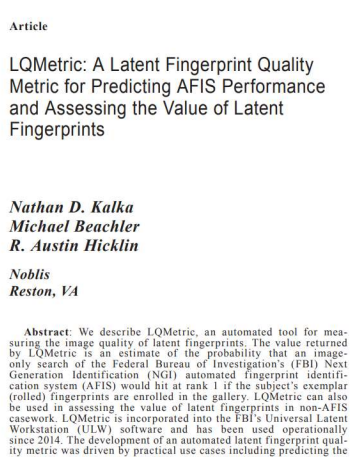
A Framework for *Practical* Implementation

Level	Name	Narrative Definition (adapted)	Human Role	Algorithm Role	Conflict Resolution	Basis for Conclusion
0	No Algorithm	No Algorithm	Evaluation	N/A	N/A	Expert Opinion
1	Algorithm Assistance	Optional Use (*after* expert opinion)	Evaluation	Supplemental Assistance (optional)	Expert Discretion	Expert Opinion
2	Algorithm Quality Control	Required Quality Control (*after* expert opinion)	Evaluation	Supplemental Quality Control	Standard Operating Procedures	Expert Opinion (Algorithm Supported)
3	Algorithm Informed Evaluation	Supporting Basis for Opinion (*before* expert opinion)	Human-Algorithm Integrated Evaluation	Human-Algorithm Integrated Evaluation	Standard Operating Procedures	Algorithm Output (Human Supported)
4	Algorithm Dominated Evaluation	Basis for Conclusion (no expert opinion)	Procedural Oversight	Evaluation	Standard Operating Procedures	Algorithm Output
5	Algorithm Only	"Lights Out" (no human oversight)	N/A	Evaluation	N/A	Algorithm Output

Adapted from: **Swofford, H. and Champod, C. Implementation of Algorithms in Pattern & Impression Evidence: A Responsible and Practical Roadmap, *Forensic Science International—Synergy*, 3:2021, 100142.**

Quality Metrics and Probabilistic Models

Quality Metric Software Tools



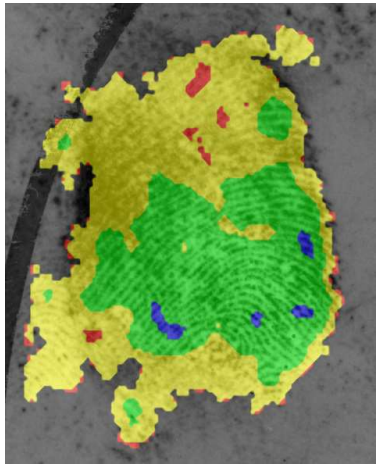
Article

LQMetric: A Latent Fingerprint Quality Metric for Predicting AFIS Performance and Assessing the Value of Latent Fingerprints

Nathan D. Kalka
Michael Beachler
R. Austin Hicklin

Noblis
Reston, VA

Abstract: We describe LQMetric, an automated tool for measuring the image quality of latent fingerprints. The value returned by LQMetric is an estimate of the probability that an image-only search of the Federal Bureau of Investigation's (FBI) Next Generation Identification (NGI) automated fingerprint identification system (AFIS) would hit at rank 1 if the subject's exemplar (rolled) fingerprints are enrolled in the gallery. LQMetric can also be used in assessing the value of latent fingerprints in non-AFIS casework. LQMetric is incorporated into the FBI's Universal Latent Workstation (ULW) software and has been used operationally since 2014. The development of an automated latent fingerprint quality metric was driven by practical use cases including predicting the



Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

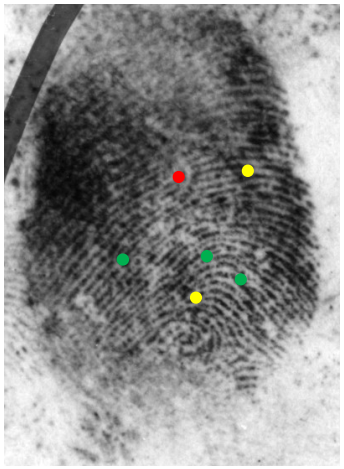
A method for measuring the quality of friction skin impression evidence: Method development and validation

H. Swofford[a,*], C. Champod[a], A. Koertner[b], H. Eldridge[a,c], M. Salyards[d]

ARTICLE INFO

ABSTRACT

The forensic fingerprint community has faced increasing criticism by scientific and legal commentators, challenging the validity and reliability of fingerprint evidence due to the lack of an empirical basis to assess the quality of the friction ridge impressions. This paper presents a method, developed as a stand-alone software application, DFIQI ("Defense Fingerprint Image Quality Index"), which measures the clarity of friction ridge features (locally) and evaluates the quality of impressions (globally) across three different scales: value, complexity, and difficulty. Performance was evaluated using a variety of datasets, including datasets designed to simulate casework and a dataset derived directly from casework under operational conditions. The results show performance characteristics that are consistent with experts' subjective determinations. This method provides fingerprint experts: (1) a more rigorous approach by providing an empirical foundation to support their subjective determinations from the Analysis phase of the examination methodology, (2) a framework for organizations to establish transparent, measurable, and demonstrable criteria for Value determinations, (3) and a means of flagging impressions that are vulnerable to erroneous outcomes or inconsistency between experts (e.g., higher complexity and difficulty), and (4) a method for quantitatively summarizing the overall quality of impressions for ensuring representative distributions for samples used in research designs, proficiency testing and error rate testing, and other applications by forensic science stakeholders.



https://forms.fbi.gov/universal-latent-workstation-ulw-software-download-request

https://doi.org/10.5281/zenodo.4426344

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

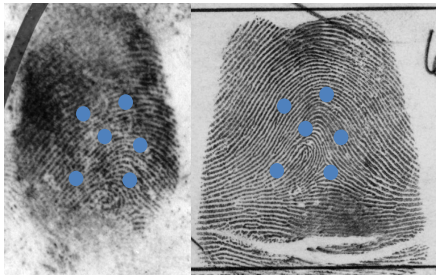
A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation

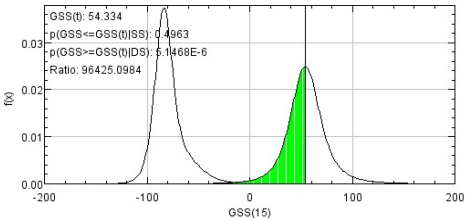
H.J. Swofford[a,*], A.J. Koertner[a], F. Zemp[b], M. Ausdemore[c], A. Liu[d], M.J. Salyards[a]

ARTICLE INFO

ABSTRACT

The forensic fingerprint community has faced increasing amounts of criticism by scientific and legal commentators, challenging the validity and reliability of fingerprint evidence due to the lack of an empirically demonstrable basis to evaluate and report the strength of the evidence in a given case. This paper presents a method, developed as a stand-alone software application, FRStat, which provides a statistical assessment of the strength of fingerprint evidence. The performance was evaluated using a variety of mated and non-mated datasets. The results show strong performance characteristics, often with values supporting specificity rates greater than 99%. This method provides fingerprint experts the capability to demonstrate the validity and reliability of fingerprint evidence in a given case and report the findings in a more transparent and standardized fashion with clearly defined criteria for conclusions and known error rate information thereby responding to concerns raised by the scientific and legal communities.





https://doi.org/10.5281/zenodo.4426484

LQMetric (Kalka et al.)
DFIQI (Swofford et al.)
LFIQ (Yoon et al.)
ESLR (Stoney et al.)
SNoQE (Richter et al.)

https://forensicstats.org/quality-metric-algorithms-for-fingerprint-images/

Looking Forward . . .

- Quantitative measures are critical for ensuring robust scientific practices and enduring admissibility of forensic evidence.
- Validated quality assessment tools and probabilistic models are free *and* available for anyone to access.
- Recommend introducing the tools as a Level 1 first, then proceeding to Level 2 as the “Target Level” for implementation.

Questions / Discussion

Henry Swofford, Ph.D., CLPE

Principal Forensic Analyst
HJS Consulting, LLC
Washington, D.C., USA