NISTIR 8351-DRAFT

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Hari Iyer
Rich Press
Melissa K. Taylor
Peter M. Vallone
Sheila Willis*
*International Associate under contract; retired director of Forensic Science Ireland

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351-draft

National Institute of Standards and Technology
U.S. Department of Commerce

NISTIR 8351-DRAFT

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Melissa K. Taylor
Sheila Willis*
Special Programs Office
Associate Director of Laboratory Programs

Hari Iyer
Statistical Engineering Division
Information Technology Laboratory

Peter M. Vallone
Biomolecular Measurement Division
Material Measurement Laboratory

Rich Press
Public Affairs
Director's Office

*International Associate under contract; retired director of Forensic Science Ireland

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351-draft

June 2021

U.S. Department of Commerce
Gina M. Raimondo, Secretary of Commerce

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Internal Report 8351-DRAFT
(June 2021)

Acknowledgments: Members of the DNA Mixture Resource Group (listed in Table 1.2) contributed helpful feedback and assistance in the early stages of drafting this report. Katherine Gettings, Nikola Osborne, and Sarah Riman provided valuable input on the text, including the data summaries used in Chapter 4. Jason Weixelbaum, Susan Ballou, Christina Reed, and Kathy Sharpless assisted with copy editing. Kathryn Miller from the NIST Library helped finalize the document for public release.

Public comment period: June 9, 2021 through August 9, 2021

The initial release of this report is a draft document, and we welcome comments and feedback from readers. All relevant submitted comments will be made publicly available and will be considered when finalizing this report. Do not include personal information, such as account numbers or Social Security numbers, or names of other individuals. Do not submit confidential business information, or otherwise proprietary, sensitive, or protected information. We will not post or consider comments that contain profanity, vulgarity, threats, or other inappropriate language or like content. During the 60-day comment period, comments may be sent to scientificfoundationreviews@nist.gov.

All comments, including commenter name and affiliation, will be published at https://www.nist.gov/dna-mixture-interpretation-nist-scientific-foundation-review.

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the authors or the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

National Institute of Standards and Technology
Attn: Special Programs Office – Scientific Foundation Review
100 Bureau Drive Stop 4701
Gaithersburg, MD 20899-4701

Email: scientificfoundationreviews@nist.gov

National Institute of Standards and Technology Interagency or Internal Report 8351-draft
Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8351-draft, 251 pages (June 2021)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351-draft

Preface

Forensic science plays a vital role in the criminal justice system by providing scientifically based information through the analysis of physical evidence. The National Institute of Standards and Technology (NIST) is a non-regulatory scientific research agency within the U.S. Department of Commerce with a mission to advance national measurement science, standards, and technology. NIST has been working to strengthen forensic science methods for almost a century. In recent years, several scientific advisory bodies have expressed the need for a review of the scientific bases of forensic methods and identified NIST as an appropriate agency for conducting them. A scientific foundation review, also referred to as a technical merit evaluation, is a study that documents and assesses the foundations of a scientific discipline, that is, the trusted and established knowledge that supports and underpins the discipline's methods. Congress has appropriated funds for NIST to conduct scientific foundation reviews in forensic science. These reviews seek to answer the question: "What established scientific laws and principles as well as empirical data exist to support the methods that forensic science practitioners use to analyze evidence?" Background information on NIST scientific foundation reviews is available in NISTIR 8225 at https://doi.org/10.6028/NIST.IR.8225.

Abstract

Improvements in DNA testing methods have allowed forensic scientists to reduce the quantity of DNA required for profiling an individual. Today, DNA profiles can be generated from a few skin cells. This increased sensitivity has extended the usefulness of DNA analysis into new areas of criminal activity beyond homicides and sexual assaults – but also the complex DNA mixtures often seen in casework. Distinguishing one person's DNA from another in these mixtures, estimating how many individuals contributed DNA, determining whether the DNA is even relevant or is from contamination, or whether there is a trace amount of suspect or victim DNA make DNA mixture interpretation inherently more challenging than examining single-source samples. These issues, if not properly considered and communicated, can lead to misunderstandings regarding the strength and relevance of the DNA evidence in a case.

This report explores DNA mixture interpretation with six chapters and two appendices. Chapter 1 introduces the topic of DNA mixtures, the difficulties behind their interpretations, and discusses the relevance of issues explored in the other chapters of this scientific foundation review. Chapter 2 provides background information on DNA and describes principles and practices underlying mixture measurement and interpretation. The likelihood ratio (LR) framework and probabilistic genotyping software (PGS) are also discussed. Chapter 3 lists data sources used in this study and strategies to locate them. Chapter 4 and Chapter 5 cover the report's core concepts: reliability and relevance issues in DNA mixture interpretation. Chapter 6 explores the potential of new technologies to assist mixture interpretation and considerations for implementation. The two appendices provide context on how the field has progressed (Appendix 1) and strategies to strengthen it going forward (Appendix 2). There are 528 references in the bibliography.

Keywords

activity level propositions
binary models
case assessment and interpretation
case context
cell separations
combined probability of inclusion
complex DNA mixture
contamination
continuous (fully continuous) models
discrete (semi-continuous) models
DNA
DNA mixture
DNA mixture interpretation
DNA transfer and persistence
forensic science
hierarchy of propositions
interlaboratory studies
internal validation studies
interpretation
likelihood ratio
massively parallel sequencing
measurement
microhaplotypes
next generation sequencing
peer-reviewed publications
principles
probabilistic genotyping
probabilistic genotyping software
proficiency tests
relevance
reliability assessment
receiver operating characteristic (ROC) curves
scientific foundation review
software reliability
technical merit evaluation
technology
validation studies

199 **Table of Contents**

200

List of Tables

List of Figures

Glossary and Acronyms

Allele: one of two or more versions of a genetic sequence; humans typically inherit one allele from each parent; however, sometimes three alleles, called tri-allelic patterns, are seen in STR analysis of a single-source DNA sample; genetic sequence at a particular location (a locus) in the genome alleles targeted in STR analysis can vary by sequence in addition to length

Allele drop-in: allele peak(s) in an electropherogram (EPG) that are not reproducible across multiple independent amplification events; also, a hypothesis/postulate for the observation of one or more allelic peaks in an electropherogram that are inconsistent with the assumed/known contributor(s) to a sample

Allele (or locus) drop-out: loss of allele (or both alleles) information from a DNA profile; failure of an otherwise amplifiable allele to produce a signal above the analytical threshold because the allele was not present, or was not present in sufficient quantity, in the aliquot that underwent polymerase chain reaction (PCR) amplification

Amplification: an increase in the number of copies of a specific DNA fragment; in forensic DNA testing laboratories, this refers to the use of the PCR technique to produce many more copies of DNA alleles at specific genetic loci

Artifact: any non-allelic product of the amplification process (e.g., a stutter product), an anomaly of the detection process, such as spectral pull-up, or a dye blob, which is by-product of primer synthesis, that may be observed in an electropherogram; may complicate interpretation of a DNA profile when they cannot be distinguished from actual allele(s) data

Bracketing approach: considers results from samples that are more complex or less complex than the casework sample of interest as a pragmatic way of understanding case-specific reliability of an interpretation system

Binary method: an interpretation scheme in which there are only two values (possible or not possible) for each decision (e.g., a peak is either "an allele" or "not an allele," or a genotype is "included" or "not included")

CE: capillary electrophoresis; an electrophoretic technique for separating DNA or other molecules by their size or charge based on migration through a narrow glass tube filled with a liquid polymer

Complex mixture: a DNA profile resulting from comingled DNA of two or more contributors that is difficult to interpret due to uncertainty in the determination of contributor genotypes; factors complicating mixture interpretation include, but are not limited to, low quantity DNA, low quality (degraded) DNA, the number of contributors, and the amount of allele sharing

Contamination: the transfer of irrelevant DNA during an investigation; inadvertent introduction of biological material including DNA alleles into a DNA sample at any stage from collection to testing; it is sometimes easy to identify but has the potential to mislead

Continuous approach: a statistical model and accompanying probabilistic genotyping method that evaluates DNA profiles using peak height information to assign weights to the observed peak heights for different combinations of contributor genotypes at all tested loci

CPI: combined probability of inclusion; the product of the probabilities of inclusion calculated for each locus; the probability of inclusion at each locus estimates the probability that a randomly selected, unrelated individual is not excluded from being one of the sources of DNA present in a mixture profile and is calculated as the square of the sum of the relative frequencies of the observed alleles at the locus; sometimes referred to as Random Man Not Excluded (RMNE); can only be appropriately used when all alleles from all contributors are present in the DNA profile

Deconvolution: separation of component DNA genotypes of contributors to a mixed DNA profile based on quantitative peak height information and any underlying assumptions (e.g., the number of contributors to the mixture, mixture ratios, or known contributors)

Discrete approach: a statistical model and accompanying probabilistic genotyping method that evaluates DNA profiles solely on the presence or absence of alleles without considering peak height information and utilizes probabilities of allele drop-out and drop-in

DNA: deoxyribonucleic acid

DNA mixture: sample that contains DNA from more than one individual

DNA mixture interpretation: an effort to (1) infer possible genotypes for detectable sample contributors (a process sometimes referred to as *deconvolution* of the mixture components) and (2) provide the strength of evidence for a person of interest being part of an evidentiary DNA profile

DNA profile: a string of values (numbers or letters) compiled from the results of DNA testing at one or more genetic markers (loci); can be single-source or a mixture from multiple contributors

EPG: electropherogram; graphic representation of the separation of molecules by electrophoresis in which data appear as "peaks" along a line; the format in which DNA typing results are presented with the horizontal axis displaying the observed peaks (which could be STR alleles or artifacts such as stutter products) in order of increasing size and the vertical axis recording the relative amount of DNA detected based on the fluorescent signal collected

Empirical (assessments/data/methods): information gathered by direct observation

Factor space and **factor space coverage:** the totality of scenarios and associated variables (factors) that are considered likely to occur in actual casework; with DNA mixture interpretation, factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested

Genotype: the variation in a DNA sequence that distinguishes one individual of a species, also described as the genetic constitution of an individual organism; the pair of alleles present at a tested STR locus

Ground truth: information provided by direct observation (i.e., empirical evidence) as opposed to information provided by inference; a situation where the correct answer is known by design

Interpretation: the process of giving meaning to findings; includes data and statistical analysis and usually produces an opinion on evidence examined

Known samples: DNA samples with known genotypes, used for validating methods and assessing proficiency

Locus (pl. Loci): a unique physical location of a gene (or a specific sequence of DNA in the case of STRs) on a chromosome; the plural form of locus is pronounced /LOW-sigh/

LR: likelihood ratio; the probability of the evidence under one proposition divided by the probability of the evidence under an alternative, mutually exclusive proposition; the magnitude of its value is commonly used to express a strength of the evidence based on the propositions proposed

Measurand: property intended to be measured

Measurement: an experimental or computational process that, by comparison with a standard, produces an estimate of the true value of a property of a material or virtual object or collection of objects, or of a process, event, or series of events, together with an evaluation of the uncertainty associated with that estimate and intended for use in support of decision-making

Microhaplotypes: regions of DNA containing two or more closely linked single nucleotide polymorphisms (SNPs) associated with multiple allelic combinations (haplotypes); these markers

have been explored for mixture deconvolution using massively parallel sequencing due to lack of stutter artifacts

Next generation sequencing: a high-throughput DNA sequencing technology where millions or billions of DNA strands can be sequenced in parallel; also called massively parallel sequencing

ng: nanogram; a billionth of a gram (10^{-9} g); there is 1 ng of DNA in ≈150 human cells

NIST: National Institute of Standards and Technology

PCR: polymerase chain reaction; an *in vitro* process that yields millions of copies of targeted DNA regions through repeated cycling of a biochemical reaction involving a DNA polymerase enzyme

pg: picogram; a trillionth of a gram (10^{-12} g); there are ≈6 pg of DNA in a single diploid human cell

PGS: probabilistic genotyping software; a computer program that utilizes statistical genetics, biological models, computer algorithms, and probability distributions to infer genotypes and assign likelihood ratios using either discrete or continuous approaches

Principles: fundamental, primary, or general scientific laws or truths from which others are derived

Proficiency test: a quality assurance measure used to monitor performance of a scientist and identify areas in which improvement may be needed; can be internal (produced by the agency undergoing the test) or external (produced by an outside test provider); external proficiency tests can be either open (where the scientist is aware the samples being tested are a proficiency test) or blind (where the scientist is unaware the samples being tested are a proficiency test)

Reliability: providing consistently accurate results

RFLP: restriction fragment length polymorphism; an analysis method used in early DNA testing

RFU: relative fluorescence unit; an arbitrary measure of the heights of peaks in an electropherogram

ROC curve: receiver operating characteristic curve; a graphical plot that examines the relationship between sensitivity (fraction of true positives) and specificity (fraction of false positives)

SRM: Standard Reference Material; a certified reference material supplied by NIST

Stochastic effects or variation: the observation of intra-locus peak imbalance and/or allele drop-out resulting from random, disproportionate amplification of alleles in low-quantity DNA samples; allele drop-in and elevated stutter product levels may also result

STR: short tandem repeat; an identical (or similar) DNA sequence arranged in direct succession where the repeat sequence unit is 2 base pairs (bp) to 6 bp in length; the number of repeat units varies among individuals

SWGDAM: Scientific Working Group on DNA Analysis Methods; formerly known as TWGDAM, Technical Working Group on DNA Analysis Methods; an FBI-sponsored group that develops quality assurance standards and guidelines for forensic DNA and DNA databasing laboratories in the United States and Canada

Uncertainty: the lack of certainty or sureness of an event; measurement uncertainty is the doubt about the true value of the measurand [property intended to be measured] that remains after making a measurement (see Possolo 2015)

Executive Summary

All scientific methods have limits. One must understand those limits to use a method appropriately. This is especially important in forensic science as critical decisions impacting life and liberty are often based on the results of forensic analysis.

Forensic DNA technology brings immense benefits to society, and new tools and techniques can increase those benefits further. But as new technologies are implemented with increased detection capabilities, we believe it is important to periodically assess their impacts on the scientific discipline. We do so in this scientific foundation review by identifying scientific principles, reviewing the scientific literature, gathering other empirical evidence from publicly available sources, and receiving input from a group of forensic DNA practitioners and researchers. This scientific foundation review explores what is known about the limits of DNA mixture interpretation methods, including probabilistic genotyping software systems.

As with any field, the scientific process (research, results, publication, additional research, etc.) continues to lead to advancements and better understanding. Information contained in this report comes from the authors' technical and scientific perspectives and review of information available to us during the time of our study. Where our findings identify opportunities for additional research and improvements to practices, we encourage researchers and practitioners to take action toward strengthening methods used to move the field forward. The findings described in this report are meant solely to inform future work in the field.

Improvements in DNA testing methods have allowed forensic scientists to reduce the quantity of DNA required for profiling an individual. In the 1990s, an evidence sample needed to contain thousands of cells, such as from a visible blood or semen stain. Today, analysts can extract a DNA profile from the few skin cells that someone might leave behind when handling an object.

This increased sensitivity extended the usefulness of DNA analysis into new areas of criminal activity beyond homicides and sexual assaults. DNA on bullets or cartridge casings can reveal clues to crimes involving firearms. Swabbing objects that a perpetrator might have handled can yield evidence in property crimes. Cold case evidence previously analyzed with less discriminating methods can be re-opened and researched again to find new insights. However, because people constantly shed small amounts of DNA into the environment, and by touching objects, people can potentially transfer small amounts of DNA from one surface to another, including someone else's DNA. Analyzing small quantities of DNA can create challenges in interpreting the data.

Highly sensitive methods, now universally used across the forensic DNA community, often detect DNA from more than one individual in a sample. But distinguishing one person's DNA from another in these mixtures, estimating how many individuals contributed DNA, determining whether the DNA is even relevant or is from contamination, or whether there is a trace amount of suspect or victim DNA make DNA mixtures inherently more challenging to interpret than single-source samples. These issues, if not properly considered and

communicated, can lead to misunderstandings regarding the strength and relevance of the DNA evidence in a case.

When laboratories analyze high-quality, single-source samples, decision-makers often have confidence in DNA test results in part because it has been demonstrated that different laboratories will arrive at the same result. This is true regardless of the specific instruments, kits, and software used. However, multiple interlaboratory studies conducted by different groups over the past two decades have demonstrated a wide range of variation in how specific *DNA mixtures* are interpreted.

This report is arranged into six chapters and two appendices. Chapter 1 introduces the topic of DNA mixtures (samples that contain DNA from more than one individual), the difficulties behind their interpretations, and the relevance of the issues explored in the other chapters of this scientific foundation review. Chapter 2 provides background information on DNA and describes principles and practices underlying mixture measurement and interpretation. The likelihood ratio (LR) framework and probabilistic genotyping software (PGS) are also discussed. Chapter 3 lists data sources used in this study and strategies to locate them. Chapters 4 and 5 cover the report's core concepts: reliability and relevance issues in DNA mixture interpretation. Chapter 6 explores the potential of new technologies to assist mixture interpretation and considerations for implementation. The two appendices provide context on how the field has progressed and strategies to strengthen it going forward. Appendix 1 presents the history of DNA mixture interpretation, while Appendix 2 considers various perspectives on training and continuing education.

A DNA Mixture Resource Group (see Table 1.2), with extensive experience in public and private forensic DNA laboratories, reviewed an early draft of our report and provided valuable feedback, insights, and suggestions. However, they were not asked to sign off on our final report or endorse its conclusions. The NIST team is grateful for their dedication and contributions to our efforts.

Chapter 1: Introduction

New tools and techniques for analyzing and interpreting minor contributors to DNA mixtures are now routinely employed in everyday casework in the United States and around the world. These tools include DNA profiling kits, genetic analyzer instruments, and probabilistic genotyping software.

DNA mixtures can be partly understood by analogy to latent print examination. If multiple fingerprints are deposited on top of one another, it would be difficult to tease apart the individual fingerprints because it may not be clear which ridge lines belong to which print. In a DNA mixture it may not be clear which genetic components, called alleles, belong to which contributor. Interpreting the mixture requires an assessment of which alleles go together to form the DNA profiles of the individual contributors.

Forensic scientists interpret DNA mixtures with the assistance of statistical models and expert judgment. Interpretation becomes more complicated when contributors to the mixture

share common alleles. Complications can also arise when random variations, also known as stochastic effects, make it more difficult to confidently interpret the resulting DNA profile.

Not all DNA mixtures present these types of challenges. We agree with the President's Council of Advisors on Science and Technology (PCAST) that "DNA analysis of single-source samples or simple mixtures of two individuals, such as from many rape kits, is an objective method that has been established to be foundationally valid" (PCAST 2016). Therefore, this scientific foundation review does not concentrate on interpretation of single-source DNA samples and two-person mixtures involving significant quantities of DNA from both contributors.

Instead, this review focuses on methods for interpreting data from complex DNA mixtures, which we define as samples that contain comingled DNA from two or more contributors in which stochastic effects or allele sharing cause uncertainty in determining contributor genotypes. The following factors contribute to increased complexity (see also Chapter 2):

- Number of contributors and the degree of overlapping alleles
- Low-quantity DNA from one or more minor contributors
- Degree of degradation or inhibition of the DNA sample.

It is important that users of forensic DNA test results understand that DNA evidence can vary greatly in complexity based on these factors, and that more complex samples involve greater uncertainty.

Chapter 2: DNA Mixture Interpretation: Principles and Practices

Successful analysis and interpretation of DNA results depends on crime scene evidence (the "Q" or questioned sample) being of suitable quality and quantity, and the availability of a reference sample (the "K" or known sample). When appropriate Q and K DNA profiles are available, forensic scientists can perform a Q-to-K comparison and report a likelihood ratio (LR) that is an evaluative interpretation of the strength of this association using specific assumptions and usually one of several statistical approaches. In testing forensic casework samples, a range of DNA profile qualities and quantities can exist. DNA mixtures are inherently more difficult to interpret than single-source DNA samples.

The process of DNA evidence analysis can be divided into two major steps: (1) *measurements* of relative abundances of polymerase chain reaction (PCR) products in a tested DNA sample that are displayed as an electropherogram (EPG), and (2) *interpretation* involving use of the EPG data to make a strength-of-evidence assessment when an evidentiary DNA profile is compared to a person of interest (POI). The outcome of interpretation includes an LR number that can range in value depending on the analyst's assumptions, protocols, algorithms, tools, and other variables. There remains a need to assess the fitness for purpose of an analyst's LR using empirical methods.

Forensic scientists interpret DNA mixtures with the assistance of statistical models and expert judgment. Interpretation becomes more complicated when contributors to the mixture

share common alleles. Complications can also arise when reduced DNA template amounts are used in PCR, where random sampling, also known as stochastic effects, makes it more difficult to confidently interpret the resulting DNA profile.

This chapter describes 16 principles and includes 6 key takeaways.

KEY TAKEAWAY #2.1: DNA mixtures, where the DNA of more than one individual is present in a sample, are inherently more difficult to interpret than single-source DNA samples.

KEY TAKEAWAY #2.2: Generating a DNA profile involves measuring the inherent physical properties of the sample. Interpreting a DNA profile involves assigning values that are not inherent to the sample. To do this, the DNA analyst uses their judgment, training, tools (including computer software), and experience, and considers factors such as case context.

KEY TAKEAWAY #2.3: The process of generating a DNA profile can produce stochastic or random variation and artifacts that contribute to the challenge of DNA mixture interpretation.

KEY TAKEAWAY #2.4: DNA mixtures vary in complexity, and the more complex the sample, the greater the uncertainty surrounding interpretation. Factors that contribute to complexity include the number of contributors, the quantity of DNA from each contributor, contributor mixture ratios, sample quality, and the degree of allele sharing.

KEY TAKEAWAY #2.5: Continuous probabilistic genotyping software (PGS) methods utilize more information from a DNA profile than binary approaches.

KEY TAKEAWAY #2.6: Likelihood ratios are not measurements. There is no single, correct likelihood ratio (LR). Different individuals and/or PGS systems often assign different LR values when presented with the same evidence because they base their judgment on different kits, protocols, models, assumptions, or computational algorithms. Empirical data for assessing the fitness for purpose of an analyst's LR are therefore warranted.

Chapter 3: Data and Information Sources

This chapter contains sources of data and information used in conducting this review along with strategies to locate them. These sources include (1) peer-reviewed articles appearing in scientific journals, (2) published interlaboratory studies, (3) laboratory internal validation study summaries that are accessible online, and (4) proficiency test data available on test provider websites.

Chapter 4: Reliability of DNA Mixture Measurements and Interpretation

In this report, we divide the challenges presented by DNA mixtures into two main categories. The first involves the *reliability* of mixture interpretation methods when used with DNA evidence of varying complexity. (Chapter 5 deals with the second challenge: *relevance*.) In this report, we use the "plain English" definition of reliability as a measure of trustworthiness. A highly reliable method is one that consistently produces accurate results. Reliability is not a yes or no question, but a matter of degree. Understanding the degree of reliability of a method can help the user of that information decide whether they should trust the results of that method when making important decisions.

This chapter considers foundational issues related to reliability of DNA mixture interpretation. Reliability centers on trustworthiness established through empirical assessments of available data to evaluate the degree of reliability of a system or its components. We use the term "factor space" to describe the factors that influence complexity, measurement, and interpretation reliability – these factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested.

We note that the degree of reliability of a DNA mixture interpretation system, such as a DNA analyst using a probabilistic genotyping software program, depends on sample complexity. Results cannot be simply categorized as "reliable" or "unreliable" without considering context. In addition, reliability cannot be established without validation tests using known samples of similar complexity. The results of such tests provide data that are considered accurate and reliable; only with such valid results can comparisons be made as to the reliability of unknown casework samples. We also emphasize that samples used in proficiency tests need to be representative of complex DNA mixtures seen in casework if these tests are intended to assess analysts' ability to conduct dependable DNA mixture interpretation.

Finally, the theme of reliability is discussed throughout this report. Note that our original goal in this review was *external* and *independent* assessment of reliability based on publicly available data that met our selection criteria. These criteria evolved during this study as we became aware of the amount and type of data available to us. Laboratories and researchers may make claims or have their own understanding of reliability as it relates to their own work, but our findings are defined by the public information available at the time of this report.

This chapter includes eight key takeaways.

KEY TAKEAWAY #4.1: The degree of reliability of a component or a system can be assessed using empirical data (when available) obtained through validation studies, interlaboratory studies, and proficiency tests.

KEY TAKEAWAY #4.2: To enable effective use of any information, responsibilities exist with both providers and users of that information. While a provider explains the relevance and significance of the information and data, only the user can assess the degree of reliability, validity, and whether that information is fit-for-purpose.

KEY TAKEAWAY #4.3: Currently, there is not enough publicly available data to enable an external and independent assessment of the degree of reliability of DNA mixture interpretation practices, including the use of probabilistic genotyping software (PGS) systems. To allow for external and independent assessments of reliability going forward, we encourage forensic laboratories to make their underlying PGS validation data publicly available and to regularly participate in interlaboratory studies.

KEY TAKEAWAY #4.4: Additional PGS validation studies have been published since the 2016 PCAST Report. However, publicly available information continues to lack sufficient details needed to independently assess reliability of specific LR values produced in PGS systems for complex DNA mixture interpretation. Even when a comparable reliability can be assessed (results for a two-person mixed sample are generally expected to be more reliable than those for a four-person mixed sample, for example), there is no threshold or criteria established to determine what is an acceptable level of reliability.

KEY TAKEAWAY #4.5: Current proficiency tests are focused on single-source samples and simple two-person mixtures with large quantities of DNA. To appropriately assess the ability of analysts to interpret complex DNA mixtures, proficiency tests should evolve to address mixtures with low-template components or more than two contributors – samples of the type often seen in modern casework.

KEY TAKEAWAY #4.6: Different analysts and different laboratories will have different approaches to interpreting the same DNA mixture. This introduces variability and uncertainty in DNA mixture interpretation. Improvements across the entire community are expected with an increased understanding of the causes of variability among laboratories and analysts.

KEY TAKEAWAY #4.7: The degree of reliability of a PGS system when interpreting a DNA mixture can be judged based on validation studies using known samples that are similar in complexity to the sample in the case. To enable users of results to assess the degree of reliability in the case of interest, it would be helpful to include these validation performance results in the case file and report.

KEY TAKEAWAY #4.8: We encourage a separate scientific foundation review on the topic of likelihood ratios in forensic science and how LRs are calculated, understood, and communicated.

Chapter 5: Context and Relevance Related to DNA Mixture Interpretation

The second major challenge posed by DNA mixtures involves the *relevance* of a DNA sample to the crime being investigated. The question of relevance arises because DNA can be transferred between surfaces, potentially more than once. This means that some of the DNA present at a crime scene may be irrelevant to the crime, and current DNA profiling methods increase the likelihood of detecting more DNA. Similarly, today's highly sensitive DNA

methods increase the risk that very small amounts of contamination might affect DNA test results.

Chapter 5 focuses on questions of context and relevance: How and when was the DNA deposited, and is that DNA relevant to the crime being investigated?

The question of relevance arises because people readily shed DNA into the environment, and they can potentially transfer DNA between surfaces when touching objects or other people. Therefore, the DNA present at a crime scene or on a piece of evidence may be irrelevant to any crime. To assess relevance, in addition to knowing specific details of the case, one would need information on what factors make DNA more or less likely to transfer and to persist in the environment. This chapter reviews the scientific literature on DNA transfer and persistence and presents strategies for assessing DNA relevance.

The fact that DNA can be transferred between surfaces upon contact is a foundational principle of forensic DNA analysis. This is what makes the discipline useful for investigating crimes in the first place. This has several implications for DNA found at a crime scene. First, that DNA might have been deposited before or after the crime was committed and therefore may not be relevant to the crime. Second, the DNA might have been deposited via secondary transfer, which occurs when DNA is picked up for one surface and deposited on another. For instance, a person might pick up DNA from a second person during a handshake, then deposit the second person's DNA onto an item or surface.

These possibilities mean that the presence of a person's DNA in an evidence sample does not necessarily mean that the DNA is relevant to the crime. Relevance should be assessed. If not, the evidence can be misleading.

By definition, highly sensitive methods are more likely to detect small quantities of DNA, including background DNA that may be present in the environment. In addition, highly sensitive methods are more likely to detect DNA mixtures, which by their nature usually include irrelevant DNA. Therefore, when assessing evidence that involves very small quantities of DNA, it is especially important to carefully consider relevance.

This report uses the word contamination to describe the transfer of irrelevant DNA during an investigation. For example, a fingerprint brush can potentially transfer minute amounts of DNA onto evidence at a crime scene. Such a small amount of DNA might have gone undetected in the past, but highly sensitive methods increase the likelihood that it might now be detected. This increases the likelihood that contamination might affect an investigation.

Forensic laboratories have been using procedures to avoid contamination since the advent of DNA methods. However, because the likelihood of detecting contaminating DNA has increased with the development of highly sensitive DNA methods, contamination avoidance in forensic laboratories is more important than ever. Furthermore, contamination avoidance procedures should be used during all stages of an investigation, including at the crime scene. Elimination databases that include DNA profiles of laboratory staff and police who go to crime scenes can help identify contamination and should be maintained.

Many interpretation methods, including probabilistic genotyping, address questions about who might have contributed DNA to a crime scene profile and express the strength of evidence in the form of a likelihood ratio. This statistic does not provide any information about how much DNA was present, or how or when the DNA was deposited. For instance, a large blood stain might produce a very similar likelihood ratio to a swab from a light switch, yet the two types of evidence might vary greatly in terms of their evidential value. Therefore, the likelihood ratio should not be used in isolation. It is imperative that the likelihood ratio be considered in the context of other evidence in the case.

The fact that DNA can transfer does not mean that DNA is useless as evidence. To the contrary, this is what makes DNA useful to criminal investigations in the first place. However, the possibility of DNA transfer may raise questions of relevance that need to be addressed, especially in cases that involve very small amounts of DNA. These questions can be addressed by considering DNA evidence in the context of case circumstances, including other evidence in the case.

More research is needed on DNA transfer and persistence. In addition, to make use of the studies that are available, individual laboratories would need to know how the sensitivity of methods used in their laboratory compares to the sensitivity of methods employed in the studies being considered.

This chapter includes six key takeaways.

KEY TAKEAWAY #5.1: DNA can be transferred from one surface or person to another, and this can potentially happen multiple times. Therefore, the DNA present on an evidence item may be unrelated (irrelevant) to the crime being investigated.

KEY TAKEAWAY #5.2: Highly sensitive DNA methods increase the likelihood of detecting irrelevant DNA. When assessing evidence that involves very small quantities of DNA, it is especially important to consider relevance.

KEY TAKEAWAY #5.3: Highly sensitive methods increase the likelihood of detecting contaminating DNA that might affect an investigation. Contamination avoidance procedures should be robust both at the crime scene and in the laboratory.

KEY TAKEAWAY #5.4: DNA statistical results such as a sub-source likelihood ratio do not provide information about how or when DNA was transferred, or whether it is relevant to a case. Therefore, using the likelihood ratio as a standalone number without context can be misleading.

KEY TAKEAWAY #5.5: The fact that DNA transfers easily between objects does not negate the value of DNA evidence. However, the value of DNA evidence depends on the circumstances of the case.

KEY TAKEAWAY #5.6: There is a growing body of knowledge about DNA transfer and persistence, but significant knowledge gaps remain.

Chapter 6: New Technologies: Potential and Limitations

New technologies are often investigated to assess whether they can provide solutions to existing problems in the forensic community. The adoption and implementation of these technologies depends upon a cost/benefit analysis within forensic laboratories. Appreciating fundamental challenges with DNA mixture interpretation can assist in considering whether new approaches can bring desired improvements to mixture interpretation.

The ability to analyze short tandem repeat alleles by sequence in addition to length promises to bring some new capabilities to forensic DNA laboratories, including the potential for improvements in DNA mixture interpretation. Next-generation sequencing platforms also enable additional genetic markers to be examined, some of which, such as microhaplotypes, have been pursued with the potential to improve DNA mixture interpretation. Additionally, cell separation techniques offer the potential to separate contributors prior to DNA extraction.

The ultimate decision to implement new technologies in forensic laboratories should be driven by a real-use case and by those responsible for producing and reporting the information. A vendor or members of the general public may encourage forensic DNA laboratories to adopt a new approach or technology without appreciating investments required to make a change. Consideration should be given to whether supporting factors and resources will be available upon implementation (e.g., allele frequencies, analysis software, interpretation methods, training, and support for potential admissibility hearings). An overall assessment of 1) how a new technology works, 2) what its limitations are, and 3) how it might specifically address the problem to be solved (e.g., DNA mixture interpretation) is important and a key component of evaluating whether implementation will be worthwhile.

This chapter includes two key takeaways.

KEY TAKEAWAY #6.1: Fundamental measurement and interpretation issues surrounding DNA mixtures, as described in Chapter 2, should be understood before attempting to apply a new technology.

KEY TAKEAWAY #6.2: Implementation requires a thorough understanding of the benefits and limitations of the new technology as well as the practical investment of time and effort put forth for its adoption by the laboratory.

1. Chapter 1: Introduction

All scientific methods have limitations. One must understand those limitations to use a method appropriately. This is especially important in forensic science as critical decisions impacting life and liberty are often based on the results of forensic analysis. This scientific foundation review explores what is known about the limitations of DNA mixture interpretation methods, including probabilistic genotyping software systems, by reviewing the scientific literature and other sources of information.

1.1. Advances in Forensic DNA

The field of forensic DNA analysis is constantly advancing. One important change involves the ability to detect and analyze very small quantities of DNA (Butler 2012, Butler 2015a). During the early decades of forensic DNA analysis, an evidence sample containing thousands of cells, such as a visible blood or semen stain, was needed to produce a DNA profile. Today, analysts can extract a DNA profile from the few skin cells that someone might leave behind when handling an object.

This increased sensitivity extends the usefulness of DNA analysis into new areas of criminal activity beyond the homicides and sexual assaults that were once the primary focus. Crimes involving firearms can be investigated by testing for DNA on bullets or cartridge casings (e.g., Montpetit & O'Donnell 2015). Property crimes can be investigated by swabbing objects that a perpetrator might have handled (Mapes et al. 2016). Cold cases that were previously analyzed with less discriminating methods can yield more useful evidence.

However, people constantly shed small amounts of DNA into the environment, and by touching objects, people can potentially transfer small amounts of DNA – including someone else's DNA – from one surface to another. Analyzing small quantities of DNA can create challenges when interpreting the data. Highly sensitive methods, now universally used across the forensic DNA community (Gill et al. 2015), often detect DNA from more than one individual in sample. Analysts know they are dealing with what is called a DNA mixture when they detect more than two alleles at multiple locations in a DNA profile. Because humans typically inherit one allele from each parent for every gene, finding more than two is one indication that more than one genotype, the variation in a DNA sequence that is unique to an individual organism, may be in the sample. As a result, more than one genotype combination may be possible at each tested location in the DNA sequence.

Distinguishing one person's DNA from another in these mixtures, estimating how many individuals contributed DNA, determining whether the DNA is even relevant or is from contamination, or whether there is a trace amount of suspect or victim DNA make DNA mixtures inherently more challenging to interpret than single-source samples. These issues, if not properly considered and communicated, can lead to misunderstanding the strength and relevance of the DNA evidence in a case.

The ability to detect small amounts of DNA has been improving for decades (Butler 2012, Butler 2015a). When forensic DNA analysis was first introduced in the mid-1980s (Gill et al.

1985), a stain about the size of a quarter was needed to generate a DNA profile. In the early 1990s, forensic laboratories started using polymerase chain reaction (PCR), a method that leverages the natural tendency of DNA to produce copies of itself, to amplify DNA. This method allowed the analysis of much smaller amounts of starting material (e.g., Saiki et al. 1989, Blake et al. 1992), though a visible stain was still generally needed. In 1997, scientists demonstrated high-sensitivity methods that allowed for recovery of DNA information from touched objects (van Oorschot & Jones 1997) and even from single cells (Findlay et al. 1997).

Highly sensitive methods began moving from research centers into crime laboratories more than ten years ago, but the application of such methods to detect minor contributors in DNA mixtures has increased rapidly in recent years. New tools and techniques for analyzing and interpreting minor contributors to DNA mixtures are now routinely employed in everyday casework in the United States and around the world (Butler 2015b, Gill et al. 2015). These tools include DNA profiling kits, genetic analyzer instruments, and probabilistic genotyping software (PGS).

Forensic DNA technology brings immense benefits to society, and these new tools and techniques can increase those benefits further. But as new technologies are implemented with increased detection capabilities, we believe it is important to periodically assess the impacts on the scientific discipline. We do so in this scientific foundation review by identifying scientific principles, reviewing the scientific literature, gathering other empirical evidence from unpublished sources, and collecting input from a group of leading forensic DNA practitioners and researchers.

As with any field, the scientific process (research, results, publication, additional research, etc.) continues to lead to advancements and better understanding. Information contained in this report comes from the authors' technical and scientific perspectives and review of information available to us during the time of our study. Where our findings identify opportunities for additional research and improvements to practices, we encourage researchers and practitioners to take action toward strengthening methods used to move the field forward. The findings described in this report are meant solely to inform future work in the field.

1.2. DNA Mixtures Vary in Complexity

DNA mixtures can be partly understood by analogy to latent print examination. If multiple fingerprints are deposited on top of one another, it would be difficult to tease apart the individual fingerprints because it may not be clear which ridge lines belong to which print. In a DNA mixture it may not be clear which genetic components, called alleles, belong to which contributor. Interpreting the mixture requires an assessment of which alleles go together to form the DNA profiles of the individual contributors.

Forensic scientists interpret DNA mixtures with the assistance of statistical models and expert judgment. Interpretation becomes more complicated when contributors to the mixture share common alleles (e.g., Clayton et al. 1998). Complications can also arise when reduced

DNA template amounts are used in PCR, where random sampling, also known as stochastic effects, make it more difficult to confidently interpret the resulting DNA profile (e.g., Gill et al. 2000).

Not all DNA mixtures present these types of challenges. We agree with the President's Council of Advisors on Science and Technology (PCAST) that "DNA analysis of single-source samples or simple mixtures of two individuals, such as from many rape kits, is an objective method that has been established to be foundationally valid" (PCAST 2016). Therefore, this scientific foundation review does not emphasize interpretation of single-source DNA samples and two-person mixtures involving significant quantities of DNA from both contributors. Instead, this review focuses on methods for interpreting data from complex DNA mixtures, which we define as samples that contain comingled DNA from two or more contributors in which stochastic effects or allele sharing cause uncertainty in determining contributor genotypes. The following factors contribute to increased complexity (see Chapter 2):

- Number of contributors and the degree of overlapping alleles
- Low-quantity DNA from one or more minor contributors
- Degree of degradation or inhibition of the DNA sample.

It is important that users of forensic DNA test results understand that DNA evidence can vary greatly in complexity based on these factors, and that more complex samples involve greater uncertainty.

1.3. Reliability

In this report, we divide the challenges presented by DNA mixtures into two main categories. The first involves the *reliability* of mixture interpretation methods when used with DNA evidence of varying complexity. In this report, we use the "plain English" definition of reliability as a measure of trustworthiness. A highly reliable method is one that consistently produces accurate results. Reliability is not a yes or no question, but a matter of degree. Understanding the degree of reliability of a method can help the user of that information decide whether they should trust the results of that method when making important decisions. In addition, the degree of reliability of a method can often be demonstrated with empirical data.

We address reliability issues by surveying available validation studies, which are meant to demonstrate how a method performs under defined sets of circumstances (e.g., varying numbers of contributors, template amounts, mixture ratios). We also consider interlaboratory studies, which provide information on the variability in test results across laboratories, and we review standards and guidelines for mixture interpretation.

In addition, we briefly discuss performance assessments that are frequently used in other sectors, such as receiver operating characteristic (ROC) curves (Green & Swets 1966, Bleka et al. 2016b) and calibration of likelihood ratios (Zadora et al. 2014). When sufficient data are available, these assessments can be used to evaluate the reliability of DNA mixture

interpretation methods and compare reliability across different PGS systems (e.g., Bleka et al. 2016b, You & Balding 2019). Laboratories might also use these assessments to set operational limits based on their validation studies.

1.4. Relevance

The second major challenge posed by DNA mixtures involves the *relevance* of a DNA sample to the crime being investigated. The question of relevance arises because DNA can be transferred between surfaces, potentially more than once (van Oorschot et al. 2019). This means that some of the DNA present at a crime scene may be irrelevant to the crime, and current DNA profiling methods increase the likelihood of detecting more DNA. Similarly, today's highly sensitive DNA methods increase the risk that very small amounts of contamination might affect DNA test results (Fonneløp et al. 2016, Szkuta et al. 2015a).

This report uses the word contamination to describe the transfer of irrelevant DNA during an investigation. For example, a fingerprint brush can potentially transfer minute amounts of DNA onto evidence at a crime scene. Such a small amount of DNA might have gone undetected in the past, but highly sensitive methods increase the likelihood that it might now be detected. This increases the likelihood that contamination might affect an investigation.

Forensic laboratories have been using procedures to avoid contamination since the advent of DNA methods. However, because the likelihood of detecting contaminating DNA has increased with highly sensitive DNA methods, contamination avoidance in forensic laboratories is more important than ever. Furthermore, contamination avoidance procedures should be used during all stages of an investigation, including at the crime scene. Elimination databases that include DNA profiles of laboratory staff and police who go to crime scenes can help identify contamination and should be maintained. Therefore, relevance should be carefully assessed and considered by both the DNA analyst and users of the DNA results, especially when an evidence item contains very small amounts of DNA.

In this report, we address relevance issues by surveying the existing literature on DNA transfer and persistence, identifying what is known about these phenomena, and highlighting knowledge gaps. We discuss several ways in which DNA transfer might mislead an investigation if DNA evidence is not considered in the context of the facts and evidence in the case. We also suggest strategies for mitigating the risks presented by DNA transfer.

Mixture interpretation methods address questions about source of a DNA sample (i.e., who the DNA came from) and provide statistical strength of evidence such as a likelihood ratio. The interpretation of a DNA profile can be useful by itself for generating leads in an investigation. However, the investigator or the trier of fact should consider not just the source of the DNA, but also what activity might have caused the DNA to be deposited as evidence (Gill et al. 2018, Taylor et al. 2018). Answering questions about activity generally requires consideration of contextual information, including other evidence in the case (Gill et al. 2020a). In Chapter 5, we argue that uncertainties about activity are usually much greater than uncertainties about source (e.g., Taylor et al. 2018), and it is therefore critical to consider DNA evidence in context. Focusing only on a statistic without considering context can be

misleading. This is especially so in cases involving very small quantities of DNA, such as when touch samples are collected from a store counter or from a firearm that many people may have handled.

1.5. Why Conduct This Scientific Foundation Review?

As described in our earlier publication (NISTIR 8225), a scientific foundation review is "a study that seeks to document and evaluate the foundations of a scientific discipline, that is, the trusted and established knowledge that supports and underpins the discipline's methods. These reviews seek to answer the question: 'What empirical data exist that speak to the reliability of the methods that forensic science practitioners use to analyze crime scene material?'"

Such a review can help identify knowledge gaps and provide guidance for future research. In addition, documenting foundational studies and core principles in a written report can assist laboratories in identifying appropriate limits for interpretation and contribute to the training of forensic practitioners. This report can also help investigators, officers of the court, and other users of forensic science to consider DNA test results in context and with awareness of their limitations so they can make informed decisions.

There is abundant forensic DNA testing literature to be explored due to the large number of active researchers and a history of publishing that surpasses many other forensic disciplines. Thousands of articles pertaining to forensic DNA methods have been published in dozens of peer-reviewed scientific journals in the past 35 years. Similar review studies have been performed by other groups on forensic disciplines like fire investigations (Almirall et al. 2017) and latent fingerprints (Thompson et al. 2017). However, DNA mixture interpretation has not been explored in the same way.

When laboratories analyze high-quality, single-source samples, decision-makers often have confidence in DNA test results in part because it has been demonstrated that different laboratories will arrive at the same result; that is, obtain the same DNA profile at the tested loci. This is regardless of the specific instruments, kits, and software used. However, multiple interlaboratory studies conducted by different groups over the past two decades have demonstrated a wide range of variation in how specific DNA *mixtures* are interpreted (Duewer et al. 2001, Crespillo et al. 2014, Benschop et al. 2017a, Barrio et al. 2018, Butler et al. 2018a). A scientific foundation review might shed light on the sources of variability observed.

1.6. Limitations of This Study

First, forensic genetics is an evolving field, and this study can only provide a snapshot of the state of the science at a particular moment in time. Therefore, the literature and empirical evidence we discuss in this review will be incomplete as soon as it is published, as is the case with evidence reviews in other evolving fields such as medicine and public health.

Second, the data available for conducting this review were limited. For instance, most laboratories do not publish data from their validation studies. We find merit in the perspective that "Dissemination is a critical part of the scientific process because it exposes our work to peer review and allows scientists to build upon the contributions of others. A study isn't complete until it's been published" (Martire & Kemp 2018). In addition, many published developmental validation studies do not include enough data for an independent assessment of performance. We believe that greater transparency through forensic laboratories openly sharing their supporting validation data, along with an independent review, would help strengthen the field of forensic DNA analysis.

Third, we may not have succeeded in identifying all of the studies relevant to our research objectives. We welcome suggestions, during the public comment period on the initial draft (see below), for additional publicly available studies that should be included in our analysis.

Again, we note that the findings of this report are meant to inform future work in the field.

1.7. NIST Review Team

The review team consisted of six individuals from the National Institute of Standards and Technology (NIST) whose diverse expertise allowed us to examine issues from many perspectives and to use lessons learned in other fields. Table 1.1 lists members of the review team, their NIST operating unit, and their expertise. Our team met regularly between September 2017 and July 2020 while conducting this review and developing the content of this report. Assistance in finalizing this report was also provided by several additional NIST employees or contractors as noted in the Acknowledgments.

Table 1.1. Members of the NIST review team and their areas of expertise.

Name	NIST Operating Unit	Areas of Expertise
John M. Butler	Special Programs Office	Forensic DNA methods and scientific literature
Hari K. Iyer	Statistical Engineering Division, Information Technology Laboratory	Mathematics and statistics
Rich Press	Public Affairs Office	Communication and science writing
Melissa K. Taylor	Special Programs Office	Human factors (previous efforts in latent fingerprints and handwriting analysis)
Peter M. Vallone	Applied Genetics Group, Material Measurement Laboratory	DNA technology, research, rapid DNA analysis, next-generation DNA sequencing
Sheila Willis	Special Programs Office (hired under contract as an International Research Associate)	Forensic laboratory management and trace evidence (retired director of Forensic Science Ireland)

1.8. DNA Mixture Resource Group

The NIST review team met regularly with a group of outside experts, the DNA Mixture Resource Group (Resource Group), which provided input and feedback that were vital to keeping this project focused on critical and relevant issues.

The Resource Group (Table 1.2) provided important perspectives based on their extensive experience in public and private forensic laboratories. This group included nine active practitioners, including five DNA technical leaders, from federal, state, and local jurisdictions in the United States and Canada, and four leading academics and consultants who have published in the forensic DNA literature.

The Resource Group reviewed an early draft of this report and provided valuable feedback, insights, and suggestions during its development. However, they were not asked to provide consensus advice or recommendations, sign off on our final report, or endorse its conclusions. The NIST team is grateful for their dedication and contributions to our efforts.

Table 1.2. Members of the DNA Mixture Resource Group.

Name	Affiliation
Jack Ballantyne	Professor of Chemistry, University of Central Florida
Todd Bille	Alcohol, Tobacco, Firearms, and Explosives (ATF) Laboratory, DNA Technical Leader
Jennifer Breaux	Montgomery County (MD) Police Crime Laboratory, DNA Technical Leader
Robin Cotton	Boston University School of Medicine (and former laboratory director of Cellmark Diagnostics)
Roger Frappier	Centre of Forensic Sciences (Toronto, Canada)
Bruce Heidebrecht	Maryland State Police, DNA Technical Leader
Keith Inman	California State University East Bay and Forensic DNA Consultant
Eugene Lien	New York City Office of Chief Medical Examiner, Department of Forensic Biology, DNA Technical Leader
Tamyra Moretti	Federal Bureau of Investigation Laboratory, DNA Support Unit
Lisa Schiermeier-Wood	Virginia Department of Forensic Sciences, DNA Supervisor

Name	Affiliation
Joel Sutton	Defense Forensic Science Center, U.S. Army Criminal Investigation Laboratory, DNA Technical Leader
Ray Wickenheiser	New York State Police Laboratory Director (and president of the American Society of Crime Laboratory Directors, 2017–2018)
Charlotte Word	Independent Forensic DNA Consultant (and former laboratory director at Cellmark Diagnostics)

We requested input from the Resource Group to: (1) make sure we were addressing real-world problems faced by the community, (2) help define the scope and direction of our project, and (3) provide a sounding board for communications before sharing them with a wider community. This included a review of an early version of our report to ensure that the document was appropriate and helpful. The group met with the NIST team eight times in person and four times by teleconference over an 18-month period (December 2017 to June 2019).

Prior to our first meeting in December 2017, two questions were asked of the invited attendees to serve as a starting point: (1) What is your main concern in DNA mixture analysis today? (2) Where is there room for improvement in DNA testing?

Responses regarding concerns in DNA mixture analysis centered around the following areas, which are listed in no particular order:

- *Defining interpretation limits* so analysts know when to stop attempting to interpret a mixture, especially when only low-level data are available and when it is difficult to differentiate stutter from true alleles of another donor;
- *Delineating interpretation accuracy and reliable use of probabilistic genotyping software* (PGS) and ascertaining whether or not laboratories are adopting new approaches with proper foundation and training needed to create new interpretation protocols;
- *Estimating the number of contributors* and establishing a cutoff for mixtures in terms of the number of contributors that can reliably be distinguished in a particular case;
- *Addressing report writing and content,* including the difficulties of communicating results to law enforcement or attorneys;
- *Recognizing the need to increase consistency/reproducibility in interpretation and report writing* in some cases, within laboratories and across the community; and
- *Acknowledging the need to increase the scope of validation studies particularly for PGS systems* and in subsequent interpretation protocols to more accurately represent the meaning and value of DNA mixture results to law enforcement, attorneys, judges, and juries.

Responses to the question about room for improvement expressed a need for:

- *Standards with "teeth"* (impact or real influence), rather than general guidelines;

- *More publication and dissemination of results* to the community, along with tools to improve;
- *More consistent training* that helps the analyst improve DNA mixture interpretation, as opposed to presentations on research projects that are years away from implementation;
- *More information on validation and implementation* of PGS tools, with training that is hands-on, interactive, and involves critical thinking exercises;
- *Improved understanding of secondary transfer possibilities;* and
- *More training and continuing education* for analysts and stakeholders.

1.9. Informing Stakeholders

While conducting this scientific foundation review, the authors made several presentations to a wide range of stakeholders, including DNA analysts, technical leaders, academic researchers, students, prosecutors, defense attorneys, and judges. These public presentations enabled the NIST team to keep members of these communities informed about plans and progress being made as well as to receive input. This included suggested topics for consideration and articles to add to the literature review.

After the first public presentation regarding this scientific foundation review at the January 2018 SWGDAM meeting, copies of slides and a draft reference list were provided to all known probabilistic genotyping software vendors or developers. Progress made after the first year was summarized in the *Proceedings of the 29th International Symposium on Human Identification* titled "DNA Mixture Interpretation Principles: Insights from the NIST Scientific Foundation Review" (Butler et al. 2018b). Progress after the second year was reported at the 2019 Congress of the International Society for Forensic Genetics (ISFG) (Butler et al. 2019).

Two of the NIST team members prepared an INTERPOL literature review covering forensic DNA articles published between 2016 and 2019, which included information on PGS and DNA mixture interpretation (Butler & Willis 2020). This effort also involved a presentation at the INTERPOL International Forensic Science Managers Symposium in October 2019.

Approximately 120 people attended a full-day workshop held in February 2019 at the American Academy of Forensic Sciences (AAFS) meeting in Baltimore, Maryland. This workshop, titled "DNA Mixture Interpretation Principles: Observations from a NIST Scientific Foundation Review," provided a detailed progress report of our findings and insights from Resource Group members about their experiences participating in the NIST review. A total of 19 presentations[1] were given by the six NIST team members and 11 Resource Group members.

In September 2019, three authors of this report – John Butler, Hari Iyer, and Sheila Willis – gave a workshop[2] entitled "DNA Mixture Interpretation Principles and Best Practices" in Palm Springs, California as part of the 30th International Symposium on Human

[1] https://strbase.nist.gov/AAFS2019-W10.htm
[2] https://strbase.nist.gov/pub_pres/ISHI2019-MixtureWorkshop.pdf

Identification (ISHI). In November 2019, John Butler and Hari Iyer gave an hour-long webinar[3] for the Center for Statistics and Applications in Forensic Science (CSAFE). Members of the NIST team[4] have provided additional workshops on validation (ISHI 2020) and useful literature regarding DNA measurement and interpretation (AAFS 2021). Further efforts to keep stakeholders informed include more than two dozen presentations at various conferences between 2018 and 2021 on aspects of DNA mixture interpretation, as well as our efforts collecting information and writing this report.

Plans for this DNA mixture interpretation review were announced to the general public in a NIST press release[5] on October 3, 2017, and through an interview and subsequent ProPublica news article[6] shortly thereafter. A plain language summary covering DNA mixtures and why they are sometimes difficult to interpret was also shared online[7] during the course of this study.

1.10. Structure of This Report

This report contains six chapters and two appendices. Following this introductory chapter, Chapter 2 provides background information on DNA and describes principles and practices involved in mixture interpretation. Chapter 3 lists data sources used and strategies to locate them. Chapters 4 and 5, which are the core of the report, discuss reliability and relevance issues in DNA mixture interpretation. Chapter 6 explores the potential of new technologies to aid DNA mixture interpretation. Finally, two appendices provide a brief history of DNA mixture interpretation (Appendix 1) and perspectives on training and continuing education (Appendix 2) to provide context for how the field has progressed and recommendations to strengthen it going forward.

The initial release of this report is a draft document, and we welcome comments and feedback from readers. All relevant submitted comments will be made publicly available and will be considered when finalizing this report. Do not include personal information, such as account numbers or Social Security numbers, or names of other individuals. Do not submit confidential business information, or otherwise proprietary, sensitive, or protected information. We will not post or consider comments that contain profanity, vulgarity, threats, or other inappropriate language or like content. During the 60-day comment period, comments may be sent to scientificfoundationreviews@nist.gov.

[3] https://forensicstats.org/portfolio-posts/dna-mixture-interpretation-thoughts-and-lessons-learned-from-a-nist-scientific-foundation-review/
[4] https://strbase.nist.gov/training.htm
[5] https://www.nist.gov/news-events/news/2017/10/nist-assess-reliability-forensic-methods-analyzing-dna-mixtures
[6] https://www.propublica.org/article/putting-crime-scene-dna-analysis-on-trial
[7] https://www.nist.gov/featured-stories/dna-mixtures-forensic-science-explainer

2. Chapter 2: DNA Mixture Interpretation: Principles and Practices

DNA mixture interpretation principles and practices are introduced in this chapter. The DNA testing process involves measurement and interpretation. Measurements reflect the physical properties of the sample while interpretation depends on the DNA analyst assigning values that are not inherent to the sample. Multiple statistical approaches are used to answer different questions. This includes strength-of-evidence interpretation, such as the random match probability (for major components of mixtures), the combined probability of inclusion, and the likelihood ratio. DNA samples are not equal in complexity and some are more difficult to analyze than others. Factors influencing the complexity include the number of contributors, DNA quantities of components, mixture ratios, sample quality, and the degree of allele sharing. In addition, artifacts created during the process of generating the DNA profile contribute to the challenge of DNA mixture interpretation. Continuous probabilistic genotyping systems, which report a likelihood ratio based on a pair of selected propositions, utilize more information from a DNA profile than binary approaches. The theory and application of likelihood ratios are introduced here in the context of probabilistic genotyping software. The chapter concludes with 16 principles related to DNA mixture interpretation. This information is intended as a precursor to topics covered in other chapters on reliability of measurements and interpretation (Chapter 4), relevance and case context (Chapter 5), and the potential of new technology (Chapter 6).

2.1. Value of DNA Evidence to Forensic Science

Forensic science processes involve collection, analysis, interpretation, and reporting of evidence. Since its introduction in the mid-1980s (Gill et al. 1985), DNA testing has been an important resource to forensic science and the criminal justice system. Forensic DNA results provide important capabilities to aid law enforcement investigations, strengthen prosecutions, and enable exoneration of the innocent. These capabilities include (1) ability to identify an individual or associate a perpetrator with a crime scene, since DNA remains unchanged throughout life and across bodily cells, (2) high sensitivity with DNA amplification techniques, (3) well-established quality assurance measures, (4) ability to provide a numerical strength of the evidence based on established genetic principles and statistical models, (5) use of close biological relatives as potential reference points through applying established characteristics of genetic inheritance, and (6) new technology development aided by biotechnology and genomics efforts (see Butler 2012, Butler 2015a, Butler 2015b).

DNA information can assist both law enforcement (investigative) and prosecutorial (evaluative) aspects of the criminal justice system. Investigative leads may be generated when a crime scene profile or a deconvoluted mixture component of a DNA profile are searched against a local, state, or national DNA database to locate a potential person of interest (POI). When writing reports or providing court testimony, the evaluative strength of available DNA evidence can be assessed when comparing a POI to an evidentiary DNA profile. Investigative and evaluative examinations serve different purposes and answer different questions (Gill et al. 2018). The evaluative uses of DNA information are held to a higher standard than investigative ones.

Successful DNA analysis and resulting interpretation depends on the quality and quantity of the crime scene evidence (the "Q" or questioned sample) and the availability of a reference sample (the "K" or known sample). When appropriate Q and K DNA profiles are available, forensic scientists can perform a Q-to-K comparison and report the strength of this association using specific assumptions and usually one of several statistical approaches. A range of DNA profile qualities and quantities can be observed in forensic casework samples.

2.1.1. DNA Basics

A biological sample collected directly from a single individual (i.e., a "single-source sample") can be analyzed to generate a *DNA profile*. This profile identifies the genetic variants (termed *alleles*) found at tested locations (*loci* or when singular, *locus*) along the human *genome*. Usually less than two dozen loci, which are each in a size range of 100 to 400 *nucleotides* in length, are examined to generate a forensic DNA profile. Thus, information from only a few thousand nucleotides in total are examined in a forensic DNA test out of the approximately three billion nucleotides across 23 pairs of *chromosomes* that comprise the human genome.

Core sets of loci have been selected for use in national DNA databases (e.g., Budowle et al. 1998, Hares 2015). These tested loci, also termed *DNA markers*, were selected from non-protein-coding regions of the genome occurring between genes. Thus, results from forensic DNA profiles are not expected to contain information on physical traits or susceptibility to genetic diseases (e.g., Katsanis & Wagner 2013).

The DNA markers used in most forensic applications include short genetic sequences that are repeated a variable number of times. These are called *short tandem repeat (STR)* markers. The number of repeats at each STR marker varies from person to person. This variability in STR alleles is what allows a DNA analyst to associate a DNA sample with an individual. A variety of commercially available STR kits have been used over the past 25 years. These kits have evolved and expanded over time permitting 6 to 10 markers in the mid- to late-1990s, 10 to 16 loci between 2000 and 2013, and 20 to 24 markers or more, presently (see Butler 2012, pp. 108-122 and Butler 2015a, pp. 17-21).

Humans are *diploid, i.e., they* possess two copies of each non-sex-determining chromosome (*autosome*) with one allele at each locus coming from an individual's biological mother and the other from their biological father. Thus, alleles at each tested locus exist in pairs, which are termed *genotypes*. Allele pairs that are indistinguishable and cannot be differentiated with the technology used are termed *homozygous.* An analyst might label these 12,12 or A,A. Those genotypes that are distinguishable from one another, in other words, differing alleles that are inherited from each parent, are called *heterozygous.* These might be labeled 12,13 or A,B.

When analyzing the DNA sample, a technique called the *polymerase chain reaction* (PCR) is used to create millions of copies of each STR marker. The purpose of this step, called *amplification*, is to generate a quantity of STR alleles sufficient for laboratory analysis. The

PCR process labels STR alleles with different colored fluorescent dyes to enable multiple markers to be examined in a single analysis.

The amplified and labeled STR alleles are then separated and detected using a technique called *capillary electrophoresis* (CE). CE instruments utilize four, five, or six dye-channels to analyze many STR markers simultaneously. Peak positions and heights are visualized by dye-channel color and DNA size in a chart format called an *electropherogram* (EPG). The location of peaks on the chart indicate which alleles (i.e., STR marker variants of different size) are present in the tested sample. The EPG is the raw data that must be interpreted to draw conclusions from the sample.

The amplification step using PCR and the separation and detection step using CE are important in the context of this report because they produce artifacts that can confound the interpretation. These artifacts are discussed in Section 2.2.1 Factors that Affect Measurement Reliability. Analysis of samples containing very small quantities of DNA tends to produce EPGs with a higher proportion of artifacts due to *stochastic variation* or random sampling of DNA molecules (see Butler & Hill 2010).

The amount of DNA recovered from crime scene evidence depends on a number of factors including the amount of biological material deposited, DNA extraction efficiencies, and environmental conditions that can contribute to DNA degradation or PCR inhibition. When degraded, DNA molecules break into smaller pieces, such that some or all of the tested loci are no longer detectable by PCR. Loss of allele information from a DNA profile is termed *allele drop-out* or, if both alleles are not present or detectable, *locus drop-out*. Swabs from so-called "touch evidence" samples, which typically have a relatively small quantity of biological material deposited (with perhaps tens of cells), are more likely to exhibit allele drop-out compared to visible blood or semen stains, which contain hundreds to thousands of cells.

Further details on DNA basics and the process for generating forensic DNA profiles are available in textbooks such as *Fundamentals of Forensic DNA Typing* (Butler 2009) or *An Introduction to Forensic Genetics, Second Edition* (Goodwin et al. 2010).

2.1.2. DNA Mixtures

A DNA mixture can occur when biological material from more than one individual is deposited on the same surface. In single-source samples, only a single genotype is possible at each locus. With DNA mixtures, however, more than one genotype combination may be possible at each locus. This ambiguity is an important reason why DNA mixture interpretation is more difficult than testing single-source samples. Interpretation of evidence, in the words of a leader in the field, "continues to be the most difficult challenge that faces scientists, lawyers, and judges" (Gill 2019b).

DNA from multiple contributors cannot be physically separated once DNA molecules are extracted from their biological cells (see Chapter 6 and Figure 6.2). Instead, DNA mixture interpretation is an effort to (1) infer possible genotypes as detectable sample contributors (a

process sometimes referred to as *deconvolution* of the mixture components) and (2) provide the strength of evidence for a POI to be included in an evidentiary DNA profile.

DNA mixtures are common, and even expected, in many evidence types coming from criminal investigations. Person-on-person crimes, such as sexual assaults or homicides, may involve DNA mixtures of biological material (e.g., semen or blood) from the perpetrator and the victim. DNA mixtures may be detected in many property crimes where items in a house or a vehicle are handled by a burglar but also touched previously by the owner(s) or other people not associated with the crime in question.

In their 2016 report, the President's Council of Advisors on Science and Technology (PCAST) differentiated between single-source samples, simple mixtures, and complex mixtures (PCAST 2016). We would point out that *DNA samples and mixtures in forensic casework exist on a continuum, and there are no hard and fast lines defining or separating particular categories.* Artificial categories have been described (e.g., Wickenheiser 2006, Schneider et al. 2006b, Schneider et al. 2009) to explain where use of different approaches to mixture interpretation may be helpful.

An analogy involving mathematics may assist in illuminating aspects of various categories that have been used for DNA profiles. If we consider that single-source DNA profiles are like basic arithmetic and simple mixtures are like algebra, then complex mixtures (e.g., profiles with three or more contributors, with low-level and/or degraded DNA where *uncertainty in assigning contributor genotypes increases*) can be considered the equivalent of calculus. In a similar manner, calculus builds upon principles of arithmetic and algebra but requires more advanced training and perspective to fully appreciate; so does DNA interpretation of complex mixtures. Validation studies and training are required to develop the necessary expertise. However, the fundamental principles must be understood before approaching complex DNA mixture interpretation.

KEY TAKEAWAY #2.1: DNA mixtures, where the DNA of more than one individual is present in a sample, are inherently more difficult to interpret than single-source DNA samples.

2.2. The DNA Testing Process

The general steps involved in forensic DNA testing are illustrated in Figure 2.1. Briefly, an item of evidence is collected or a sample is obtained by swabbing a surface containing possible crime scene evidence. DNA, which could be from one or more contributors, is extracted from the sample. Following DNA extraction, DNA quantitation (with adjustments for amount of human DNA present), and PCR amplification with predefined DNA marker sets of STR loci, the amplification products are separated and detected. Results are then interpreted, compared to reference sample profiles along with a statistical estimate of the strength of evidence, and reported in a written summary. If a case goes to trial, then the analyst might be asked to provide testimony as an expert witness.

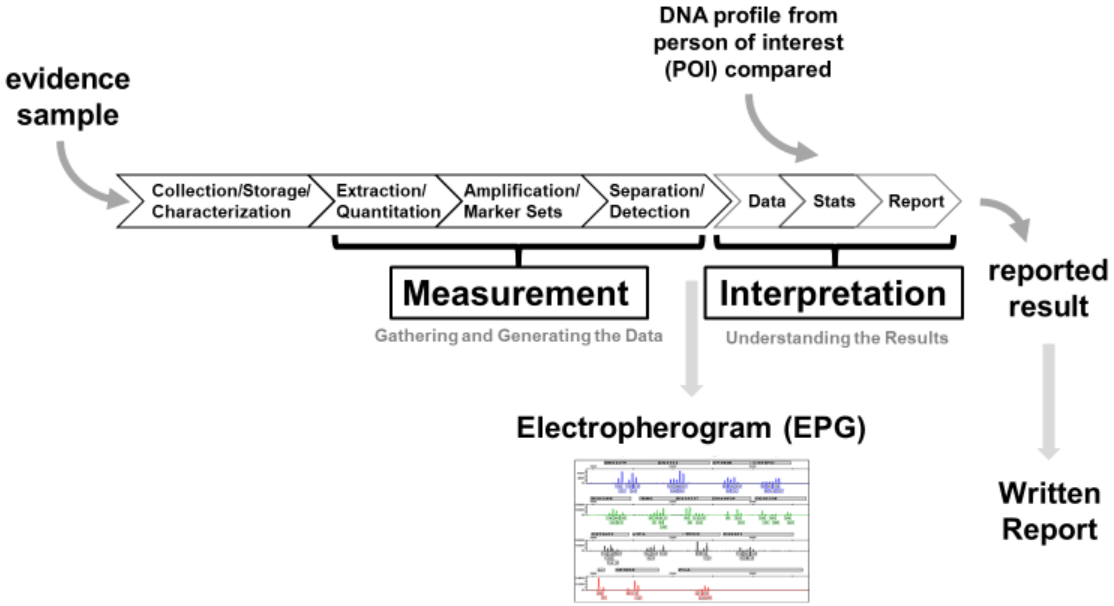

Figure 2.1. Illustration of the general steps involved in processing an evidence sample containing DNA (either single-source or mixture). The output of the measurement steps is an electropherogram. The output of interpretation is a reported result in a written report.

This overall process can be divided into two parts (Figure 2.1): (1) ***measurement*** that involves a series of steps to generate a DNA profile and (2) ***interpretation*** of the DNA profile to help fact finders understand the value of the evidence. The measurement steps result in an *electropherogram* (EPG), which is a representation of the DNA profile observed from the test sample at specific DNA locations. Interpretation of the EPG concludes with a written report describing a strength-of-evidence statistic for Q-to-K comparison with the POI(s), and in some cases, court testimony.

Figure 2.1 outlines general steps; however, the details of measurement and interpretation steps may vary between laboratories. For example, different STR kits, PCR cycle numbers, and CE instruments may be used in different laboratories. Likewise, interpretation approaches may differ among analysts and, more often, laboratories. Therefore, we discuss general practices and principles involved in measurement and interpretation rather than one specific protocol.

Measurements reflect the physical properties of the sample while interpretation depends on the DNA analyst assigning values that are not inherent to the sample. These interpretations are based on case context and their own training and experience. In part, because interpretation of the same evidence may vary from person to person, it is described as an opinion (see Gill 2019b). Complex DNA mixtures are challenging because they require more interpretation than a high-quality, single-source sample.

When a POI is available for comparison to the evidence, DNA analysts render their opinions (often in the form of likelihood ratios) in written reports drawing upon (1) empirical data from the evidence sample compared to a POI's DNA profile, (2) available relevant case context information (e.g., location from which the sample originated, body fluid screening results, quantity of DNA extracted, and overall quality of the DNA profile) and (3) their training and experience (see SWGDAM 2017a).

Further details are available in textbooks such as *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists* (Evett & Weir 1998), *Forensic DNA Evidence Interpretation* (Buckleton et al. 2005), and *Forensic Practitioner's Guide to the Interpretation of Complex DNA Profiles* (Gill et al. 2020b).

KEY TAKEAWAY #2.2: Generating a DNA profile involves measuring the inherent physical properties of the sample. Interpreting a DNA profile involves assigning values that are not inherent to the sample. To do this, the DNA analyst uses their judgment, training, tools (including computer software), and experience, and considers factors such as case context.

2.2.1. Factors that Affect Measurement Reliability

The measurement portion of the DNA testing process produces an EPG (see Figure 2.1). DNA mixture interpretation (as well as single-source DNA interpretation) is conducted in the presence of PCR amplification and CE analysis artifacts. These artifacts influence the complexity of the DNA profile to be interpreted and validation studies are performed to characterize them.

Artifacts that may be observed in an EPG include the non-allelic products of the PCR amplification process (e.g., stutter products, non-templated nucleotide addition, or other non-specific products), anomalies of the detection process (e.g., single or multichannel voltage spikes or "pull-up" from spectral channel bleed-through), or by-products of primer synthesis (e.g., "dye blobs") (see Butler 2015a, pp. 183-210).

There are several quantifiable factors that affect measurement reliability.

The first is peak position. The DNA profile peaks observed in an EPG are fluorescently labeled PCR products (STR alleles) that differ in length due to variation in the number of STR repeats. Use of an internal size standard with each tested sample along with calibration to an allelic ladder enables accurate STR allele designations with electrophoresis separation and detection systems (Butler 2015a, pp. 48-58). Peak positions are measured as migration time (raw data), nucleotides (against the size standard), and allele designations (against an allelic ladder). This factor is important because the accurate determination of peak locations is necessary for reliable STR allele designations.

Another measurable factor includes peak morphology or resolution. This is when wide peaks result in poor resolution and the inability to fully separate STR alleles that differ by as little

as a single nucleotide. Capillaries fail and resolution is lost after many CE sample injections. Peak resolution can be monitored by examining separation of the alleles in an allelic ladder (Butler 2015a, pp. 201-202). This factor is important because failure to resolve similar length STR alleles may result in missing true contributor genotypes. Wide peaks may also size inaccurately.

Peak heights are measured in relative fluorescence units (RFUs) and are generally proportional to the amount of PCR product detected. While an RFU value does not necessarily correspond to a specific number of picograms of DNA, variation in peak heights matters because this information is used to deconvolute mixture components into contributor genotype possibilities. On-scale data are essential when calculating information impacted by peak heights, such as stutter percentages and peak height ratios (Butler 2015a, pp. 30-33).

Stutter products, another measurable factor, are produced during PCR amplification from slippage of the DNA strands while being copied, and are typically one repeat shorter or longer than their originating STR allele (Walsh et al. 1996, Butler 2015a, pp. 70-79). The relative heights of stutter products correlate in large measure to the length of sequence composed of the same repeat pattern of the corresponding STR allele (Brookes et al. 2012). Stutter products are the most influential artifacts in an EPG because they can be indistinguishable from true alleles of minor contributors and therefore impact DNA interpretation (Gill et al. 2006b).

Spectral artifacts are a measurable factor, as well. This is an anomaly of the detection process where fluorescent signal from one spectral channel bleeds through into an adjacent color channel (e.g., green into blue). Pull-up occurs from a saturating signal on the instrument detector (see Butler 2015a, pp. 32, 200-201). Artifacts matter because when low quantities of DNA are tested, it can be challenging to differentiate true alleles from amplification or detection artifacts. Spectral artifacts may also signal off-scale data in an EPG that should be avoided, as the stutter ratio will not be accurate.

Relative peak heights of allele pairs within a locus are another measurable factor. Heterozygous STR loci possess two alleles that differ in overall PCR product size. The peak heights of these two "sister" alleles can be compared in single-source samples to enable genotype assumptions in samples containing more than one contributor (Butler 2015a, pp. 87-93). This factor is important in order to determine the limits of pairing alleles into genotypes with binary approaches and also helps define parameters used for assigning potential genotypes and mixture ratios with PGS systems.

Assessing relative peak heights across loci in a DNA profile provides an indication of the quality of a sample. With degraded DNA, peak heights decrease from left to right across an EPG (small-size to large-size STR alleles) (Butler 2015a, pp. 121-123). This factor is important because ratios between mixture components may differ across tested loci.

Finally, baseline noise is also a measurable factor in this context. Noise exists in all measuring systems. In a DNA profile EPG, noise is represented as jitter in the baseline signal

(Butler 2015a, p. 33). Characterizing the level of baseline noise enables an analytical threshold to be set and a lower limit of reliability to be established for peak heights.

These measurable factors in DNA profile EPGs can affect measurement reliability. Table 2.1 lists validation experiments typically conducted and the purpose of each factor in DNA mixture interpretation. For foundational purposes, we need to consider what we know about uncertainty around each of these measurements as well as other factors that can influence interpretation, including artifacts. For this reason, studies regarding stutter product variation (e.g., Bright & Curran 2014) and allele drop-in (e.g., Moore et al. 2020) are valuable.

Table 2.1. Measurable factors and features in a short tandem repeat (STR) DNA profile electropherogram (EPG) that influence DNA mixture interpretation with binary or probabilistic genotyping software (PGS) approaches. Assessment for some of these factors are more qualitative than quantitative. Validation experiments (SWGDAM 2016) to demonstrate measurement reliability are typically performed using single-source DNA samples (e.g., Moretti et al. 2001a, Moretti et al. 2001b, Butler et al. 2004, Rowan et al. 2016).

Measurable Factor (units)	**Validation Experiments to Demonstrate Reliability**	**Purpose in DNA Mixture Interpretation**
1a) Peak Position (nucleotides)[a]	Accuracy and precision studies to verify consistency in peak sizing and STR allele calls	To determine limits of peak sizing and accurate allele calls compared to an allelic ladder
1b) Peak Morphology or Resolution	Examination of peak height and width in allelic ladders and inspecting separation of similar length allelic ladder alleles (e.g., TH01 alleles 9.3 and 10) as quality control of kit and instrumentation	To examine CE separation resolution that can influence ability to accurately designate similar length STR alleles (e.g., Butler et al. 2004)
2a) Peak Height (RFU)[b]	Precision studies to verify consistency in allele calls; variability is typically studied in terms of presence or absence; repeatability of peak heights can be investigated with replicate injections and reproducibility of peak heights with replicate PCR amplifications of sample aliquots	To determine the presence of stochastic effects such as allele drop-out (only when examining ground-truth samples); presence of contamination including allele drop-in (only when examining ground-truth samples); help infer parameters used for assigning potential genotypes and mixture ratios with PGS systems

Measurable Factor (units)	**Validation Experiments to Demonstrate Reliability**	**Purpose in DNA Mixture Interpretation**
2b) Stutter Products	Calculation of stutter peak height to STR allele peak height ratio	To determine stutter thresholds applied in binary approaches or to develop and inform stutter models for PGS; multiple types of stutter (e.g., n-1, n-2, n+1) and approaches (e.g., allele-specific, locus-specific, or profile-wide) have been used
2c) Spectral Artifacts	Visual inspection of EPGs for signal bleed-through between dye channels (e.g., green into blue) with overloaded peaks; calculation of bleed-through to parent peak height ratio; quality control for spectral calibration of system	To determine upper limits of DNA quantities used to generate profile EPG; to help define parameters for distinguishing bleed-through from true peaks
2d) Relative Peak Heights of Allele Pairs within a Locus	Calculation of heterozygote balance or peak height ratios from heterozygous allele pairs in single-source samples	To determine the limits of pairing alleles into genotypes with binary approaches and to help infer parameters used for assigning potential genotypes and mixture ratios with PGS systems and calculating probability of allele drop-out
2e) Relative Peak Heights Across Loci in a DNA Profile	Calculation of interlocus balance to determine if peak heights are significantly reduced for longer length PCR products (on the right side of the EPG)	To estimate the level of DNA degradation or PCR inhibition (some new STR kits have quality sensors included in the STR profile) and to help infer parameters used for assigning potential genotypes and mixture ratios with PGS systems
2f) Baseline Noise (RFU)	Examination of variation in baseline noise from negative controls and extraction blank samples	To determine the analytical threshold so that noise can be distinguished from true peaks (that can be alleles or artifacts); multiple approaches have been used (e.g., Bregu et al. 2013)

[a] in nucleotides relative to an internal size standard with allele calls made in comparison to an allelic ladder run simultaneously or sequentially with the same internal size standard

[b] relative fluorescence units

A series of single-source samples and negative controls are commonly examined to assess observed variability of these measurable factors including artifact behavior. Higher variability in peak heights leads to greater uncertainty in the possible genotype combinations for contributors in mixture interpretation. These measurable factors are mathematically modeled to create probability distributions with probabilistic genotyping software (e.g., Taylor et al. 2016c, Kelly et al. 2018).

2.2.2. Steps in the Interpretation Process

Interpretation begins with separate evaluations of EPGs from a Q (the evidentiary DNA profile) and a K (the DNA profile of a POI). Data interpretation decisions are made separately for Q and K EPGs, in accordance with validation-based interpretation protocols, which includes questions such as "is this a peak or part of baseline noise?," "is this an allele or an artifact?," "could this DNA profile have come from more than one contributor?," etc. Increasingly, these decisions, which respond to the above questions, are made with assistance from suitable computer software. If the Q profile appears to be a mixture, then the DNA analyst assesses possible genotype combinations of contributors and compares these possible genotypes with one (or more) POIs.

In 2006, the DNA Commission of the International Society for Forensic Genetics (ISFG) published nine recommendations on DNA mixture interpretation (Gill et al. 2006b). These recommendations, which are summarized in Appendix 1 (Box A1.4), serve as core fundamental principles for working with DNA mixtures. The ISFG recommendations build upon previous work (e.g., Weir et al. 1997, Clayton et al. 1998, Bill et al. 2005) and provide a framework built around the steps shown in Figure 2.2. This framework was initially developed for manual interpretation methods with simple, two-person mixtures. However, the concepts also apply to software programs used for examining complex mixtures.

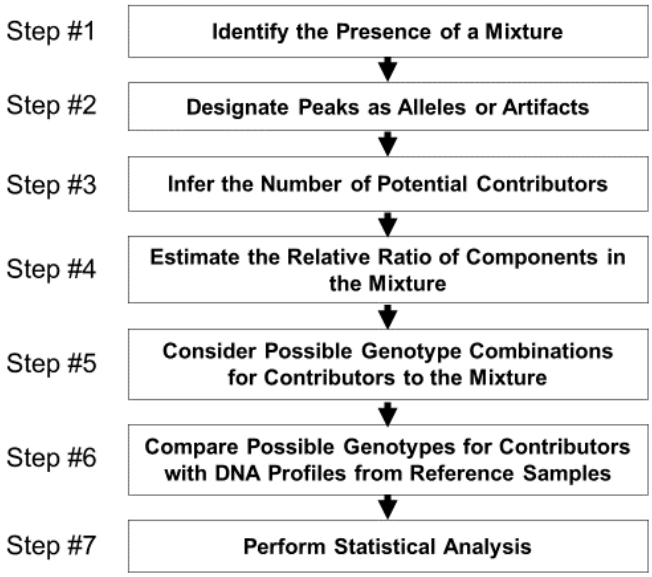

Figure 2.2. Steps in DNA mixture interpretation first outlined by the UK Forensic Science Service (Clayton et al. 1998) and endorsed by the ISFG DNA Commission (Gill et al. 2006b).

The 2006 ISFG DNA Commission noted that there are three kinds of alleles in a crime scene profile. Those are alleles that (1) are unmistakable, (2) may be masked by an artifact such as stutter, and (3) have dropped out completely and are therefore not detected (Gill et al. 2006b). When assessing possible genotype combinations of contributors to a mixture, a DNA analyst may encounter any or all of three situations.

Alleles may contain components from more than one contributor that are shared and need to be deconvoluted (i.e., separated out into component genotypes). More possible contributors mean more possible genotype combinations with any of the observed set of alleles. The creation of computer software to explore possible genotype combinations has been an important development in DNA mixture interpretation (Coble & Bright 2019).

For a detailed analysis of these interpretation steps using an example DNA mixture and the various statistical approaches discussed later in this chapter, see *Advanced Topics in Forensic DNA Typing: Interpretation* (Butler 2015a, pp. 129-158 and pp. 537-567).

2.3. Complexity and Ambiguity with Mixture Interpretation

DNA samples recovered from crime scenes vary in quality and may be challenging to analyze and interpret (Word 2011). The types of cases being submitted to a laboratory will impact the complexity of mixtures observed (e.g., Torres et al. 2003, Mapes et al. 2016). Over the past decades as DNA testing methods have become more sensitive (see Appendix 1), more challenging evidence types (e.g., touch evidence with limited quantities of DNA and complex DNA mixtures) have been submitted to forensic laboratories (Mapes et al. 2016). A "complex" DNA mixture sample is one in which uncertainty exists in the genotype assignments at tested STR loci in a DNA profile.

2.3.1. Factors that Contribute to Increased Complexity

There are at least three challenges that are fundamental to DNA mixture interpretation: (1) ***stochastic variation***, which impacts recovered quantities of alleles from contributors and can lead to uncertainty in assigning alleles to genotypes and uncertainty in assigning genotypes to contributor profiles when examining small amounts of DNA, (2) ***stutter products***, which create uncertainty through minor contributor(s) with alleles in the stutter positions of major contributor(s) alleles, and (3) ***sharing of common alleles***, which influences the ability to estimate the number of contributors, particularly when combined with stochastic variation and the existence of stutter products that create uncertainty in deconvoluting mixture components.

Ambiguity in DNA mixture interpretation arises when (1) small quantities of DNA are tested that, when copied, may not fully represent the original sample (i.e., the recovered DNA profile is incomplete and missing information), (2) a mixture of DNA from more than one individual may make it hard to deconvolute or separate information from each individual contributor depending on the contributor ratios, amounts, and degree of allele overlap, (3) the DNA molecules may be damaged or destroyed (i.e., the recovered DNA profile is incomplete

and may be missing information), (4) environmental contamination may impact the ability to recover the original sample (DNA may come from a transfer not related to the crime or PCR inhibitors that lead to an incomplete recovered DNA profile), or (5) any combination of the previous four issues.

2.3.2. Improved Sensitivity Methods Can Result in Higher Complexity Profiles

As techniques for generating DNA profiles become more sensitive, smaller amounts of DNA can be detected, analyzed, and interpreted. DNA testing sensitivity has increased due to improvements in STR kits (e.g., Ensenberger et al. 2016, Ludeman et al. 2018), introduction of new CE instruments, use of higher PCR cycle numbers (e.g., Whitaker et al. 2001), reduced volume PCR (e.g., Leclair et al. 2003), PCR product desalting (e.g., Smith & Ballantyne 2007), and higher CE injection (e.g., Westen et al. 2009). "High" sensitivity DNA testing has become the new normal (Gill et al. 2015).

When analyzing small quantities of DNA, stochastic (random sampling) effects can cause alleles that are present in the sample to "drop out" of the detected profile (e.g., Lohmueller & Rudin 2013). Stochastic effects can also cause alleles that are not present in the sample to "drop in" to the profile (e.g., Moore et al. 2020). In other words, with low-quantity DNA samples, the resulting profile and EPG vary in how accurately they reflect the original sample, which can lead to loss of genotype information from a true contributor to the mixture.

Furthermore, in part due to stochastic variation, two low-quantity DNA samples collected from the same surface can produce DNA profiles with different peak heights and therefore different ratios of alleles and possible genotype combinations. Analyzing the same low-quantity DNA mixture two or more times can also produce dissimilar DNA profiles (e.g., Benschop et al. 2013). Interpretation methods need to be able to account for this ambiguity.

KEY TAKEAWAY #2.3: The process of generating a DNA profile can produce stochastic or random variation and artifacts that contribute to the challenge of DNA mixture interpretation.

2.3.3. Mixture Complexity Increases as Number of Contributors Increase

The challenge of genotype assignment increases with the number of contributors in a mixture due to the possibility of allele sharing (Alfonse et al. 2017). In addition, estimating the number of contributors in a DNA mixture becomes more uncertain when there are more contributors as noted in several publications (Paoletti et al. 2005, Buckleton et al. 2007, Coble et al. 2015). The frequency of occurrence for an allele from population data correlates to the degree of allele sharing that is expected if that allele is present in the crime scene DNA mixture. If mixture contributors are related, then even more allele sharing between contributors is expected. Thus, with more contributors to a mixture, more allele sharing occurs, which increases the complexity and ambiguity of interpretation (e.g., Dembinski et al. 2018, Lynch & Cotton 2018).

KEY TAKEAWAY #2.4: DNA mixtures vary in complexity, and the more complex the sample, the greater the uncertainty surrounding interpretation. Factors that contribute to complexity include the number of contributors, the quantity of DNA from each contributor, contributor mixture ratios, sample quality, and the degree of allele sharing.

2.4. Approaches and Models for Dealing with Complexity

DNA mixture interpretation can be divided into two general approaches: (1) binary (e.g., Budowle et al. 2009) or (2) probabilistic genotyping (e.g., Gill et al. 2012). Both approaches generally follow the seven steps outlined in Figure 2.2 with an important difference at step five, where possible genotype combinations of contributors are considered.

2.4.1. Binary Statistical Approaches

Statistical analysis provides a quantitative expression of the strength or value of the evidence when K is considered as a possible contributor to Q. When a DNA analyst believes that a *major* component can be confidently separated from a *minor* component of a mixture, then a random match probability (RMP) or modified RMP (mRMP) method has been used on the major component – treating it statistically as a single-source sample (DAB 2000, Bille et al. 2013, SWGDAM 2017a). Likewise, conditioning on the donor of an intimate sample under the assumption of a defined number of contributors has been used to perform mRMP calculations on the foreign profile even if it is not the major component (see SWGDAM 2017a). For mixture contributors that cannot be confidently distinguished because of allele overlap or similar mixture ratios, then "manual" likelihood ratio (LR) methods have been used (e.g., Weir et al. 1997, Evett & Weir 1998, Gill et al. 2006b). Either of these approaches can be applied with simple, two-person mixtures, such as sexual assault intimate samples.

A commonly used statistical approach in the United States has been the combined probability of inclusion (CPI), which is defined as the probability that a randomly chosen (unrelated) individual would be included as a possible contributor to the mixture (NRC 1992, Bieber et al. 2016). Once a K is included as a possible contributor to Q, the CPI, which is sometimes referred to as random man not excluded (RMNE), indicates the statistical value of all possible genotypes present in a mixture (giving them equal weight) based on observed alleles (NRC 1992, p. 59).

As seen in Table 2.2, different statistical approaches answer different questions using the data available. Each approach has strengths and weaknesses (e.g., Buckleton & Curran 2008). A trier of fact in a court of law is typically interested in what DNA results mean in a particular case, with regard to a specific POI and set of case circumstances. For this reason, likelihood ratio methods (Question 4 in Table 2.2), as will be discussed later in this chapter, have been considered a valuable tool in DNA mixture interpretation and recommended by the ISFG DNA Commission (Gill et al. 2006b, see also Appendix 1).

Table 2.2. Different approaches used in statistical analysis of DNA and the questions addressed. RMP and MP are calculated for single-source DNA profiles (or deduced major profiles). CPI and LR are calculated for mixtures.

	Question	Approach (Reference)	Specific Requirements
1	What is the probability of observing this profile in the population? (i.e., what is the rarity of the profile?)	Profile Probability (or random match probability, RMP) (NRC 1996 for single-source samples; Bille et al. 2013 for mixtures)	For mixtures, an assumption that the major contributor can be distinguished from minor components so that specific genotypes in the major can be inferred
2	What is the probability of observing this profile in the population if we have already observed one person with this profile in this population?	Match Probability (MP) (Balding & Nichols 1994, Weir 2001)	Use of conditional probabilities and a subpopulation correction
3	What is the probability that a person selected randomly in the population would be included (or not excluded) as a possible donor of the DNA typing result?	Combined Probability of Inclusion (CPI) (Bieber et al. 2016)	All alleles for all contributors are all present at the reported loci (i.e., cannot cope with allele drop-out that is expected with low quantities of DNA)
4	By how much do the DNA typing results support the person of interest (POI) being the donor under specific assumptions and propositions?	Likelihood Ratio (LR) (Evett & Weir 1998)	An assumption as to the number of contributors and a specific pair of propositions

2.4.2. Limitations with Binary Methods

Traditional *binary* methods and approaches to DNA mixture interpretation (e.g., Clayton et al. 1998) work under the assumption that a specific genotype of interest is either present or absent. Statistical approaches include LR (e.g., Weir et al. 1997), CPI (e.g., Budowle et al. 2009), and mRMP (Bille et al. 2013). However, binary approaches cannot account for the possibility of missing information (i.e., allele drop-out) when testing small quantities of DNA, nor can they account for the possibility of allele drop-in, which is more common with high-sensitivity methods (Balding & Buckleton 2009).

As noted in a recent textbook:

> "These [CPI] calculations found favor and were widely used, because they were very easy to implement and assumptions about the number of contributors were not

> needed. There are two drawbacks however: (1) There is an implicit assumption that all of the contributors have all alleles fully represented in the EPG. There is no allele drop-out present, i.e., the calculation is not valid for minor contributors with drop-out that is or may be present. (2) The calculation exists by itself and is unchanged by the suspect's profile, i.e., the calculation is unmodified by the presence of a suspect who matches or does not match … When an RMNE is reported, then it is necessary to make a binary decision about whether a suspect could have contributed to a crime stain. Either he has (probability = 1) or he has not (probability = 0)" (Gill et al. 2020b, p. 386).

Thus, proper application of a CPI calculation is dependent on all possible alleles being present and therefore commonly involves use of a stochastic threshold to provide confidence that loci used in statistical calculations are not missing alleles (Moretti et al. 2001a, Moretti et al. 2001b, Budowle et al. 2009, SWGDAM 2017a). In addition to the CPI statistic not accounting for the possibility of allele drop-out when testing small quantities of DNA, this same limitation exists for minor components of complex mixtures, even when the total DNA input is optimal. Guidance on the appropriate application of CPI has been published (e.g., Bieber et al. 2016, Buckleton et al. 2016, pp. 238-247).

In a binary approach, measurement limitations and stochastic effects can make it difficult to identify which of the peaks in an EPG correspond to alleles, which are stutter products, and which are noise peaks. During the PCR amplification process, certain alleles present in the original sample may not have a corresponding peak in the EPG (failure to amplify) or may be judged as absent (below a predetermined analytical threshold), and certain peaks in the EPG that are artifacts may be judged to be real alleles from the original sample (e.g., stutter products, allele drop-ins, spectral pull-up peaks).

To address the complexity that comes with increased DNA sensitivity (Gill et al. 2000), leaders in the forensic DNA community have looked to probabilistic genotyping in recent years (see Appendix 1).

2.4.3. Advantages with Probabilistic Genotyping Approaches

Probabilistic genotyping approaches represent a way to address complexity in DNA profiles. In their 2006 publication, the ISFG DNA Commission concluded:

> "A future approach would elaborate the combinatorial approaches by taking into account all aspects including stutter, contamination and other artefacts, allelic drop-out, such as using a probabilistic weighting for each possible genotype rather than just using a weighting of zero or one, as is inherent in the restricted combinatorial (binary) approach" (Gill et al. 2006b).

The first three authors of this publication (Peter Gill, Charles Brenner, and John Buckleton) have been involved in developing probabilistic genotyping software systems over the past decade.

Probabilistic genotyping enables weighting (based on the probability of) specific genotype contributions through biological and statistical models informed by probabilities of missing alleles (Kelly et al. 2014, Gill et al. 2020b). These methods incorporate mathematical

modeling that can reflect uncertainty in the mixture interpretation. PGS uses LR calculations, where the probability of the data being observed are compared under two hypotheses or propositions. Depending on the propositions used and probabilistic genotyping models applied, different LRs can be produced (see Gill et al. 2018).

Probabilistic genotyping considers possible genotype combinations for contributors where information may be missing in a crime scene DNA profile (Gill et al. 2012). Two different probabilistic genotyping approaches have been used: discrete or continuous (Kelly et al. 2014, Gill et al. 2015). Table 2.3 compares binary and probabilistic genotyping approaches to DNA mixture interpretation.

Table 2.3. Comparison of approaches used in DNA mixture interpretation. CPI = combined probability of inclusion, mRMP = modified random match probability, LR = likelihood ratio. Adapted from ISFG 2015 workshop by John Butler and Simone Gittelson available at https://strbase.nist.gov/training/ISFG2015-Basic-STR-Interpretation-Workshop.pdf.

	Takes into account		***Mathematically models***	
	Presence/ absence of alleles	Possible genotypes based on peak heights	Allele drop-out and allele drop-in	Peak heights
Binary Approaches				
CPI	X			
mRMP	X	X		
LR (binary)	X	X		
Probabilistic Genotyping				
LR (discrete)	X		X	
LR (continuous)	X	X	X	X

Discrete approaches (sometimes referred to as semi-continuous) require the analyst to determine the presence of alleles and artifacts prior to use in their models. Potential allele drop-out or allele drop-in are accommodated without considering parameters such as peak heights, peak height ratios, mixture ratios, or stutter percentages (e.g., Balding & Buckleton 2009, Inman et al. 2015).

Continuous approaches (sometimes called fully continuous) use all observed alleles and their corresponding peak height information and accommodate potential allele drop-out or allele drop-in, while also incorporating information regarding peak height ratios, mixture ratios, and stutter percentages. Some continuous models even consider amplification efficiencies, degradation, and other factors (e.g., Perlin et al. 2011, Taylor et al. 2013, Cowell et al. 2015).

KEY TAKEAWAY #2.5: Continuous probabilistic genotyping software (PGS) methods utilize more information from a DNA profile than binary approaches.

2.5. Likelihood Ratios: Introduction to Theory and Application

Dennis Lindley, a modern pioneer in using Bayesian statistics, introduced the concept of likelihood ratios (LRs) to forensic science more than four decades ago (Lindley 1977). LRs were first applied to DNA mixture interpretation about 14 years later (Evett et al. 1991; see Appendix 1). The LR involves a ratio of two conditional probabilities: the probability of the evidence given that one proposition (or hypothesis, narrative) is true and the probability of the evidence given an alternative proposition is true. The magnitude of the LR value is commonly used to express a strength of the evidence in favor of one proposition versus an alternative proposition.

Numerical results obtained from performing LR calculations are dependent on the evidence available, statistical models applied, propositions selected, and the scientist making various judgments. LR results vary based on amount of information available and assumptions made. With less information (e.g., results from a partial DNA profile possessing fewer loci), a lower LR number should be obtained with a well-calibrated system (Meuwly et al. 2017).

2.5.1. Likelihood Ratio Framework

The LR framework or paradigm is linked to Bayes Theorem, which is attributed to an eighteenth-century clergyman named Thomas Bayes (Bayes 1763). Bayesian statisticians[8] define the probability of an event as *the degree of belief* in the truth of the proposition that asserts it will occur. An individual's degree of belief is updated, in light of any new information, by multiplying the individual's prior degree of belief the event will occur (expressed as odds) by their LR to obtain their posterior degree of belief (expressed as odds). The Bayesian framework is based on the philosophical viewpoint that all probabilities are *personal*, meaning[9] "of, relating to, or coming as from a particular person." Probabilities quantify a personal state of uncertainty regarding the truth of propositions (see Lindley 2014, pp. 1 and 19, Kadane 2011, p. 1).

The term *assigning* is used when describing LR results (e.g., Bright & Coble 2020) rather than "calculating" to reflect dependence on subjective judgments. That is, different people may assign different values to the same evidence. Concerns have been raised that the LR framework applies only to personal decision making and cannot automatically be used for the transfer of information from one expert to a separate decision maker (Lund & Iyer 2017). Comments on these concerns have also been published (Aiken et al. 2018, Aiken & Nordgaard 2018, Gittelson et al. 2018).

In recent years, the LR framework (Jackson et al. 2006) has gained widespread acceptance in DNA mixture interpretation (e.g., NRC 1996, Gill et al. 2006b) as a way of reporting the strength of evidence (E) in support of one proposition (H_1 or H_p) over an alternative proposition (H_2 or H_d or H_a). For example, that the POI (and in some cases, specific other

[8] See https://en.wikipedia.org/wiki/Bayesian_statistics
[9] See https://www.dictionary.com/browse/personal

individuals) contributed to the crime sample, against a chosen alternative proposition stating, among other things, that the POI is a non-contributor to the mixture.

An LR is defined as the ratio of the probability of the findings given H_1 is true versus the probability of the findings given H_2 is true. Note that a reported LR value is *not* the odds that a particular proposition is true. The probabilities are assessed considering other relevant background information, often denoted as *I*.

Symbolically,

$$LR = \frac{\Pr(E|H_1, I)}{\Pr(E|H_2, I)}.$$

Different approaches and statistical models can be used within the LR framework. For DNA mixture interpretation, these include binary, discrete (semi-continuous), and continuous (fully continuous) models and approaches (e.g., Kelly et al. 2014, Bille et al. 2014).

2.5.2. LR Results, Transposed Conditionals, and Verbal Scales

Likelihood ratios are often thought of in terms of evidence scales. When an LR result is greater than one, the scale tips in the direction of having data that favor support if the hypothesis or proposition in the numerator (H_1) is true (Figure 2.3, left). When the LR result is less than one, the scale tips in the direction of having data that favor support if the hypothesis or proposition in the denominator (H_2) is true (Figure 2.3, right). The magnitude of the LR result is a reflection of how far the scale has tipped in support of one proposition over the other. An LR numeric value is not a measurement of a physical quantity. Rather, it is a ratio of probabilities and is dependent on the specific propositions used to formulate it.

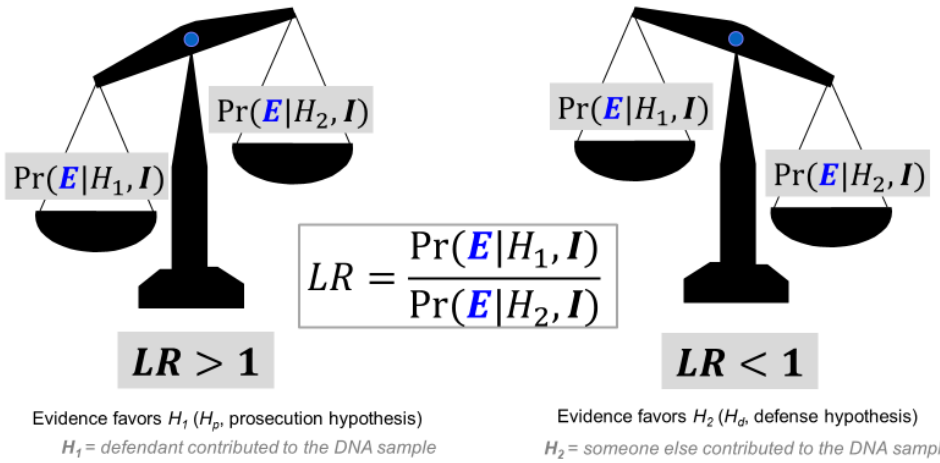

Figure 2.3. Illustration of likelihood ratio (LR) as a ratio of two likelihoods and tipping of scales. Abbreviations: E = evidence, H_1 = hypothesis (proposition) 1, H_2 = hypothesis (proposition) 2, I = information available, Pr = probability.

A common problem known as "transposing the conditional" (Evett 1995) or committing the "prosecutor's fallacy" (Thompson & Schumann 1987) can lead to a misunderstanding of the

meaning of an LR result. In these situations, a user confuses "the probability of the evidence given the propositions" with "the probability of the propositions given the evidence." This confusion comes from misinterpreting the conditional probabilities used: rather than Pr(E|H), or the probability of the evidence if (or given) the proposition is true, the terms are effectively reversed to Pr(H|E), or the probability of the proposition given the evidence.

A commonly used example illustrates the impact of transposing the conditional:

> "*The probability that an animal has four legs if it is a cow is one* does not mean the same thing as *the probability that an animal is a cow if it has four legs is one*" (Evett 1995).

If rewritten in symbols, Pr(four legs|cow) = 1 is not equivalent to Pr(cow|four legs) = 1. The second statement is false since horses, dogs, cats, and other animals also have four legs. Even the first statement, Pr(four legs|cow) = 1, assumes that rare situations of cows with missing limbs are not considered.

With DNA evidence, a statement such as "DNA evidence found on the item is one million times more likely to have come from Person X than anyone else" transposes the conditional. This statement emphasizes the proposition rather than the evidence. An appropriate way to report this LR result would be "DNA evidence found on the item is one million times more likely to be observed *if* the evidence came from Person X than *if* the evidence came from Person Y." The inclusion of the word "if" emphasizes the conditional probabilities and assumptions made in assigning the LR value. It is always the trier-of-fact's final decision whether the DNA originates from a specific person or not and the relevance of this information.

In an effort to describe the relative significance of their results, some forensic scientists use a verbal scale in conjunction with the LR to communicate the probative value of the evidence (e.g., Marquis et al. 2016). In their recent book *Forensic DNA Profiling: A Practical Guide to Assigning Likelihood Ratios*, authors Jo-Anne Bright and Michael Coble note (pp. 30-31):

> "There has been some justifiable criticism that LRs are not understood by our audience. The use of words to represent the strength of evidence has been proposed as a way to supplement numerical LR evidence. *The assignment of words to a numerical LR scale is, of course, arbitrary*…and there are a number of different scales used around the world for different jurisdictions" (Bright & Coble 2020; emphasis added; see also Thompson & Newman 2015).

A verbal scale recommended by the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios includes categories of *uninformative* (LR=1), *limited support* (LR = 2 to <100), *moderate support* (LR = 100 to <10,000), *strong support* (LR = 10,000 to < 1 million), and *very strong support* (LR > 1 million) (SWGDAM 2018). This SWGDAM verbal scale was adopted in September 2018 as part of the Department of Justice Uniform Language for Testimony and Reports for Forensic Autosomal DNA Examinations Using Probabilistic Genotyping Systems[10].

[10] https://www.justice.gov/olp/page/file/1095961/download

2.5.3. Probabilistic Genotyping Software

A number of software programs have been developed in recent years to assist analysts in performing DNA mixture interpretation by computing LR results using discrete or continuous approaches (Coble & Bright 2019, Butler & Willis 2020). Probabilistic genotyping software (PGS) systems utilize statistical genetics, biological models, computer algorithms, and probability distributions to infer possible genotypes and calculate LRs using either discrete or continuous approaches. Examples of discrete PGS systems include LRmix (Gill & Haned 2013), likeLTD (Balding 2013), Lab Retriever (Inman et al. 2015), or LiRa (Puch-Solis & Clayton 2014). Examples of continuous models include EuroForMix (Bleka et al. 2016a), STRmix (Taylor et al. 2013), and TrueAllele (Perlin et al. 2011).

A PGS system assists a DNA analyst with deconvolution of information in mixtures and provides an estimate of the statistical strength of evidence in the data and "stats" portion of the interpretation process illustrated in Figure 2.1. Weighted genotype possibilities can be estimated using Markov chain Monte Carlo (MCMC) simulations to assess possible combinations of parameters considered in deconvoluting potential contributor genotypes (e.g., Curran 2008, Buckleton et al. 2016, p. 287-293).

A PGS system computes LR values based on the information provided (Figure 2.4), including (1) *modeling choices* made by the system architect(s), (2) *data input choices* made by the analyst regarding an analytical threshold for calling peaks as alleles, selecting the number of contributors to the mixture for use in PGS calculations, and sometimes categorizing artifacts (e.g., pull-up peaks), (3) *proposition choices and assumptions* made by the analyst (e.g., use of unrelated individuals versus relatives, conditioning on a victim when analyzing an intimate sample, and underestimating or overestimating the number of contributors), and (4) *population database choices* used by the laboratory to provide allele and genotype frequency estimates including using or not using subpopulation correction and if using, what value is selected.

An increasing number of forensic laboratories are beginning to use PGS for DNA mixture interpretation. The UK Forensic Science Regulator shared seven perceived benefits of PGS compared to manual calculations (UKFSR 2018b, p. 8): (1) increased consistency within and between organizations utilizing the same software, (2) information available in the profile is used more efficiently, (3) deconvolution of genotypes enabling database searches that would not otherwise be feasible, (4) improved reliability due to increased automation in processing, (5) reduced variability between analysts in deciding whether peaks are true alleles or artifacts, (6) increased range of DNA profiles suitable for interpretation, and (7) publication of statistical models in peer-reviewed journals.

While PGS can assist in interpretation of complex DNA mixtures, "a computer program does not replace the need to think carefully about the case" (Gill et al. 2015). Thinking carefully about a case involves assigning an LR using propositions that address case-relevant questions.

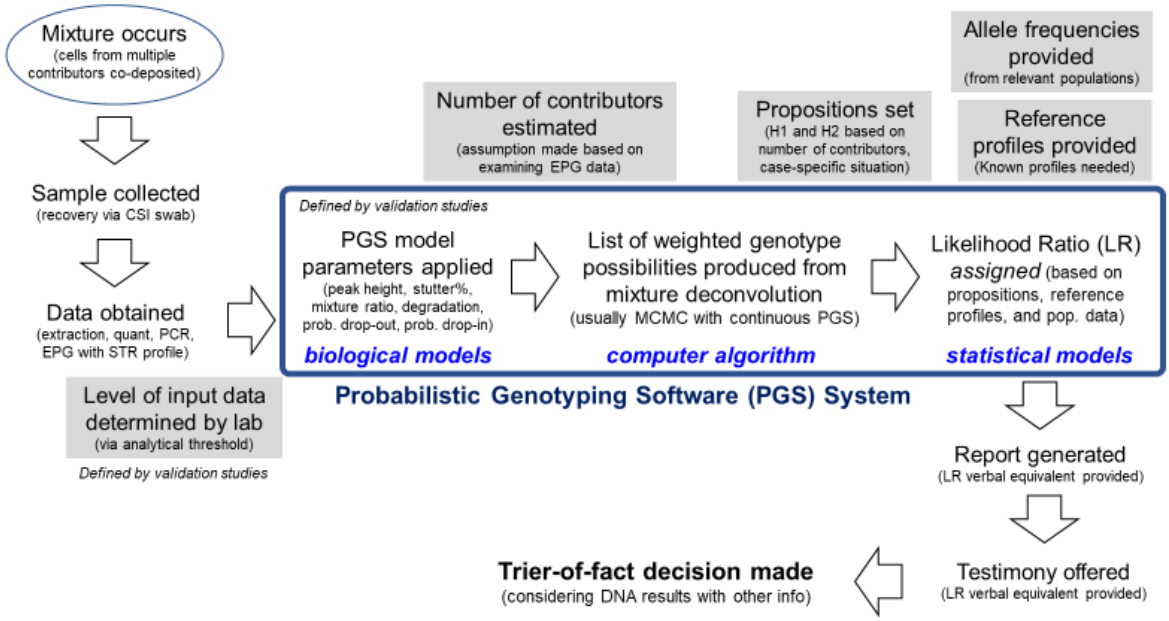

Figure 2.4. Illustration of aspects of a PGS system along with inputs needed (grey shaded boxes). Abbreviations: EPG = electropherogram, LR = likelihood ratio, MCMC = Markov chain Monte Carlo, PGS = probabilistic genotyping software, STR = short tandem repeat. Adapted from Butler & Willis 2020.

2.5.4. Propositions Impact LR Results

As noted by a group of statisticians and forensic scientists, selection of propositions is a vital part of LR assignment:

> "…the choice of these propositions depends on the case information and the allegations of each of the parties. This dependence is unavoidable for the forensic scientist to be able to accomplish his/her duty of presenting what the DNA results mean with regard to the issue of interest to the court" (Gittelson et al. 2016).

LR results vary when different propositions and assumptions are used. The guidance from the UK Forensic Science Regulator on DNA mixture interpretation emphasizes the need to record in the case file the reasoning used by the analyst to support the propositions selected (UKFSR 2018a). The magnitude of this variation can be observed with worked examples using the same data set (Table 2.4). With PGS, propositions are typically arranged as follows, assuming a number of contributors (*N*) who are unrelated to each other and to the POI:

H_1: POI + (*N*-1) unknown, unrelated contributors to the crime sample
H_2: *N* unknown, unrelated contributors to the crime sample

In Chapter 7 of the 2020 book *Forensic DNA Profiling: A Practical Guide to Assigning Likelihood Ratios* (Bright & Coble 2020), the authors provide detailed, worked examples using a two-locus DNA profile (involving D16S539 and D2S1338) with all observed alleles above the analytical threshold. Assuming two contributors, genotype weights were estimated using a PGS system. A person of interest was typed at these loci and could not be excluded

as a possible contributor to the mixture. Caucasian allele frequencies from a published data set (Moretti et al. 2016) were used in calculations performed. The same EPG data were examined under four different sets of propositions and assumptions. The LR results varied from over 4,000 (*moderate support* on SWGDAM 2018 verbal scale) to less than 10 (*limited support*) depending on the propositions and assumptions made (Table 2.4). These LR results were all determined at the sub-source level on the *hierarchy of propositions* (see Gill 2001, Taylor et al. 2018).

The highest LR result in Table 2.4 occurred when conditioning on the victim, meaning that the victim's genotypes are expected to be present at each locus in the EPG. This conditioning removes some ambiguity in the possible genotype combinations, which leads to a higher LR result for the POI under consideration.

Another possible source of variation in LR results comprises the estimated degree of co-ancestry in observed alleles, which involves using a subpopulation correction factor typically symbolized by the Greek letter theta (Balding & Nichols 1994, NRC 1996). Using different assumptions in the genetic model (e.g., without or with a 1% subpopulation correction, $\theta = 0.01$), the LR changes from 2895 to 1144.

Table 2.4. Summary of LR results from worked examples with two STR loci using different propositions and assumptions (information from Bright & Coble 2020). For information on NRC II 4.2, see NRC 1996.

Pages in book with worked example details	Summary of Propositions and Assumptions Used	LR Result
pp. 160-161	Conditioning on the victim	4143
pp. 148-150	Using the product rule ($\theta = 0$)	2895
pp. 150-153	Using NRC II 4.2 ($\theta = 0.01$)	1144
pp. 151,154-160	With possible untested brother	7.7

Finally, the lowest LR result in Table 2.4 comes from considering a possible untested brother rather than an unrelated individual in the assumptions made and calculations performed. Even considering only two loci, LR assignments can differ by several orders of magnitude.

Providing relevant answers depends on asking the right questions. In a review of the 1996 NRC II report (NRC 1996), several authors note:

> "At best DNA profiling can provide very strong evidence of association between people and places. It does not address ultimate questions of guilt or innocence" (Chambers et al. 1997).

Earlier in their article, these authors point out:

> "It should be accepted that there is now no dispute about the potential for DNA analysis to identify individuals, *subject to the constraints imposed by the quality of the evidential samples*" (Chambers et al. 1997, emphasis added).

More recently the following suggestion has been provided by a group of statisticians and forensic scientists:

> "The need to formalize one's propositions for assigning an LR may act as a beneficial restraint. If it is simply not possible to form propositions, then maybe the situation is beyond interpretation" (Gittelson et al. 2016).

DNA mixture interpretation is performed in the face of uncertainty. As noted by Ian Evett and Bruce Weir in their 1998 book:

> "The origins of crime scene stains are not known with certainty, although these stains may match samples from specific people. The language of probability is designed to allow numerical statements about uncertainty, and we need to recognize that *probabilities are assigned by people rather than being inherent physical quantities*" (Evett & Weir 1998, p. 21, emphasis added).

KEY TAKEAWAY #2.6: Likelihood ratios are not measurements. There is no single, correct likelihood ratio (LR). Different individuals and/or PGS systems often assign different LR values when presented with the same evidence because they base their judgment on different kits, protocols, models, assumptions, or computational algorithms. Empirical data for assessing the fitness for purpose of an analyst's LR are therefore warranted.

2.6. DNA Principles

This chapter concludes with a list of 16 important DNA principles. A principle has been defined as "a fundamental, primary, or general law or truth from which others are derived"[11]. An understanding of foundational principles can provide the basis for why something is important and can assist in deciding what actions should be taken in specific situations. The principles and concepts described here, which are not necessarily exhaustive, have been distilled out of various publications and aspects of DNA mixture interpretation. They are grouped by theme and ordered arbitrarily. With each principle, which is numbered and displayed in bold font, additional information is provided concluding with a statement in italics that describes why that principle is important to DNA interpretation.

We believe that a shared understanding of fundamental principles described in this chapter will benefit all stakeholders and help users of DNA information appreciate the potential and the limitations of DNA mixture interpretation (see Schneider et al. 2006a, Morling et al. 2007, Stringer et al. 2009). Training and continuing education can assist in acquiring this understanding (see Appendix 2). These principles are not new but may need to be re-emphasized because once a process becomes more complex, fewer people may understand the details and their origins.

Principle 1 [Biology]: Our DNA generally remains unchanged across time and cell type. Each cell of the human body contains DNA, except for mature red blood cells (Grasso & Woodard 1967). The DNA sequence and patterns found in the human genome of an individual generally remain unchanged over time (Jeffreys 1987). Likewise, DNA samples originating from the same individual will yield, with very rare exceptions, the same DNA profile independent of the type of cells examined (e.g., sperm vs. epithelial) (e.g., Cotton et al. 2000, Holt et al. 2002). Thus, a sample from an individual collected at different times over his/her lifetime is expected to yield equivalent DNA profiles. *This*

[11] https://www.dictionary.com/browse/principle

principle enables meaningful comparison of DNA from a reference sample to an evidence sample deposited and/or collected at a different time and to verify identity in a "biometric" sense, where a previously analyzed DNA profile is checked against a new one for "authentication" purposes.

Principle 2 [Biology]: DNA transfers and persists and can be collected and analyzed.

Human cells can be transferred to a surface through a variety of means, such as touching or coughing (van Oorschot & Jones 1997). DNA transfers and persists (e.g., van Oorschot et al. 2019) – and when collected and analyzed, can assist investigations. *This principle of direct or primary transfer enables results to be generated from evidentiary DNA profiles to assist in crime-to-crime and crime-to-individual associations.*

Principle 3 [Biology]: Forensic DNA profiles examine a limited number of specific sites in the human genome.

Current forensic DNA tests used in crime laboratories examine only a small portion of the human genome. A DNA profile comes from examining specific sites (*loci*) that are known to vary between individuals and do not code for genetic traits (Katsanis & Wagner 2013). Short tandem repeat (STR) markers, which possess multiple (e.g., 10 to 20) possibilities (*alleles*) that vary in the number of repeats, are the primary loci used today in forensic DNA tests (Butler 2007). The ability to distinguish DNA profiles from two unrelated individuals increases as more DNA sites are tested. *This principle is a reminder that the entire DNA sequence is not examined with forensic tests. Statistical assessments of profile rarity are used based on inheritance patterns and population genetics.*

Principle 4 [Genetics]: DNA passes from parent to offspring according to established genetic inheritance patterns.

Half of an individual's autosomal nuclear DNA comes from each of their biological parents. Each child can inherit different combinations of their parents' DNA (e.g., Roach et al. 2010). For this reason, the genetic characteristics shared among siblings can vary. Lineage markers, such those found on Y-chromosomes and mitochondrial DNA, typically pass from parent to offspring unchanged although an occasional mutation may occur (Kayser 2007). DNA results from biological relatives can be associated using the expected genetic inheritance patterns of various DNA markers. *This principle enables missing persons investigations, familial searching, relationship testing, and genetic genealogy.*

Principle 5 [Genetics]: Genetic inheritance patterns and population genetics enable strength of evidence statistical calculations.

A statistical weight can be calculated because of probabilities associated with genetic inheritance expectations. The statistical model for these population genetics calculations was described more than a century ago (Hardy 1908, Weinberg 1908) and is known as Hardy-Weinberg equilibrium (Crow 1999). The random match probability (RMP) is a measure of a DNA profile's rarity and reflects an estimate of the probability of drawing one individual with a specific DNA profile at random from a group of unrelated individuals in a population (NRC 1996). The rarity of a specific DNA profile can be

calculated using allele frequency estimates for individual markers along with sub-population adjustments and combining genotype frequency estimates across each marker deemed to be independent from other markers in the DNA profile (Balding & Nichols 1994). *This principle supports population frequency calculations made when a known is considered as a possible contributor to an evidence profile.*

Principle 6 [Genetics]: DNA profiles from close relatives are more similar than DNA from unrelated people.

DNA profiles from close relatives are expected to be more similar than DNA profiles from unrelated individuals (Li et al. 1993). There are a limited number of alleles at each locus, and even individuals who are not closely related will share alleles and genotypes. The frequency of occurrence of specific alleles and genotypes varies. *This principle is a reminder that while statistical models typically assume individuals are unrelated, if case context suggests closely related individuals may have contributed to the sample in question, then performing calculations assuming individuals are related may be helpful to decision makers.*

Principle 7 [Relevance]: Answers from DNA results depend on questions asked and circumstances of the evidence.

The FBI DNA Advisory Board stated: "Proper statistical inference requires careful formulation of the question to be answered. Inference must take into account how and what data were collected, which, in turn, determine how the data are analyzed and interpreted" (DAB 2000). DNA results typically address questions at the sub-source level of the hierarchy of propositions (i.e., who could be the source of the DNA or is the DNA from the person of interest, Taroni et al. 2013). *This principle is a reminder to users that DNA information by itself can only answer "who" questions, that is, questions of source not activity.*

Principle 8 [Measurement]: PCR amplification is a process needed to enrich the starting DNA material into measurable amounts. However, when small amounts of DNA are amplified, the results may not exactly represent the original DNA sample, including the relative quantities of each allele and genotype. In addition, the PCR process with STR alleles introduces artifacts, such as stutter products, that complicate interpretation of the resulting DNA profile.

PCR relies on replicating specified areas of the available DNA template to generate a detectable DNA profile at multiple STR markers. This DNA profile, which is depicted as an EPG, is influenced by DNA template amount and degradation level, the presence of inhibitors, and primer binding region sequence – all of which can influence the overall balance of the DNA profile. STR kits from different manufacturers may target slightly different regions of the same STR markers. PCR enables sensitive detection of even small amounts (e.g., 10 or fewer cells) of DNA, but also introduces artifacts such as stutter products into the test results that can influence the uncertainty of an interpretation (Gill et al. 2006b). *This principle is a reminder that STR results are a copy of the recovered DNA in a tested sample and depend on the accuracy and efficiency of the copying process. PCR artifacts increase uncertainty for the genotype possibilities of contributors to complex DNA mixtures.*

Principle 9 [Measurement]: Peak positions more accurately reflect allele calls than peak heights represent relative allele amounts.

Use of an internal size standard with each tested sample along with calibration to an allelic ladder enables accurate STR allele designations with electrophoresis separation and detection systems (e.g., Gill et al. 1997, Lazaruk et al. 1998). Peak heights and relative peaks heights, which do not use internal size standards to normalize stochastic variation, are not as reproducible as peak positions but do show trends by locus (e.g., Leclair et al. 2004, Debernardi et al. 2011). *This principle is a reminder that while alleles may be either present or absent (impacted by their peak heights and instrument detection thresholds), detected alleles are reproducible in terms of their designation (i.e., replicate testing does not show alleles shifting to a different allele, e.g., a "12" cannot become a "14" because peak position/sizing is stable).*

Principle 10 [Measurement]: Relative fluorescence unit (RFU) variance (uncertainty) is inversely proportional to DNA profile peak height.

Low peak heights are a function of starting amount and quality of the DNA template. When sufficient quality and quantity of DNA template exist, reliable and unambiguous DNA profiles can be generated from crime scene evidence. However, PCR amplification of low amounts of DNA template result in stochastic variation including severe peak imbalance of paired alleles in a genotype, allele drop-out, high stutter, and allele drop-in (Butler & Hill 2010). The chance of failing to replicate alleles that are present in the original sample during the PCR process, referred to as the probability of drop-out, increases when attempting to copy small amounts of DNA or highly fragmented DNA. Replicate amplification from aliquots of the same DNA extract have been used to improve the degree of reliability (Taberlet et al. 1996, Gill et al. 2000, Benschop et al. 2011). *This principle relates particularly to minor contributors in DNA mixtures.*

Principle 11 [Interpretation]: Although there is a single physical mixture ratio created at the time of deposition, it may be manifested differently at each tested locus due to stochastic variation in the PCR amplification process and potential variable DNA degradation across the contributors' genome sequences.

Stochastic variation in the PCR amplification process or sampling of template influences heterozygote balance and variation in mixture proportion (Bill et al. 2005). Assumptions are commonly made that allele peak heights are approximately linearly proportional to the amount of DNA prior to amplification and that contributions from two separate alleles are additive. Some studies have suggested that the estimated mixture proportion at each locus was highly variable at different loci within the same sample with variance at a locus from the overall profile estimate as high as 35% (Bill et al. 2005). *This principle emphasizes the need for interpretation methods or computer algorithms to account for variations in mixture ratios based on peak height variability and relative peak heights differences between loci in a DNA profile.*

Principle 12 [Interpretation]: Stutter products should be considered in interpretation when minor contributor alleles and stutter products of major alleles possess similar peak heights.

STR allele stutter products can complicate DNA mixture interpretation particularly when it comes to estimating the number of contributors. Depending on the ratio of contributor amounts in the mixture, peaks in the stutter position may need to be considered as possible alleles from a minor contributor (Gill et al. 2006b, Budowle et al. 2009). *This principle recognizes the impact of artifacts, such as STR allele stutter products, on mixture interpretation.*

Principle 13 [Interpretation]: Accurate estimates of the number of contributors to a DNA mixture are impacted by and may be underestimated when (a) the number of contributors increases, (b) the amount of DNA tested decreases, or (c) the degree of allele overlap in mixture contributors increases, such as when the contributors are related.

Estimating the number of contributors in a DNA mixture becomes more uncertain with more contributors (Paoletti et al. 2005, Buckleton et al. 2007, Coble et al. 2015). The more alleles observed at a tested locus, the greater the chance for allele overlap. As noted in Principle #6, biologically related contributors are expected to share alleles. When alleles overlap and are shared between contributors, it becomes more difficult to definitively estimate the number of donors to the DNA mixture. Missing alleles from true contributors can also impact estimation of the number of contributors. Low-quantity and low-quality DNA templates are subject to allele drop-out as well as stochastic variation that can skew normal stutter product amounts and heterozygote balance (Butler & Hill 2010). *This principle emphasizes that factors impacting sample complexity, such as allele sharing and allele drop-out, influence reliable estimates for the number of contributors to a DNA mixture.*

Principle 14 [Interpretation]: Mathematical models can provide a list of possible genotype deconvolutions with associated weights or probabilities for mixture components that cannot be physically separated. Continuous models use more information than discrete or binary approaches.

A DNA mixture arises when cells from multiple contributors are present in a sample. Following the extraction process, DNA from these cells commingles and mixes – and this mixture cannot be chemically separated into its original components. Instead, mathematical models are used on EPG data to deconvolute or infer possible genotype combinations for detectable contributors. Then an assessment can be performed of the strength of evidence whether a person of interest contributed to a mixed DNA profile or not. The inclusion of peak height information with continuous models increases the strength of evidence for true donors especially for major contributors (Taylor 2014, Slooten 2018). *This principle recognizes that continuous models involving allele peak height information can discriminate better between true contributors and non-contributors than discrete or binary approaches only involving allele information.*

Principle 15 [Statistics]: Different statistical approaches can produce different numerical results as they utilize different information and/or models and answer different questions.

Multiple statistical approaches have been used for DNA mixture interpretation. Questions addressed and information used by these approaches can differ (see Tables 2.2 and 2.3).

For example, different LR approaches will yield different results because these approaches may utilize different information (e.g., modeling different types of stutter products) or process the same information differently (e.g., using a log normal model versus a gamma model). Thus, the 2018 ISFG DNA Commission concludes: "*There are no true likelihood ratios, just like there are no true models.* Depending on our assumptions, our knowledge and the results we want to assess, different models will be adopted, hence different values for the LR will be obtained. It is therefore important to outline in our [reporting] statements what factors impact evaluation (propositions, information, assumptions, data, and choice of model)" (Gill et al. 2018, emphasis added). *This principle recognizes that answers obtained are dependent on information and statistical models utilized and questions asked (see also Principle #7).*

Principle 16 [Statistics]: Assessing the strength of evidence in favor a proposition (hypothesis) H_1 requires at least one other proposition (hypothesis) H_2. These propositions H_1 and H_2 are required to be mutually exclusive and exhaustive. Strength of evidence assessments depend on the framework of circumstances within which they are evaluated.

The three principles of evidence interpretation that were described in the 1998 book by Ian Evett and Bruce Weir (Evett & Weir 1998, pp. 23-29) and restated in the 2020 book by Jo-Anne Bright and Michael Coble (Bright & Coble 2020, pp. 23-24) are combined here. *Principle 1*: To evaluate the uncertainty of any given proposition, it is necessary to consider at least one alternative proposition. *Principle 2*: Scientific interpretation is based on questions of the kind: "What is the probability of the evidence given the proposition?" *Principle 3*: Scientific interpretation is conditioned not only by the competing propositions, but also by the framework of circumstances within which they are to be evaluated. The framework of circumstances includes the hierarchy of propositions with offense, activity, source, sub-source, and sub-sub-source levels (Cook et al. 1998b, ENFSI 2015, Taylor et al. 2018, Gill et al. 2018, Gill et al. 2020a). *This principle emphasizes the foundational elements of the likelihood ratio framework.*

3. Chapter 3: Data and Information Sources

This scientific foundation review seeks to document and independently assess the empirical evidence that supports the reliable use of DNA mixture interpretation methods. The sources of data and information used in conducting this review are described in this chapter. These sources include (1) peer-reviewed articles appearing in scientific journals, (2) published interlaboratory studies, (3) laboratory internal validation studies that are accessible online, and (4) proficiency test data available on test provider websites.

3.1. Information Sources

This scientific foundation review focused on DNA mixture interpretation involving autosomal short tandem repeat (STR) markers. To assess reliability and relevance issues related to DNA mixture interpretation, we sought empirical data from a variety of publicly available sources.

The resources we examined include (1) publications in the peer-reviewed scientific literature and (2) data or information located on the internet, such as proficiency test (PT) results from PT provider websites or publicly available internal validation data summaries from individual laboratories. PT data provide insights into how individual analysts performed on specific tests while internal validation studies offer insights into how laboratories performed when analyzing a range of DNA mixtures of varying complexity. Published interlaboratory studies enable an important assessment of analyst and laboratory performance. This is because the same samples and/or data are evaluated among the participants to examine reproducibility and reliability across methods.

By searching and studying the peer-reviewed literature on forensic DNA, we collected and examined articles on DNA mixture interpretation and DNA transfer studies.

We recognize that there are information and data collected in forensic laboratories that may not yet be publicly available or published. However, we believe for information to be considered foundational, it needs to be reasonably accessible to anyone who wishes to review it.

3.1.1. Peer-Reviewed Publications

We performed a literature search on articles related to DNA mixture interpretation using PubMed (https://www.ncbi.nlm.nih.gov/pubmed), Google Scholar (https://scholar.google.com/) and Web of Science (http://apps.webofknowledge.com/). Knowledge distilled from the examination of these articles informed the entire report.

As part of our review, we specifically examined titles and abstracts for articles in the following journals: *Journal of Forensic Sciences*, *Forensic Science International*, *Forensic Science International: Genetics*, *Science & Justice*, *Legal Medicine*, *Australian Journal of Forensic Sciences, Electrophoresis*, *International Journal of Legal Medicine*, and *Forensic Science Medicine and Pathology*. In addition, we considered over 1500 extended abstracts

published in the 2009, 2011, 2013, 2015, 2017, and 2019 *Forensic Science International: Genetics Supplement Series,* representing the proceedings of the biennial meetings of the International Society for Forensic Genetics.

Search parameters impact the number and types of articles that can be located on any particular topic. The challenge of locating relevant articles is illustrated in Table 3.1, which contains a summary of PubMed searches for articles containing the words "DNA" and "mixture" in the text.

The number of articles listed for each entry in Table 3.1 corresponds to the year of print, rather than electronic publication. For example, a PubMed search using dates between January 1, 2009 and December 31, 2009 with search terms "Forensic Science International Genetics" along with "DNA" and "mixture" provides six search results, yet three were electronic publications that were published in print in 2010. In Table 3.1 this example is highlighted in red font. An examination of the remaining three articles in this example finds only one that falls in the scope of this review (Cowell 2009), as the other two describe Y-chromosome STR analysis or tri-allelic single nucleotide markers (SNP) markers.

Table 3.1. Numbers of articles published with "DNA" and "mixture" in the text across the listed forensic science journals from 2009 to 2018 based on PubMed searches (https://www.ncbi.nlm.nih.gov/pubmed/) conducted May 10, 2019.

Journal	Total	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
PLoS ONE	**187**	7	7	15	26	40	30	20	11	13	18
Forensic Sci. Int. Genet.	**135**	**3**	3	7	12	4	22	16	15	26	27
Int. J. Legal Med.	**30**	2	2	0	3	3	2	2	4	5	7
J. Forensic Sci.	**27**	4	3	6	3	2	1	4	1	1	2
Electrophoresis	**25**	5	3	1	1	1	2	4	3	2	3
Sci Justice	**11**	1	1	2	2	2	0	0	0	1	2
Legal Med.	**9**	0	1	2	1	1	0	0	2	1	1
Forensic Sci. Int.	**4**	0	0	1	0	0	0	0	1	1	1
For. Sci. Med. Pathol.	**2**	0	1	0	0	0	0	0	0	1	0
TOTAL	**430**	**22**	**21**	**34**	**48**	**53**	**57**	**46**	**37**	**51**	**61**

Table 3.1 illustrates a steady stream of new literature and is a reminder that information gathered to compile this report on DNA mixture interpretation represents a snapshot in time. The PubMed search results reported in Table 3.1 are missing some relevant publications (e.g., ones cited in this report bibliography) within these journals or in other journals not listed. In addition, many of the search results provided articles that have "DNA" and "mixture" within the text but are not relevant to DNA mixture interpretation involving autosomal STR markers. This is the case with many of the *PLoS ONE* articles.

By examining online search results, we identified publications dealing specifically with DNA mixtures and aspects of DNA interpretation. Each located article was first assessed by reviewing the title and abstract. Articles of interest were downloaded and studied further. We also examined citation lists in the articles we examined to see whether a relevant article may have been missed in initial searches. Information used in Chapter 5 regarding DNA transfer studies was located with similar types of search strategies. Hundreds of relevant articles were

collected and are cited throughout this report. However, an original goal of this project – to develop a comprehensive, curated bibliography on DNA mixtures – proved unfeasible as a result of the constantly growing literature.

3.1.2. Available Internal Laboratory Data

Forensic laboratories conduct internal validation experiments before implementing a new technique to assess method performance under specific conditions. Data from these studies are not typically shared outside the laboratory except in response to a discovery request connected to a specific legal proceeding. With an understandable focus on casework production in forensic laboratories, information from internal validation studies or related research experiments may not be prepared in a manner conducive to sharing with a wider community. Even if prepared, manuscripts reporting internal validation analysis are unlikely to be considered unless they provide a new insight that has not been previously reported. We performed Google searches for data from internal validation studies searching for the state, city, and agency (if known) and the phrase "forensic DNA laboratory validation data." We then reviewed laboratories' public websites for available standard operating procedures (SOPs) and/or validation documents. Eight laboratory probabilistic genotyping software (PGS) internal validation summaries were located on https://johnbuckleton.wordpress.com/strmix/strmix-validations/.

Internal validation summaries from eight U.S. forensic laboratories were located with our online searches (Table 3.2). Generally speaking, we have found that sufficient data of this sort are not publicly available for an independent assessment of reliability (see Chapter 4). Some laboratories provide summary information from their validation studies, but detailed data are often unavailable, in part because of privacy concerns around releasing genotype information from individuals. The same is true for most peer-reviewed articles that describe validation experiments.

Information included in these summaries is related to the PGS system being validated and the types of DNA mixture samples being used. However, we recognize that additional internal validation data likely exists within individual laboratories. This scientific foundation review is limited to publicly available information.

Table 3.2. Publicly available internal validation data from forensic laboratories located in Google searches performed March 23, 2020. Updated February 8, 2021. See Table 4.5 for analysis of DNA mixtures examined.

Laboratory	Information Available and Website
California Department of Justice DNA Laboratory (Richmond, CA)*	STRmix v2.06 (Identifiler Plus, ABI 3130/3500) https://epic.org/state-policy/foia/dna-software/EPIC-16-02-02-CalDOJ-FOIA-20160219-STRmix-V2.0.6-Validation-Summaries.pdf

Laboratory	Information Available and Website
Erie County Central Police Services Forensic Laboratory (Buffalo, NY)	STRmix v2.3 (PowerPlex Fusion, ABI 3500) https://johnbuckleton.files.wordpress.com/2016/09/strmix-implementation-and-internal-validation-erie-fusion.pdf STRmix v2.3 (Identifiler Plus, ABI 3500) https://johnbuckleton.files.wordpress.com/2016/09/strmix-implementation-and-internal-validation-erie-id-plus.pdf
Michigan State Police (Lansing, MI)	STRmix v2.3.07 (PowerPlex Fusion, ABI 3500/3500xl) https://johnbuckleton.files.wordpress.com/2016/09/strmix-summary-msp.pdf
Office of Chief Medical Examiner Forensic Biology Laboratory (New York City, NY)	STRmix v2.4 (PowerPlex Fusion, ABI 3130xl) https://www1.nyc.gov/site/ocme/services/validation-summary.page
Palm Beach County Sheriff's Office (West Palm Beach, FL)	STRmix v2.4.06 (PowerPlex Fusion, ABI 3500xl) http://www.pbso.org/qualtrax/QTDocuments/4228.PDF STRmix v2.6.2 (PowerPlex Fusion 6C, ABI 3500xl) https://www.pbso.org/qualtrax/QTDocuments/10787.PDF
San Diego Police Department Crime Laboratory (San Diego, CA)	STRmix (GlobalFiler, ABI 3500), STRmix v2.3.07; STRmix v2.4.06 https://www.sandiego.gov/police/services/crime-laboratory-documents
Virginia Department of Forensic Science (Richmond, VA)*	TrueAllele Casework (PowerPlex 16, ABI 3130xl) https://epic.org/state-policy/foia/dna-software/EPIC-15-10-13-VA-FOIA-20151104-Production-Pt2.pdf
Department of Forensic Sciences (Washington, DC)	STRmix v2.3 parameters & validation report (Identifiler Plus, ABI 3500) https://dfs.dc.gov/page/fbu-validation-studiesperformance-checks STRmix v2.4 parameters & validation report (GlobalFiler, ABI 3500) https://dfs.dc.gov/page/fbu-validation-studiesperformance-checks

*Information available online via a Freedom of Information request by the Electronic Privacy Information Center (epic.org)

3.1.3. Available Proficiency Test Data

Proficiency test (PT) data can also be useful when assessing the reliability of DNA mixture interpretation methods. The DNA Identification Act of 1994 and the FBI Quality Assurance Standards require semiannual proficiency tests for all DNA analysts working in a U.S. laboratory that receives federal funding or supply data to the national DNA database (DNA Identification Act 1994, QAS 2020). Over the years, a variety of DNA mixture tests have been provided to participating forensic DNA analysts. In the United States, PT providers offering DNA mixture tests include Collaborative Testing Services (Sterling, VA), Bode Technology (Lorton, VA), and Forensic Assurance (Northville, MI). In addition, the German DNA Profiling Group (GEDNAP) provides DNA proficiency tests for many laboratories in Europe. PT provider websites were searched for available information.

3.1.3.1. CTS Forensics

Collaborative Testing Services, Inc. (CTS; Sterling, VA; https://cts-forensics.com/), offers several DNA mixture proficiency tests. Participants are rated on their ability to return results that agree with a consensus result.

CTS has reported a steady state of enrollment and about 80% return rates for their DNA PT exams from 2004 to 2015 (Kolowski et al. 2016). Currently, CTS offers a DNA mixture test twice a year (5801 and 5806 series) and a DNA interpretation test (588 and 589 series) twice a year. The CTS forensic biology tests (until 2017 the 571, 572, 573, 574, 575, and 576 series and since 2017, the 5701, 5702, 5703, 5704, 5705, and 5706 series) also contain mixtures of human whole blood and semen.

The DNA mixture test samples contain two known bloodstains provided on Whatman FTA cards or clean white fabric, and two questioned stains where one or both contains a mixture of body fluids. This is typically blood and semen mixed in a 1:1 volume ratio before applying to the substrate (see Chapter 4).

The CTS DNA interpretation tests are intended for the technical reviewers and consultants who may not have access to laboratory equipment or data analysis software. These tests are distributed via digital download in the form of electropherogram files (.pdf, .fsa, or .hid formats) with results from a variety of common autosomal and Y-STR typing kits. Participants with the DNA interpretation study evaluate and report DNA profiles of four samples, consisting of two known and two question samples, using their existing protocols. Mixtures present in question samples are usually two-person and sometimes three-person mixtures with components in the range of 1:1 to 1:4 or 2:1:1 or 3:1:1 (mixed by body fluid volume rather than predetermined DNA quantity).

In Chapter 4 of this report, we provide a summary of CTS DNA mixture data sets along with analysis of their contents.

3.1.3.2. Bode Technology

Bode Technology, formerly known as Bode Cellmark Forensics (Lorton, VA), offers International Quality Assessment Scheme (IQAS) PT kits (https://bode-labs.com/iqas). Two kits (IQAS-50 and IQAS-60) provide the ability to assess DNA mixture interpretation results from a simple mixture of semen and white blood cells. Summary reports of participant results are provided to the ANSI-ASQ National Accreditation Board (ANAB).

We did not find these PT results or reports to be publicly available for our examination or review.

3.1.3.3. Forensic Assurance

In an effort to provide PT samples that are more like casework situations, Forensic Assurance (Northville, MI; https://forensicassurance.com/) has begun offering a PGS proficiency test. Their design includes supplying data files for two evidentiary mixture samples (two-, three-, or four-person mixtures) and four known reference samples. Participants are required to estimate the number of contributors in the mixture profiles and compare the reference profiles to the mixture profiles using their laboratory's PGS and interpretation protocols. Participants return their likelihood ratio (LR) value for each comparison along with the

propositions used and a determination of which proposition is favored (i.e., H_1 versus H_2 or the numerator versus the denominator in their LR calculation).

We did not find these PT results or reports to be publicly available for our examination or review.

3.1.3.4. GEDNAP Studies

The German DNA Profiling Group (GEDNAP; https://www.gednap.org/) provides regular DNA PT exams for quality-assurance purposes (Rand et al. 2002, Rand et al. 2004). A GEDNAP "Stain Commission" designs the studies, which commonly contain challenging samples and mixtures. Each GEDNAP PT consists of three reference samples and four "stains" designed to mimic crime scene samples. Samples are prepared and sent out twice a year from a DNA laboratory in Münster, Germany. Each February, an annual Stain Workshop meeting is held ("Spurenworkshop" in the German language) to review the overall results obtained in the two studies from the prior year.

Typical errors are examined in an anonymous fashion to encourage quality improvements. Successful laboratories receive proficiency certificates. Over 200 laboratories from more than 40 different countries regularly participate in the GEDNAP PT DNA studies. Correct results are shared with each participating laboratory along with their score and a summary of any errors made.

We did not find these PT results or reports to be publicly available for our examination or review.

3.1.4. Interlaboratory Studies on DNA Mixture Interpretation

Interlaboratory studies provide an opportunity to assess variations across laboratory protocols and can be useful barometers regarding the reproducibility and reliability of various approaches.

Nineteen interlaboratory studies examining various aspects of DNA mixture interpretation and performance (see Chapter 4) have been conducted over the past two decades. These studies have been conducted by researchers at the National Institute of Standards and Technology, the Spanish-Portuguese Working Group of the International Society for Forensic Genetics, the European Forensic Genetics Network of Excellence, the UK Forensic Science Regulator, the Defense Forensic Science Center, the Netherlands Forensic Institute, and developers of the STRmix PGS system. Most of these studies have been published (see citations in Chapter 4).

3.1.5. Available Research Data Sets

Research data sets have been produced to aid current and future DNA mixture studies. The largest and most widely used to date is the PROVEDIt (Project Research Openness for Validation with Empirical Data) data set maintained by Professor Catherine Grgicak at

Rutgers University, which contains almost 25,000 DNA profiles (Alfonse et al. 2018). Table 3.3 summarizes the PROVEDIt data set, which contains DNA profiles amplified with three STR kits (Identifiler Plus, PowerPlex 16HS, and GlobalFiler) and analyzed on two capillary electrophoresis (CE) platforms (ABI 3130 and ABI 3500). These data were generated under 144 laboratory conditions and are classified by total DNA amount, DNA treatment, contributor numbers, and mixture proportions.

Table 3.3. Summary of PROVEDIt data set collected by researchers at Boston University and Rutgers University. Available at https://lftdi.camden.rutgers.edu/provedit/files/.

Sample Preparation	Data Set	STR Kit	# PCR Cycles	CE	# Profiles	Single-Source	2p Mixture	3p Mixture	4p Mixture	5p Mixture
DNA extract mixtures	RD12	Identifiler Plus	29	ABI 3500	3212	2280	366	209	147	210
DNA extract mixtures	RD12	PowerPlex 16HS	32	ABI 3130	1024	795	57	52	60	60
Whole blood mixtures	RD14	Identifiler Plus	28	ABI 3130	10,261	8267	524	487	520	463
Whole blood mixtures	RD14	GlobalFiler	29	ABI 3500	10,195	8190	526	484	527	468
				TOTAL	24,692	19,532	1473	1232	1254	1201
				ABI 3500	13,407					
				ABI 3130	11,285					

The PROVEDIt data can be downloaded as raw data (.fsa and .hid files) or exported genotypes table (.csv files) from the Laboratory for Forensic Technology Development and Integration (LFTDI; https://lftdi.camden.rutgers.edu/provedit/files/). Among the 5160 mixture profiles, ranging from two-person (2p) up to five-person (5p) profiles, 76% contain a contribution of at least one individual of less than 20% of the total DNA content. Many of the samples, which were prepared with 37 different genotype combinations, were subjected to PCR inhibitors or purposely degraded to produce partial profiles (Alfonse et al. 2018).

The funding to generate this data set represents a substantial and important investment by the U.S. government over multiple years. In their article describing the PROVEDIt data set, the authors express their hope that "a large dataset would play a critical role in demonstrating the foundational validity and robustness of new or existing DNA identity testing technology" (Alfonse et al. 2018). Samples from the PROVEDIt data set have been used in PGS comparisons (e.g., Riman et al. 2019b) and interlaboratory studies (e.g., Bright et al. 2019a).

4. Chapter 4: Reliability of DNA Mixture Measurements and Interpretation

This chapter considers foundational issues related to reliability of DNA mixture interpretation. Reliability centers on trustworthiness established through empirical assessments of available data to evaluate the degree of reliability of a system or its components. The degree of reliability of a system can be assessed through validation data, interlaboratory studies, and proficiency tests. To enable effective use of any information, responsibilities exist with both providers and users of that information. We use the term "factor space" to describe the factors that influence complexity, measurement, and interpretation reliability – these factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested. Available data from published or publicly accessible validation studies, proficiency tests, and interlaboratory studies are examined; limitations of available information and factor spaces assessed are considered. This information includes data from 60 published articles and 11 internal validation summaries involving probabilistic genotyping software, 7 years of proficiency test data involving more than 100,000 comparisons, and 18 interlaboratory studies over the past 2 decades. We note that the degree of reliability of a DNA mixture interpretation system, such as a DNA analyst using a probabilistic genotyping software program, depends on sample complexity. Results cannot be simply summarized into "reliable" or "unreliable" without considering context of the factor space explored and supporting validation data using ground truth samples of similar complexity. We also emphasize that proficiency tests need to be representative of complex DNA mixtures seen in casework if these tests are intended to assess analysts' ability to conduct dependable DNA mixture interpretation.

4.1. Introduction to Reliability

The 'plain English' meaning of the word *reliability* is trustworthiness, which is determined by the degree with which a result is consistently accurate.[12] This is the sense in which we use the term reliability in our report. Reliability implies consistency, but consistency of repeated measurements alone does not indicate reliability. Reliability requires being consistently accurate. The word reliable can sometimes be treated as though it has a binary meaning (i.e., something is reliable or not reliable). However, from a scientific perspective, it is more appropriate to speak in terms of a degree of reliability, reflecting the frequency with which a result is accurate.[13]

An important hallmark of science is to develop reliable theories and methods *based on empirical data,* so that users of scientific knowledge or methods can have a high degree of trust in its claims, results, or predictions. Reliability is born out of demonstrations of accuracy along with logical inference where appropriate. Logic can lead an analyst from a set of initial assumptions to final conclusions; but logic, by itself, cannot and should not support the initial assumptions. Logic is a necessary component in the conduct of science, but

[12] Oxford Dictionary (https://www.lexico.com/en/definition/reliability): a) The quality of being trustworthy or of performing consistently well. b) The degree to which the result of a measurement, calculation, or specification can be depended on to be accurate.
[13] We recognize that in legal settings, binary decisions (e.g., guilty or not guilty) need to be made. However, our focus is on the nonbinary scientific aspects of reliability rather than the binary legal ones.

empirical knowledge is what allows for trust in both the initial assumptions as well as in the resulting claims.

In their September 2016 report, the President's Council of Advisors on Science and Technology (PCAST) associated reliability with test results that have been demonstrated to be repeatable, reproducible, and accurate (PCAST 2016, p. 47). PCAST used the phrase "foundational validity" to reflect whether something was based on reliable principles and methods and "validity as applied" to reflect whether the individual performing the work was applying these principles and methods reliably (PCAST 2016, pp. 42-66). In this chapter, we explore the basis for reliability in DNA mixture measurements and interpretation with a focus on what PCAST termed foundational validity.

It is generally accepted that measurement and interpretation of high-template, high-quality, single-source DNA samples have a high degree of reliability (NRC 2009, PCAST 2016). This reliability comes from testing and observing consistently accurate results when assigning allele pairs into genotypes. At the other extreme, measurement and interpretation of samples involving a large number of contributors, consisting of very small amounts of DNA from some, make it harder to assign allele pairs for specific contributors without ambiguity and uncertainty (e.g., Benschop et al. 2012, Benschop et al. 2015a, Taylor & Buckleton 2015). This is likely the reason some laboratories adopt a policy of not interpreting highly complex mixtures (e.g., more than three contributors).

In this chapter, we review available data, concepts, and methods for assessing reliability of DNA mixture measurement and interpretation systems. Reliability relates to the whole system – not just a portion of the process, such as the performance of a software program used as part of DNA mixture interpretation system.

4.1.1. System Reliability vs. Component Reliability

With current laboratory methods, it is impossible to physically separate the DNA within a complex mixture into its constituent parts. To interpret a DNA mixture, an analyst uses their best judgment to estimate the number of contributors based on the observed DNA profile and then proceeds as described in Chapter 2 (see Figure 2.2).

The process of DNA evidence analysis (see Figure 2.1) can be divided into two major steps: (1) ***measurements*** of relative abundances of PCR products in a tested DNA sample that are displayed as an electropherogram (EPG), and (2) ***interpretation*** involving use of the EPG data to make a strength of evidence assessment when an evidentiary DNA profile is compared to a person of interest (POI). The outcome of interpretation includes a numeric output in the form of a likelihood ratio (LR). In recent years, DNA analysts have increasingly relied on one of several available probabilistic genotyping software (PGS) systems to assign a numerical value to their mixture result based on a pair of propositions selected by the analyst (see Chapter 2 and Appendix 1). Some PGS are proprietary and others are open-source.

The reliability of the entire process – starting from sample acquisition, to its analysis and generation of an EPG, and ending with an interpretation of results and expressing the strength of evidence in the form of an LR value – is of interest to the stakeholders in criminal proceedings. We refer to this as *system reliability*.

After conducting an internal validation study to establish parameter values to be used with the laboratory-selected PGS system, the interpretation step can be further divided into the following sub-steps:

(a) curating the EPG (removal of PCR artifacts, determining which peaks are allelic and which are not, etc.),
(b) estimating the apparent number of contributors,
(c) submitting the curated EPG to the PGS system and checking the output using various diagnostic analyses to ensure the result makes sense, and
(d) reporting a strength-of-evidence value in the form of an LR for a specific pair of propositions.

Each step or sub-step within the system may also be subjected to a reliability assessment. Reliability of any particular step in the entire system is referred to as *component reliability*. Component reliability is of interest particularly when exploring opportunities for improving the overall system reliability.

4.1.2. Definitions of Measurement, Uncertainty, Assessment, and Interpretation

In a guide for evaluating and expressing measurement results, NIST Fellow and Chief Statistician, Antonio Possolo, defines measurement, measurement uncertainty, and measurement result as follows:

> "**Measurement** is an experimental or computational process that, by comparison with a standard, produces an estimate of the true value of a property of a material or virtual object or collection of objects, or of a process, event, or series of events, together with an evaluation of the uncertainty associated with that estimate, and intended for use in support of decision-making" (Possolo 2015, p. 12).
>
> "**Measurement uncertainty** is the doubt about the true value of the measurand [property intended to be measured] that remains after making a measurement. Measurement uncertainty is described fully and quantitatively by a probability distribution on the set of values of the measurand. At a minimum, it may be described summarily and approximately by a quantitative indication of the dispersion (or scatter) of such distribution" (Possolo 2015, p. 14).
>
> Chapter 2.5 in Possolo's guide emphasizes: "The evaluation of measurement uncertainty is an essential part of measurement because it delineates a boundary for the reliability (or trustworthiness) of the assignment of a value (estimate) to the measurand and suggests the extent to which the measurement result conveys the same information for different users in different places and at different times (Mari & Carbone 2012). For this reason, a

> **measurement result** comprises both an estimate of the measurand and an evaluation of the associated uncertainty" (Possolo 2015, p. 13).

Since definitions for assessment and interpretation were not found in the NIST guide on measurement results, we turned to the Merriam-Webster dictionary. **Assessment** is "the action or an instance of making a judgment about something; the act of assessing something."[14] **Interpretation** is "the act or the result of interpreting"[15] where the definition of interpret includes "(1) to explain or tell the meaning of; to present in understandable terms, or (2) to conceive in the light of individual belief, judgment, or circumstance"[16].

In the context of DNA mixture interpretation using PGS (see Chapter 2 in this report), a DNA analyst assesses the probability of the findings if one proposition (H_1) were true and also the probability of the findings if another proposition (H_2) were true. This assessment is typically accomplished with the help of specialized knowledge of the discipline, training and experience, and the assistance of statistical models and computer programs.

A forensic scientist's evidential assessments may be summarized in the form of a numerical value called the likelihood ratio. LR assessments, which involve a ratio of two probabilities, do not involve comparison to any reference standard. Assertions have been made that there is no true LR (e.g., Steele & Balding 2014, Gill et al. 2018). Some even hold the view that there is no uncertainty associated with an LR assessment (Berger & Slooten 2016; see also Biedermann et al. 2016a, Curran 2016, Morrison & Enzinger 2016, Taylor & Balding 2020).

Although evidence assessments and interpretation have a greater subjective component than measurements do, the concept of reliability applies equally to assessments and interpretations as well as to measurements. This is not a new idea. As Ian Evett and Bruce Weir summarized in their 1998 book *Interpreting DNA Evidence*: "The interpretation of DNA evidence has to be made in the face of uncertainty. The origins of crime scene stains are not known with certainty, although these stains may match samples from specific people. The language of probability is designed to allow numerical statements about uncertainty, and *we need to recognize that probabilities are assigned by people rather than being inherent physical quantities*" (Evett & Weir 1998, p. 21, emphasis added).

4.1.3. Empirical Assessments of Reliability

Reliability is a term that can be meaningfully applied to any process or method for accomplishing a task or a goal. It also applies to any claim, opinion, quantitative assessment, or measurement result. In each instance, the focus is on the degree of trustworthiness. In this chapter, our interest is in the reliability of the system that is used to measure DNA samples and interpret the results by making quantitative assessments on the strength of the evidence.

[14] https://www.merriam-webster.com/dictionary/assessment
[15] https://www.merriam-webster.com/dictionary/interpretation
[16] https://www.merriam-webster.com/dictionary/interpret

Empirical assessments of reliability require that the process of interest be tested in ground-truth[17] known situations. For DNA mixture interpretation, this means that samples with known genotypes, known number of contributors, known mixture ratios, known degrees of degradation, etc., have been tested using the process of measurement and interpretation, and results from such tests are available to provide the basis for stakeholders to assess the degree of reliability of the process. Empirical assessments of the degree of reliability can be made from developmental and internal validation experiments (method-focused), proficiency tests (analyst-focused), and interlaboratory studies (community-focused). Each type of assessment addresses different questions.

Systematic approaches for analyzing the results of validation studies, using statistical tools for summarization and visualization, and relevant concepts such as accuracy, bias, precision, and calibration, are discussed in various textbooks (e.g., Vosk & Emery 2014). For example, histograms are a convenient way to visualize the statistical distributions of measurement variation when the quantity being measured is *continuous* (i.e., a real number versus a count in a histogram bin) and a sufficient number of data points are available.

Common numerical summaries for statistical distributions of variation include their average values and their standard deviations. At the other extreme, when the quantity of interest is binary (e.g., whether a proposition is true or false), differences from the expected value are summarized using error rates, which involve calculating a percentage of the times *true* is incorrectly classified as *false* (false negative errors) or *false* is incorrectly classified as *true* (false positive errors).

The 2016 PCAST Report emphasized that "the *only* way to establish scientifically that an examiner is capable of applying a foundationally valid method is through appropriate empirical testing to measure how often the examiner gets the correct answer" (PCAST 2016, p. 57, emphasis in the original). This point was reiterated in the January 2017 *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*: "While scientists may debate the precise design of a study, there is no room for debate about the absolute requirement for empirical testing. Importantly, the test problems used in the empirical study define the specific bounds within which the validity and reliability of the method has been established (e.g., is a DNA analysis method reliable for identifying a sample that comprises only 1% of a complex mixture?)" (PCAST 2017, p. 2). The answer to PCAST's question depends on which laboratory conducted the test and what their internal validation results can support.

Again, from the 2017 PCAST Addendum: "Forensic scientists rightly cite examiners' experience and judgment as important elements in their disciplines…However, experience and judgment alone – no matter how great – can *never* establish the validity or degree of reliability of any particular method. Only empirical testing of the method can do so" (PCAST 2017, p. 3, emphasis in the original).

Later in this chapter, a few tools are discussed that are particularly useful in the context of assessing reliability of DNA mixture measurement and interpretation. An understanding of

[17] Ground-truth requires knowing the correct answer before testing is performed and therefore is not possible with samples arising from crime-scene evidence.

these concepts can help in the design of studies for collecting information relevant for reliability assessments of measurements and interpretations.

4.1.4. Factor Space and Factor Space Coverage

The overall reliability of DNA mixture measurement and interpretation is influenced by many things. We use the term *factor space* to describe the totality of scenarios and associated variables (*factors*) that are considered likely to occur in actual casework. While this totality of scenarios and variables may never be fully known or explored, previous casework experience encountered by forensic DNA laboratories permits an approximate collection of possible scenarios to guide validation studies performed.

Factors influencing DNA mixture measurement and interpretation include (a) STR kits, instruments, and PCR parameters used, (b) actual or apparent number of contributors, (c) degradation levels of DNA from contributors, (d) mixture ratios of DNA from contributors, (e) total DNA template amount, (f) relatedness of potential contributors and degree of allele sharing, (g) statistical models used to perform interpretation, etc. See Table 4.1 for a more complete (but not exhaustive) list of factors.

Table 4.1. Factor space that influences DNA mixture measurements and interpretations with probabilistic genotyping software (PGS) systems. See also Table 2.1.

Portion of Factor Space	Influencing Factors
Measurement of STR Alleles and Genotypes	• Peak position for short tandem repeat (STR) alleles • Peak morphology or resolution for STR alleles • Peak height for STR alleles • Relative peak heights for STR allele pairs • Presence of stutter products and their relative heights compared to associated STR alleles
Sample Complexity	• Number of contributors, degree of allele sharing among contributors, and presence of stutter products • Total DNA template and contributor template amounts • Mixture ratio of DNA from contributors • Sample quality including degree of degradation • Presence of stutter products and potential minor contributors in a DNA mixture

Portion of Factor Space	Influencing Factors
Laboratory Specific Decisions	• STR typing kit(s) used • Capillary electrophoresis (CE) instrument used • Sample processing methods (e.g., extraction, quantitation, target DNA template levels tested) • Number of PCR cycles • Replicate testing • Analytical threshold • Population allele frequencies • Co-ancestry coefficient (i.e., theta value) • Analyst training and experience (with lab protocols)
PGS Model Decisions	• PGS model used (i.e., discrete or continuous) • Laboratory-specific parameters for use in the PGS model (e.g., probability of allele drop-out, probability of allele drop-in) • Non-contributor data construction and testing
Software Implementing the PGS Model	• Choice of numerical methods for computing likelihood ratios (e.g., MCMC, numerical integration) • Choice of the number of iterations or numerical integration parameters (e.g., grid size) • Choice of diagnostic checks on the results
Case Specific Decisions	• Propositions and assumptions

The set of scenarios that has been explored in a laboratory's internal validation experiments represents *factor space coverage* for that laboratory. These validation experiments are performed with known samples of varying degrees of complexity that permit exploration of the factor space and allow for assessing performance with ground truth samples. Data from such experiments can then be used to investigate case-specific reliability of their system through first identifying a collection of their tested samples which used known samples "similar" to the casework sample and then studying these results.

If the factor space coverage explored by a laboratory is only a small portion of the entire factor space, then this coverage influences what can be said about the degree of reliability for the types of samples analyzed in that laboratory. The so-called *bracketing approach*, discussed later in this chapter, is a sensible way of understanding case-specific reliability and limitations of the system. It is important to keep in mind that the entire system being considered involves both measurement and interpretation with PGS being only a component of the overall system.

A noteworthy portion of the factor space with DNA mixture interpretation involves sample complexity. As described in Chapter 2, sample complexity is influenced by the number of contributors, the degree of allele sharing and ratios of mixture components, and the amount and quality of the DNA tested. The presence of more contributors increases the number of possible genotype combinations and thus more opportunity for allele sharing. Mixtures containing DNA from closely related individuals (e.g., siblings or a parent and child) mean more opportunities for allele sharing. Stochastic variation when testing small amounts of DNA also impacts sample complexity (see Section 2.3 in Chapter 2).

With higher-order DNA mixtures, the potential factor space becomes vast (e.g., consider one aspect of the factor space with possible genotyping combinations as described in Lynch & Cotton 2018). Therefore, it is unlikely that laboratories have explored every possible region of this factor space and may not be comfortable commenting on the degree of reliability with especially complex samples. For example, a casework scenario might involve a two-person mixture with a mixture ratio of approximately 1:1 that involves a total DNA template amount of 1 ng where one of the components has been partially degraded.

Validation experiments from similar portions of the factor space can be used to assess the degree of reliability expected in this region of the factor space. If a casework scenario is encountered with an eight-person mixture involving only 10 pg total template DNA, then DNA analysts might refrain from interpreting such a sample because it has not been covered in any of their validation experiments. If only a handful of samples, similar to casework sample, have been tested during internal validation, this will typically result in a lower level of confidence in the casework result than if a large number of samples, similar to casework sample, have been tested during internal validation. The level of "coverage" is also critical; a laboratory has to have tested more than one sample of a particular type.

To assess reliability of any system, the factors that impact that system's performance need to be studied and evaluated. In attempting to address the question of reliability, we need to first understand what portions of the factor space have been explored and what were the experimental outcomes. Thus, in this scientific foundation review we assess what information and data are available, what portion of the factor space this information and data cover, and what can be learned about reliability of DNA mixture interpretation from the available information and data. It is recognized that each laboratory has to demonstrate their own degree of reliability and that we must be careful not to pool data from different sources that may come with different assumptions and caveats. However, if we know the extent to which different labs give different LR results for the same sample, then we may be able to "transfer" the experience of lab A to a different lab B, based on interlaboratory trials, provided A and B consistently produced very similar LR values on identical samples during such trials.

KEY TAKEAWAY #4.1: The degree of reliability of a component or a system can be assessed using empirical data (when available) obtained through validation studies, interlaboratory studies, and proficiency tests.

4.1.5. Provider-User Responsibilities and Examples

When information and data are shared, there are two sides to this interaction: a provider and a user. To enable effective use of any information, responsibilities exist with both providers and users. A provider of information delivers this information and accompanying data in an accessible format to be used for assessment by the user. The provider also explains the relevance and significance of the information and data. However, the user decides what to accept. Thus, a user of information assesses the degree of reliability (trustworthiness) and determines validity (e.g., whether a method is fit-for-purpose). The user, not the provider, decides whether sufficient information exists for judgment of reliability relative to the intended application.

In some settings, a forensic scientist may be the user of information and in other settings, they may be the provider of information. For example, when deciding on which method to utilize and when performing an internal validation study, the forensic scientist may be the *user* of information provided by a product developer of an instrument, commercial kit, or software program. As a user performing an internal validation study, the forensic scientist determines whether sufficient data have been collected to demonstrate that a method is fit for its intended purpose. On the other hand, when serving as an expert witness in a court setting, a forensic scientist is the *provider* of information while a trier of fact (judge or jury) and lawyers asking questions in the admissibility hearing or trial are users of the provided testimony. In this case, the judge, jury, and lawyers determine whether sufficient information has been provided.

With this scientific foundation review, the authors of this report serve as *both users and providers* in examining what data and information are publicly available (user role) and in describing our findings and their significance (provider role). Thus, there may be times when we state that there is insufficient information to externally assess the degree of reliability and others where we explain the relevance and significance of what information and data are available.

KEY TAKEAWAY #4.2: To enable effective use of any information, responsibilities exist with both providers and users of that information. While a provider explains the relevance and significance of the information and data, only the user can assess the degree of reliability, validity, and whether that information is fit-for-purpose.

4.2. Data Sources Used to Examine Reliability

Chapter 3 in this report describes data sources explored in our scientific foundation review and strategies to locate information from validation experiments, proficiency tests, and interlaboratory studies. Hundreds of articles on DNA mixture interpretation were collected from peer-reviewed journals, and many of them are cited throughout this report. As part of our assessment of the foundations of DNA mixture interpretation methods and practices, we examined factor space coverage in published articles describing STR kit developmental validation, PGS validation data, publicly available PGS internal validation summaries, DNA mixture proficiency test sets, and interlaboratory studies assessing DNA mixture interpretation.

4.3. Review of Publicly Available Data and Factor Space Coverage

Publicly available data on DNA mixture interpretation performance were examined from five sources: (1) published developmental validation studies from STR kits, (2) published PGS studies, (3) accessible PGS internal validation studies or summaries from forensic laboratories, (4) proficiency test results, and (5) interlaboratory studies.

4.3.1. Published Developmental Validation Data

Validation studies and underlying experiments assist in assessing and understanding the degree of reliability of scientific methods. As described in Appendix 1, the FBI Quality Assurance Standards (QAS) and guidelines from the Scientific Working Group on DNA Analysis Methods (SWGDAM) have historically provided requirements and guidance on studies to perform. For the forensic DNA community, levels of validation have been divided into developmental validation, often performed under the auspices of the developer, and internal validation, performed within each user laboratory or laboratory system before employing a method for casework.

Developmental validation studies are more likely to be published in the peer-reviewed literature compared to internal validation studies. The secondary internal validation studies may not be viewed as novel enough for many scientific journals as has been previously noted (Buckleton 2009).

Developmental validation studies for STR typing kits typically focus on measurement aspects important for reliable genotyping of single-source DNA samples and parameters that can inform mixture interpretation guidelines, such as heterozygote balance (peak height ratios) and stutter ratios. When publishing developmental validation results with a new STR typing kit, the goal of mixture studies is typically *to demonstrate detection of minor alleles rather than accuracy with interpreting and/or deconvoluting mixture profiles* (see Table 4.2). In these situations, conducting mixture studies may be viewed as a necessity to meet published guidelines or QAS requirements as described elsewhere (see Table A1.2 in Appendix 1).

Table 4.2. Summary of factor space coverage and findings for measurement experiments and DNA mixture studies from three developmental validation studies of commonly used commercial STR typing kits. Abbreviations: SD = standard deviation; RFU = relative fluorescence units; nt = nucleotide.

#	Reference STR Kit (PCR Cycle #) Instruments Used	Measurement Experiments and Findings	Factor Space Coverage for DNA Mixture Studies and Findings
1	Ludeman et al. (2018) **GlobalFiler** (29 cycles) ABI 3130xl, 3500, 3500xL	*Sensitivity*: Tested a single sample (007) from 3000 pg to 15.6 pg; found full profiles at ≥125 pg across 4 replicates; no significant saturation at 3 ng *Sizing precision*: Not reported *Reproducibility*: (see concordance) *Concordance*: Consistent genotypes with 1194 population samples against Identifiler and NGM SElect kits *Heterozygote balance*: Average ratios >80% (with 1 ng input DNA) *Stutter*: From 1092 population samples (table 4 in article); used mean + 3 SD	**Tested a single two-person mixture** (Raji & 007); genotypes were provided (28 of 43 alleles in 007 were non-overlapping); 1 ng total DNA used for all mixtures; 3 mixture ratios examined (1:1, 1:5, 1:8) and run in triplicate; detected all non-overlapping minor contributor alleles at the 1:5 ratio (167 pg minor) in six runs and in three of six runs at the 1:8 ratio (111 pg minor) using a 150 RFU analytical threshold

#	Reference STR Kit (PCR Cycle #) Instruments Used	Measurement Experiments and Findings	Factor Space Coverage for DNA Mixture Studies and Findings
2	Kraemer et al. (2017) **Investigator 24plex QS & Investigator 24plex GO!** (30 cycles) ABI 3500, 3130	*Sensitivity*: Tested a single sample (9948) from 1000 pg to 8 pg; found full profiles consistently at ≥125 pg; for 8 pg, 50% of expected alleles were detected; no saturation at 1 ng *Sizing precision*: Sized alleles in 96 allelic ladders (max SD ≤0.08 nt) *Reproducibility*: Consistent genotypes in a single control DNA sample across 3 sites, 8 replicates, 2 types of instruments *Concordance:* No null alleles from 656 NIST samples (99.997% with 29,520 alleles compared against 6 other STR kits) *Heterozygote balance*: decreased towards lower template amounts (see fig. 10) *Stutter*: From 656 NIST population samples (table 1 in article); used max %	**Tested a single two-person mixture** (9948 & XX107); no genotypes or degree of allele overlap described; 500 pg total DNA used for all mixtures; 9 mixture ratios examined (1:15, 1:10, 1:7, 1:3, 1:1, 3:1, 7:1, 10:1, 15:1,) and run in replicates of four; 100% of expected alleles were identified for minor components of 3:1, 7:1, and 10:1 mixtures; 97% of minor component alleles for 15:1 (31 pg minor) were identified using a 50 RFU analytical threshold
3	Ensenberger et al. (2016) **PowerPlex Fusion6C** (29 cycles) ABI 3130, 3130xl, 3500, 3500xL *Results from 8 laboratories*	*Sensitivity*: Tested in 7 laboratories (7 3500s, 2 3130s) two DNA samples serially diluted from 2 ng to 31.25 pg with each amount run in replicates of four; with ABI 3500s, 99.7% of expected alleles were detected at 125 pg, 82% alleles at 62.5 pg, and 44% alleles at 31.25 pg; saturation at 2 ng on 3130s *Sizing precision*: Sized alleles from two injections of allelic ladders (8 to 48 depending on instrument; max SD ≤0.1 nt) *Reproducibility*: Concordant genotypes across 6 laboratories with NIST SRM 2391c and 2800M control DNA *Concordance:* Two discordant calls from 652 NIST samples (99.994% concordance in 33,558 alleles compared) *Heterozygote balance*: Not reported *Stutter*: From 652 samples (table 7 in article); used average + 1 SD	**Tested a single two-person mixture** in 3 laboratories; no genotypes or degree of allele overlap described; 1 ng total DNA used for all mixtures; 9 mixture ratios examined (1:19, 1:9, 1:5, 1:2, 1:1, 2:1, 5:1, 9:1, 19:1) in replicates of four; detected all non-overlapping minor contributor alleles at the 1:2 ratio (333 pg minor), 99% at 1:5 ratio (167 pg minor), 96% at 1:9 ratio (100 pg minor), and 74% at 1:19 ratio (50 pg minor) using analytical thresholds of 175 RFU for the 3500s and 50 RFU for the 3130s

Published developmental validation studies of STR typing kits generally contain a detailed coverage of STR allele measurement aspects but a limited coverage of DNA mixture factor space. For each of the three published studies listed in Table 4.2, only a single two-person mixture combination was explored with three to nine different mixture ratios, usually with replicate testing of each mixture ratio sample. These three studies are representative of other STR kit developmental validation studies (e.g., Krenke et al. 2002, Collins et al. 2004, Ensenberger et al. 2010, Wang et al. 2012, Green et al. 2013, Ensenberger et al. 2014, Oostdik et al. 2014). With these developmental validation studies, rarely is more than a single two-person mixture examined with the mixture ratio being the primary variable explored. Overall success rate of detecting non-overlapping minor contributor STR alleles is a commonly used metric in these publications. Yet the degree of allele overlap, which depends on the genotype compositions of the mixture components, is not always described (e.g., rows 2 and 3 in Table 4.2).

4.3.2. Published PGS Validation Data

At least 60 articles involving probabilistic genotyping software have been published in the peer-reviewed literature with some form of validation data (Table 4.3). Eight articles in this table were examined and cited by PCAST in their September 2016 report (PCAST 2016). Thus, a great deal more information is now available to assess the use of PGS in DNA mixture interpretation. Data summarized in Table 4.3 help understand what factor space coverage exists for the experiments reported in these publications.

For each examined article, we considered the following information: publication year, author and title, PGS system and version number, STR typing kit used to generate the DNA profiles, study type and measured variables (e.g., developmental validation), whether results from multiple PGS systems were compared, number of samples, number of contributors, number of replicates, whether known samples were used for ground truth, source of DNA, amount of DNA, mixture ratios, sample condition (e.g., degraded DNA), degree of allele sharing in tested samples, total number of different individual samples contributing to the sample sets, non-contributor data construction and population(s) explored, and whether likelihood ratios data points were reported. Only a portion of this information is displayed in Table 4.3 as many of the publications did not contain all of the information sought for preparation of this report. What is provided here summarizes those aspects most common in the publications examined.

Table 4.3. Factor space coverage for published PGS validation data from peer-reviewed literature. Studies are grouped by PGS system and publication date. Studies listed on row #6, #7, #10, #11, #12, #13, #14, and #49 were part of the PCAST 2016 review. Nikola Osborne and Sarah Riman (NIST Associates) assisted with early versions of these summaries. NoC = number of contributors; N.E.S. = not explicitly stated in the referenced publication; N/A = not applicable; *comparison of multiple PGS systems are discussed in Table 4.4. †inclusion of ranges is not meant to imply that all combinations of DNA quantities and mixture ratios were covered. §a 31-laboratory compilation (Bright et al. 2018) contained data from eight different STR kits: GlobalFiler, Identifiler Plus, NGM SElect, PowerPlex Fusion 5C, PowerPlex Fusion 6C, PowerPlex ESI17 Pro, PowerPlex ESI17 Fast, and PowerPlex 16 HS.

#	Reference	PGS System *STR Kit*	NoC Range	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
1	Perlin & Sinelnikov 2009	TrueAllele *PowerPlex 16*	2	40	125 to 1000	1:1 to 9:1
2	Perlin et al. 2011	TrueAllele *Pro+Cofiler*	2	16 *adjudicated cases*	N.E.S.	N.E.S.
3	Perlin et al. 2013	TrueAllele *Pro+Cofiler*	2| 3	73| 14 *adjudicated cases*	N.E.S.	N.E.S.
4	Ballantyne et al. 2013 (proof of concept)	TrueAllele *Identifiler*	2	2	N.E.S.	1:1
5	Perlin et al. 2014	TrueAllele *PowerPlex 16*	2| 3| 4	40| 65| 8 *adjudicated cases*	N.E.S.	N.E.S.
6	Perlin et al. 2015	TrueAllele *Identifiler Plus*	2| 3| 4| 5	10| 10| 10| 10 (5 donors)	200, 1000	1:1 to 32:16:15:2:1
7	Greenspoon et al. 2015	TrueAllele *PowerPlex 16*	1| 2| 3| 4	11| 18| 15| 7 (11 donors)	10 to 1000	1:1 to 17:1:1:1

#	Reference	PGS System *STR Kit*	NoC Range	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
8	Bauer et al. 2020	TrueAllele *PP Fusion 5C*	2| 3| 4| 5| 6| 7| 8| 9| 10	2| 2| 2| 2| 2| 2| 2| 2| 2	500	2:1 to 25:19:14:13:12:6:5:3:1:1
9	Taylor et al. 2013	STRmix *Identifiler & NGM SElect*	Ex. 1: 2 Ex. 2: 2| 3	Ex. 1: 127 (ID) Ex. 2: 4| 6 (NGM)	100 to 500	1:1 to 5:1, 3p mixes (N.E.S.)
10	Bright et al. 2014	STRmix *Identifiler*	2| 3	1| 9	1500	1:1, 1:1:1, 10:5:1
11	Taylor 2014	STRmix *GlobalFiler*	2| 3| 4	15| 6| 10 (4 donors)	10 to 400	1:1 to 10:1; 1:1:1 to 4:3:2:1
12	Taylor & Buckleton 2015	STRmix *GlobalFiler*	4	29 profiles (Taylor 2014 data)	10 to 400	1:1:1:1 or 4:3:2:1
13	Taylor et al. 2015	STRmix *GlobalFiler & Profiler Plus*	1| 2| 3| 4	4| 1| 1| 3 (3 GlobalFiler & 6 Profiler Plus tests)	10 to 500	1:1 to 4:3:2:1
14	Bright et al. 2016	STRmix *GlobalFiler*	2| 3| 4	93 profiles (Taylor 2014 data)	10 to 400	1:1 to 10:1; 1:1:1 to 4:3:2:1
15	Taylor et al. 2016a	STRmix *GlobalFiler*	N.E.S.	205 profiles	N.E.S.	N.E.S.
16	Taylor et al. 2016b	STRmix *6 different kits*	N.E.S.	1867 profiles in 14 datasets	N.E.S.	N.E.S.
17	Taylor et al. 2017a	STRmix *multiple kits*	1| 2	N.E.S.	N.E.S.	N.E.S.
18	Taylor et al. 2017b	STRmix *GlobalFiler*	4	29 profiles (Taylor 2014 data)	10 to 400	1:1:1:1 or 4:3:2:1
19	Taylor et al. 2017c	STRmix *GlobalFiler & Profiler Plus*	1| 2| 3	1| 3| 1	50 to 1000	1:1 to 10:1; 3:2:1
20	Moretti et al. 2017	STRmix *Identifiler*	1| 2| 3| 4| 5	>1400| 105| 64| 84| 24	19 to 4000 (their table 1)	1:1 to 10:1:1:2:2 (their table 1)
21	Bright et al. 2018 *(combined data from 31 labs)*	STRmix *8 different kits*§	3| 4| 5| 6	1315| 1263| 182| 65 (combined data)	N.E.S.	N.E.S.
22	Kelly et al. 2018	STRmix *GlobalFiler*	2| 3	35| 36 (PROVEDIt data)	6 to 750	1:1, 4:1, 9:1; 1:1:1, 1:4:1, 4:4:1
23	Bille et al. 2019	STRmix *GlobalFiler*	3| 4| 5	24| 73| 50 (60 mixtures, 147 interpretations)	250 to 1000	98:1:1 to 75:20:2:2:1
24	Bright et al. 2019b	STRmix *GlobalFiler*	2| 3| 4| 5	6| 6| 6| 6 (PROVEDIt data)	126 to 750 (their table 1)	1:1 to 1:9:9:9:1 (their table 1)
25	Noël et al. 2019	STRmix *Identifiler Plus*	4	24 = 12 known + 12 casework	160 to 3260	1:1:1:1 to 10:5:2:1
26	Duke & Myers 2020	STRmix *GlobalFiler*	1| 2| 3| 4	1| 2| 4| 4 (4 donors)	250 to 1000 (degraded DNA)	1:1 to 7:1:1:1
27	Lin et al. 2020	STRmix *GlobalFiler*	3	40 profiles tested (3 related donors)	100 to 500	10 : 1-10 : 5,10
28	Schuerman et al. 2020	STRmix *GlobalFiler*	3| 4	26| 33	100 to 1000	1:1:1 to 1:1:1:1 to 20:4:4:1
29	McGovern et al. 2020	STRmix *PP Fusion 5C*	2| 3| 4	Ex. 1: 2| 3| 5 Ex. 2: 11| 10| 10	150 to 1500	1:1 to 20:1; 5:1:1:1 to 10:5:5:1

#	Reference	PGS System *STR Kit*	NoC Range	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
30	Kalafut et al. 2018	ArmedXpert *GlobalFiler*	1| 2| 3| 4	368| 64| 54 | 54 (67 donors)	100 to 1000	1:1 to 80:1; 1:1:1:1 to 20:4:4:1
31	Mitchell et al. 2012	FST *Identifiler*	1| 2| 3| 4	15| 214| 232| 31 (85 donors)	25 to 500	1:1, 4:1 1:1:1, 5:1:1
32	Balding 2013	likeLTD *Identifiler, SGM Plus*	1| 2| 3| 4	3| 5| 1| 1	N.E.S.	N.E.S.
33	Steele et al. 2014	likeLTD *SGM Plus*	1| 2| 3	3| 2| 4 (5 donors)	15 to 500	17:1 to 1:1:1
34	Steele et al. 2016	likeLTD *NGM SElect*	1| 2| 3	36| 24| 12 (36 donors)	4 to 328	1:1 to 16:1; 1:1:1 to 16:4:1
35	Puch-Solis et al. 2013	DNA Insight *SGM Plus*	1| 2	560 profiles (14 donors)	50 to 1500	1:1 to 9:1
36	Swaminathan et al. 2016	CEESIt *Identifiler Plus*	1| 2| 3	303 total	8 to 1000	1:1 to 49:1; 1:1:1 to 9:9:1
37	Gill & Haned 2013	LRmix *SGM Plus*	N.E.S.	3 examples with non-contributor performance tests	N.E.S.	N.E.S.
38	Benschop et al. 2012	LRmix *NGM*	2| 3| 4	1| 2| 1 (8 donors)	180 to 390	5:1 to 10:1:1:1
39	Benschop et al. 2015a	LRmix Studio *NGM*	1| 2| 3	64| 64| 64	3 to 36	1:1 to 1:1:1
40	Benschop et al. 2015b	LRmix *NGM*	3| 4| 5	12| 12| 12 (60 donors)	1250 to 1750	2:2:1 to 2:2:1:1:1
41	Haned et al. 2015	LRmix *NGM*	3| 4| 5	76| 74| 61	50 to 500	2:1:1 to 10:10:5:5:5
42	Haned et al. 2016	LRmix *NGM*	N.E.S.	77 mixtures; 1095 LRs	N.E.S.	N.E.S.
43	Benschop & Sijen 2014	LoCIM tool *NGM*	2| 3| 4	Training: 5| 13| 6 Testing: 70| 34| 27	60 to 1200	1:1 to 15:7:1:1
44	Benschop et al. 2019a	EuroForMix *PP Fusion 6C*	2| 3| 4| 5	30| 30| 30| 30	180 to 900	1:1 to 20:1:2:1:1
45	Bleka et al. 2019	CaseSolver *PP Fusion 6C*	2| 3| 4	9| 12| 4 (14 donors)	1000	1:1 to 13:1:1 to 4:4:1:1
46	Benschop et al. 2017b	SmartRank *NGM+SE33*	2| 3| 4| 5	155| 155| 16| 17	N.E.S.	N.E.S.
47	Benschop et al. 2019b	DNAxs *PP Fusion 6C*	1| 2| 3| 4	20| 10| 10| 10 (simulated profiles)	N/A (simulated data)	N/A (simulated data)
48	Benschop et al. 2020	DNAxs *PP Fusion 6C*	1| 2| 3| 4| 5	17| 38| 38| 37| 12 (71 donors)	180 to 5350	1:1 to 20:2:1:1:1
49	Bille et al. 2014	*multiple *Identifiler*	2	50 (2 donors)	100 to 500	1:1 to 5:1
50	Puch-Solis & Clayton 2014	*multiple *SGM Plus*	1| 2| 3| 4	10 replicates| 5| 1| 1 (Balding 2013 data)	N.E.S.	N.E.S.
51	Bright et al. 2015	*multiple *Identifiler*	2	Simulated profiles (2 donors)	N/A (simulated data)	1:1; 3:1
52	Bleka et al. 2016a	*multiple *PP ESX17*	1| 2| 3| 4	N.E.S.	N.E.S.	1:1 to 9:1; 5:4:1; 5:2:2:1
53	Bleka et al. 2016b	*multiple *NGM*	2| 3	4| 55 (33 donors)	180 to 1000	5:1 to 10:5:1

#	Reference	PGS System *STR Kit*	NoC Range	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
54	Manabe et al. 2017	*multiple *Identifiler Plus*	2| 3| 4	27| 27| 18	250 to 1000	1:1 to 7:1:1:1
55	Swaminathan et al. 2018	*multiple *Identifiler Plus*	1| 2| 3	30| 41| 30	16 to 1000	1:1 to 9:9:1
56	Alladio et al. 2018	*multiple *7 kits*	2| 3	3| 4	500 (1 diluted to 4)	1:1, 8:1, 19:1; 1:1:1 to 20:9:1
57	Buckleton et al. 2018	*multiple *Identifiler Plus*	2| 3| 4	2| 2| 1 (NIST MIX13 data)	N.E.S.	1:1 to 1:1:1:1
58	Rodriguez et al. 2019	*multiple *PowerPlex 21*	2	102	500	1:1 to 19:1
59	You & Balding 2019 (data from Steele et al. 2016)	*multiple *NGM SElect*	1| 2| 3	36| 24| 12 (36 donors)	4 to 328	1:1 to 16:1; 1:1:1 to 16:4:1
60	Riman et al. 2021	*multiple *GlobalFiler*	2| 3| 4	154| 147| 127 (PROVEDIt data)	30 to 750	1:1 to 1:9; … 1:1:1:1 to 1:9:9:1

Demonstrating the degree of reproducibility in measurements is a foundational principle of science. Replicate testing was performed in many of these publications, and reports describing interlaboratory studies are described later in Section 4.3.5. In addition, 12 studies published in peer-reviewed journals have compared results across more than one PGS system (Table 4.4).

Table 4.4. Summary of published PGS comparison studies. For details on PGS systems, see Coble & Bright 2019 and Butler & Willis 2020.

PGS Systems Compared Reference	Samples Tested	Observations Made
Lab Retriever (v.1.2.1), STRmix (assume v.2.0) Bille et al. 2014	Examined **a single Identifiler two-person mixture** with a low degree of allele sharing (10 of 15 loci displayed non-overlapping four alleles) at mixture ratios of 1:1, 1:2, 1:3, 1:4, and 1:5 with total template quantities of 100, 200, 300, 400, and 500 pg DNA amplified in duplicate (resulting in 50 mixture samples)	Plotted LR data points from the discrete (Lab Retriever) and continuous (STRmix) PGS systems along with random match probability (RMP) and combined probability of inclusion (CPI) mixture statistics (their Figure 1); reproducibility improves with higher qualities of total DNA; "information content associated with height is limited for the 1:1 mixtures but increases as we proceed toward the 2:1, 3:1, 4:1, and 5:1 mixtures"; the authors conclude: "It is noted that this trial was conducted on a relatively easy type of mixed DNA profile, two person mixtures. Further comparison with three and four person mixtures and profiles where the person of interest is potentially masked is warranted."

PGS Systems Compared **Reference**	**Samples Tested**	**Observations Made**
LiRa, likeLTD (v.4.4), LRmix (v.4.3) Puch-Solis & Clayton 2014	Examined **ten replicates of a single SGM Plus profile** (simulating a single-source, low-level DNA from Balding 2013 Table S1) where allele drop-out, drop-in, and uncertain designations are possibilities	Differences were observed with each PGS system (all are discrete models) even when only alleles (and no peak height differences) are considered; authors introduced concept of "ban evidential efficiency"; four experiments were conducted: (1) one person profiles with no replicates, (2) one person profiles consisting of two and three replicates, (3) two person profiles, and (4) three and four person profiles; more variation was observed between PGS systems as profiles became more complex; with an example involving propositions of three people, results were (in *bans*): likeLTD (*9.3*), LiRa (*8.98*), and LRmix (*3.99*) – meaning that LiRa and LRmix were five *bans* or five orders of magnitude different in this example
Lab Retriever (v.1.2.4), LRmix (v.4.3), STRmix (v.2.0) Bright et al. 2015	Used **two artificial Identifiler profiles** to create major/minor, balanced, and stochastic profiles (profiles are provided in supplementary material)	Performed four experiments: (1) comparison to the expected LR with no drop-out or drop-in, (2) the effect of drop-out, (3) the effect of drop-in, and (4) reproducibility; comparison of results identified a difference in how Lab Retriever calculated their population genetic model compared to the other two PGS systems; the authors suggest: "an essential feature of validation is the ability to specify exactly what the software is doing at least with regard to routine matters such as the population genetic model and the allele probabilities"
EuroForMix, *DNAmixtures* Bleka et al. 2016a	Examined **a two-person mixture amplified with PowerPlex ESX 17**; also simulated three random DNA profiles where one, two, three, or four individuals contributed	Compared likelihood values between EuroForMix and *DNAmixtures* by randomly generating single source profiles and two- and three-person mixtures; observed identical log likelihood values up to 11 decimal places for each considered proposition
EuroForMix, LRmix Studio Bleka et al. 2016b	Examined **four two-person and 55 three-person mixtures amplified with NGM**; see Table 1 in their article; full dataset available at http://www.euroformix.com/data	Used receiver operating characteristic (ROC) plots to examine the rate of false positives versus true positives across different conditions; the authors reported: "*LRmix* still gave a high LR for true contributors up to four dropouts for a person of interest (POI) in a three-person mixture. However, the main benefit of *EuroForMix* was with the interpretation of major/minor mixtures where the minor was evidential. Here up to 11 allele dropouts for the POI in a three-person mixture could provide probative evidence, whilst *LRmix* may return a much lower LR or a false negative result. The two models are expected to return similar LR results when contributors have equal mixture proportions or for mixtures of higher order."

PGS Systems Compared **Reference**	**Samples Tested**	**Observations Made**
Kongoh, LRmix Studio (v.2.1.3), EuroForMix (v.1.7) Manabe et al. 2017	Examined **18 mixtures** (6 two-person, 6 three-person, and 6 four-person) **amplified with Identifiler Plus**; see Tables S1 and S2 in their article	Used bar charts to compare LR values from a binary model, *LRmix Studio*, *EuroForMix*, and *Kongoh* for two-person (Figure 6), three-person (Figure 7), and four-person (Figure 8); the authors reported: "LR values of *Kongoh* tended to be similar to those of *EuroForMix* even in four-person mixtures…[except with a] minor POI of 7:1:1:1 mixtures with 0.25 ng DNA and with three drop-out alleles of the POI"
DNA•VIEW (v.37.17), EuroForMix (v.1.9.3), Lab Retriever (v.2.2.1), LRmix Studio (v.2.1.3), STRmix (v.2.3.06) Alladio et al. 2018	Examined **7 mixtures** (3 two-person and 4 3-person) plus a dilution series of a 1:1:1 mixture from 500 pg total down to 4 pg **amplified with seven STR kits** (GlobalFiler, NGM SElect, MiniFiler, PowerPlex Fusion, Fusion 6C, ESI 17 Fast, and ESX 17 Fast); mixtures were made with NIST SRM 2391c components A, B, and C	Plotted log(LR) data points from the five PGS systems by mixture ratio, NIST component, and STR kit; also plotted averaged log(LR) values from the two discrete PGS systems versus the three continuous PGS systems; created histograms to compare averaged discrete vs averaged continuous LR results for each NIST component against the overall DNA quantity in the dilution series; the authors reported: "[continuous PGS] results were always higher than the [discrete PGS] ones, regardless of the DNA amplification kit that was adopted" and "LR results provided by both [discrete PGS] models were very similar or identical" while "log(LR) results provided by [continuous PGS] models proved similar and convergent to one another, with slightly higher within-software differences (i.e., approximately 3-4 degrees of magnitude)"
EuroForMix (v.1.10.0 and v.1.11.4), Lab Retriever (v.2.2.1), LRmix Studio, STRmix (v.2.5.11) Buckleton et al. 2018	Examined **one Identifiler and four Identifiler Plus profiles** and reference samples from five NIST MIX13 mock cases; data available at https://strbase.nist.gov/interlab/MIX13.htm	Provided LR values from each PGS system compared to 1/RMP for each reference sample in case 1 (Table 4), case 2 (Table 5), case 3 (Table 7), case 4 (Table 9), and case 5 (Table 11); the authors reported on the case 1 results: "All four [PGS] tested also included reference 1A with as much as four orders of magnitude difference between software systems (see Table 4). The continuous model software systems reported the larger LRs and the [discrete] software systems essentially reported the same LR"; these general trends were observed for cases 2, 3, and 4, namely (1) that the two discrete PGS systems yielded similar results (usually less than an order of magnitude part) as did the two continuous PGS systems to one another and (2) continuous systems assigned higher LR values than discrete ones; the assigned LR results differed in case 5, which were discussed by the authors as an "over engineered" challenge involving a non-contributor reference profile possessing extensive allele overlap and that inclusion of this reference "should be termed an adventitious match not a false inclusion"

PGS Systems Compared Reference	Samples Tested	Observations Made
CEESIt (four models labeled A, B, C, D); see their Table 1 for model assumptions Swaminathan et al. 2018	Examined **101 Identifiler Plus profiles** (30 single-source, 41 two-person, 30 three-person samples) five times each; see Table S2 and Table S3 in their article	Provided summaries of minimum and maximum LR values for each model with some other statistics (Table 2); for each model 1010 LRs were produced (150 single-source, 410 two-person, and 450 three-person); the authors reported: "In all four models, intramodel variability in the LRs increased with an increase in the number of contributors and with a decrease in the contributor's template mass."
likeLTD (v.6.3.0), EuroForMix (v.1.11.4) You & Balding 2019	Examined **72 NGM SElect profiles** (36 single-source, 24 two-person, 12 three-person samples); see Table 1 in their article	Used ROC plots with different thresholds and an information gain ratio (IGR) compared to the inverse match probability (1/RMP) that serves as an upper bound; the authors reported: "Overall results from likeLTD and [EuroForMix] were similar, despite being based on different modelling assumptions."
LRmix Studio (v.2.1.3), STRmix (v.2.5.11) Rodriguez et al. 2019	Examined **102 two-person mixtures amplified with PowerPlex 21**; see Table 1 in their Supplemental file	Provided LR values for each sample and PGS system with H1 true LRs (Table 2 in Supplemental file) and H2 true test results (Table 3 in Supplemental file); also plotted log(LR) values against the number of drop-outs in the POI; the authors reported: "The capacity of the LR approach to discriminate between true and false propositions increases with the amount of correct information provided."
EuroForMix (v2.1.0) STRmix (v2.6) Riman et al. 2021	Examined **154 two-person, 147 three-person, and 127 four-person mixtures** from the PROVEDIt dataset; see Supplemental Table 4 in their article	Provided LR values for 1279 Hp-true tests (Supplemental Table 4) and 1279 Hd-true tests (Supplemental Table 5) for each software; explored LR distributions observed and used ROC plots, scatter plots, histograms with distribution of differences; evaluated apparent discrepancies between PGS models, adventitious exclusionary and inclusionary support, and verbal equivalent discordance; the authors reported: "in certain cases differences in numerical LR values from both software resulted in differences in one or more than one verbal categories (Table 8). These differences were substantially more with low template minor contributors and higher [number of contributors]…"

4.3.3. Publicly Available PGS Internal Validation Data

During our discussions on the topic of available data to assess PGS systems for DNA mixture interpretation performance, the DNA Resource Group (see Table 1.2) underscored that additional PGS data exists in forensic laboratories as part of their internal validation studies. As described in Chapter 3, internet searches were made to locate publicly available internal validation data or information (see Table 3.2 for links to the eleven publicly available

internal validation summaries that could be found when these searches were performed). Table 4.5 summarizes factor space coverage described in these validation studies.

Table 4.5. Factor space coverage of information in internal validation studies listed in Table 3.2. Initial summary completed by Sarah Riman (NIST Associate). NoC = number of contributors; N.E.S. = not explicitly stated in the referenced public source; N/A = not applicable; F = female; M = male. †inclusion of ranges is not meant to imply that all combinations of DNA quantities and mixture ratios were covered.

Laboratory PGS (version) *STR Kit* ABI CE	**NoC Range**	**# samples**	**Total DNA Quantity Range (pg)†**	**Mixture Ratios Range†**
California Department of Justice DNA Lab (Richmond, CA) STRmix (v2.0.6) *Identifiler Plus* ABI 3130 & 3500	1	N.E.S.	16, 31, 62, 125, 250, 500, 1000, 2000	N/A
	2	N.E.S.	500 | 1000	9:1, 4:1, 1:1 | 19:1, 9:1, 4:1, 2:1, 1:1
	3	N.E.S.	250, 375, 500, 750, 1000, 1500	1:1:1, 4.5:4.5:1, 6:3:1, 8:1:1
Erie County Forensic Laboratory (Buffalo, NY) STRmix (v2.3) *PowerPlex Fusion* 30 cyles ABI 3500	1	95	N.E.S.	N/A
	2	N.E.S.	500	19:1, 9:1, 3:1, 1:1
	3	N.E.S.	37, 75, 150, 300, 600 | 12, 25, 50, 100, 200, 400 | 500	3:2:1 | 1:1:1 | 5:1:1, 10:4:1, 1:5:1, 4:1:10, 1:1:5, 1:10:4
	4	N.E.S.	62, 125, 250, 500, 1000 | 500	4:3:2:1 | 17:1:1:1; 14:3:2:1; 1:1:1:1
Erie County Forensic Laboratory (Buffalo, NY) STRmix (v2.3) *Identifiler Plus* 29 cycles ABI 3500	1	94	N.E.S.	N/A
	2	N.E.S.	19, 37, 75, 150, 300 | 12, 25, 50, 100, 200, 400 | 500	2:1 | 1:1 | 1:1, 1:2, 1:3, 1:5, 1:10, 2:1, 3:1, 5:1, 10:1 (with F:M)
	3	N.E.S.	37, 75, 150, 300, 600 | 12, 25, 50, 100, 200, 400 | 500	3:2:1 (with M:F:M) | 1:1:1 (with M:F:M | 1:1:1, 3:1:1, 3:1:0.5, 3:1.5:1 (with F:M:F)
	4	N.E.S.	62, 125, 250, 500, 1000 | 12, 25, 50, 100, 200, 400 | 500	4:3:2:1 (with F:M:F:M) | 1:1:1:1 (with F:M:F:M) | 1:1:1:1, 3:1:1:1, 3:2:1:0.5 (with F:M:F:F)
Michigan State Police Forensic Science Division (Lansing, MI) STRmix (v2.3.07) *PowerPlex Fusion* 30 cycles ABI 3500	1	1	N.E.S.	N/A
	2	N.E.S.	Generally targeted 500 to 1000; the 2.5:1 mixture was examined at 1000 and 3000 pg	10:1, 7.5:1, 5:1, 2.5:1, 1:1
	3	N.E.S.	Generally targeted 500 to 1000; the 3:2:1 mixture was examined with minor donor at 117, 78, 58, 26 pg	10:1:1, 10:2:1, 10:5:1, 10:10:1, 10:10:2, 10:10:5, 10:10:10
	4	N.E.S.	Generally targeted 500 to 1000; the 4: 3:2:1 mixture was examined with minor donor at 117, 78, 58, 26 pg	10:1:1:1, 10:5:1:1, 10:10:5:1

Laboratory PGS (version) *STR Kit* ABI CE	**NoC Range**	**# samples**	**Total DNA Quantity Range (pg)†**	**Mixture Ratios Range†**
NYC OCME Forensic Biology Laboratory (New York City, NY) STRmix (v2.4) *PowerPlex Fusion* 29 cycles ABI 3130	1	3 | 30 | 5	10, 25, 50, 100, 200 | 750, 1000, 1500 | 2000	N/A
	2	N.E.S.	500	15:1, 10:1, 4:1, 2:1, 1:1
	3	N.E.S.	N.E.S.	N.E.S.
	4	N.E.S.	N.E.S.	N.E.S.
Palm Beach County Sheriff's Office (West Palm Beach, FL) STRmix (v2.4.06) *PowerPlex Fusion 5C* - 30 cycles ABI 3500xl	1	N.E.S.	30, 60, 125, 250, 500	N/A
	2	N.E.S.	100, 250, 500 | 100, 250, 500, 1000	19:1, 10:1, 5:1, 2.5:1, 1:2.5, 1:5, 1:10, 1:19 | 1:1
	3	N.E.S.	100, 250, 500, 1000 | 100, 250, 500, 1000	1:1:8, 6:3:1, 5:5:1, 1:3:3 | 1:1:1
	4	N.E.S.	100, 250, 500, 1000 | 100, 250, 500, 1000	4:4:1:1, 1:1:3:6, 1:3:3:9 | 1:1:1:1
Palm Beach County Sheriff's Office (West Palm Beach, FL) STRmix (v2.6.2) *PowerPlex Fusion 6C* - 29 cycles ABI 3500xl	1	N.E.S.	12, 25, 50, 100, 200, 400	N/A
	2	N.E.S.	100, 250, 500, 1000	20:1, 10:1, 5:1, 2:1, 1:2, 1:5, 1:10, 1:20
	3	N.E.S.	100, 250, 500, 1000	10:5:1, 8:1:1, 3;2:1, 1:1:1
	4	N.E.S.	100, 250, 500, 1000	10:5:2:1, 9:3:3:1, 6:3:1:1, 4:4:1:1, 4:3:2:1, 1:1:1:1
San Diego Police Department Crime Laboratory (San Diego, CA) STRmix (v2.3.06) *GlobalFiler* 29 cycles ABI 3500	1	N.E.S.	N.E.S.	N/A
	2	42	N.E.S.	8:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:8
	3	66	N.E.S.	33.3:33.3:33.3, 70:20:10, 60:30:10, 50:40:10, 50:30:20, 45:45:10, 40:40:20, 35:35:30, 60:20:20, 50:25:25, 40:30:30
	4	66	N.E.S.	25:25:25:25, 60:20:10:10, 50:20:20:10, 70:10:10:10, 40:20:20:20, 40:40:15:5, 35:35:20:10, 40:40:10:10, 35:35:25:5, 30:30:20:20, 30:30:30:10
	5	12	N.E.S.	20:20:20:20:20, 60:10:10:10:10
Virginia Department of Forensic Science (Richmond, VA)	1	17	10, 30	N/A
	2	18	N.E.S.	N.E.S. (mixture weight in Table 1)
	3	15	N.E.S.	N.E.S. (mixture weights in Table 2)

Laboratory PGS (version) *STR Kit* ABI CE	**NoC Range**	**# samples**	**Total DNA Quantity Range (pg)†**	**Mixture Ratios Range†**
TrueAllele Casework (v3.25.4441.1) *PowerPlex 16* ABI 3130xl	4	7	N.E.S.	N.E.S. (mixture weights in Table 3)
Department of Forensic Sciences (Washington, DC) STRmix (v2.3) *Identifiler Plus* ABI 3500	1	N.E.S.	25, 50, 100 200, 400, 2000, 4000, 8000	N/A
	2	N.E.S.	500, 1000	20:1, 15:1, 10:1, 7:1, 3:1, 1:1
	3	N.E.S.	N.E.S.	N.E.S.
	4	N.E.S.	N.E.S.	N.E.S.
Department of Forensic Sciences (Washington, DC) STRmix (v2.4) *GlobalFiler* 29 cycles ABI 3500	1	32	6, 8, 12, 15, 23, 31, 47, 63, 94, 125, 188, 250, 375, 500, 750, 1000	N/A
	2	42	600	25:1, 20:1, 15:1, 10:1, 7:1, 5:1, 3:1, 2:1, 1:1
	3	20	N.E.S.	N.E.S.
	4	20	N.E.S.	N.E.S.
	5	20	N.E.S.	N.E.S.

KEY TAKEAWAY #4.3: Currently, there is not enough publicly available data to enable an external and independent assessment of the degree of reliability of DNA mixture interpretation practices, including the use of probabilistic genotyping software (PGS) systems. To allow for external and independent assessments of reliability going forward, we encourage forensic laboratories to make their underlying PGS validation data publicly available and to regularly participate in interlaboratory studies.

4.3.4. Proficiency Tests

Proficiency test (PT) results are analyst-focused rather than method-focused like validation studies. PT results provide a means to assess participant performance and to examine trends in DNA interpretation methods. If proficiency tests are representative of commonly seen casework in a forensic laboratory, then these results can also help assess what PCAST termed "validity as applied" (PCAST 2016).

As described in Chapter 3, Collaborative Testing Services, Inc. (CTS) is currently the only proficiency test provider to publicly share their results. These results are coded to anonymize participants and yet permit a view of variation across individual submissions. In each of the CTS PTs, four samples are provided (either as samples or profiles): Item 1 and Item 2 serve

as references for comparison to "evidence" Item 3 and Item 4. CTS also provides a mock case scenario for context. Participants conduct their analyses and interpretations according to their laboratory protocols and report their results.

For each item, participants return results for (1) body fluid screening (e.g., "positive," "negative," "inconclusive," or "not tested" for the presence of blood along with listing test(s) conducted), (2) allele calls for autosomal STR loci analyzed with one or more STR kits (and Y-chromosome STR loci and mitochondrial DNA sequencing, if performed), (3) interpretation, and (4) additional comments that may assist in review of their results. A differential extraction (see Box A1.1 in Appendix 1) can be performed to separate DNA components into sperm and epithelial fractions. In the past few years, participants have been asked to report whether a PGS system was used to assist in their DNA mixture interpretation.

Interpretation typically involves answering a question like: "Based on results obtained from DNA analysis, could the Victim (Item 1) and/or the Suspect (Item 2) be a contributor to the questioned samples (Item 3 and Item 4)?" Thus, the assessment is simply "Yes" or "No" (i.e., inclusion or exclusion) and does not include a statistical evaluation of the strength of evidence. Some participants may respond with "inconclusive" or "no interpretation" as well. The summary report from CTS provides manufacturer information about how the samples were created along with the "correct" result, which is determined by consensus of participants. A minimum of 10 participants is required for a result (e.g., genotype at a STR locus) to be graded. This consensus approach impacts some of the results with DNA Interpretation PTs, which typically do not have as many participants (e.g., compare Table 4.6 to Table 4.7).

Table 4.6. Analysis of 69 available data sets from Collaborative Testing Services (CTS) Forensic Biology, DNA Mixture, and DNA Semen proficiency tests between 2013 and 2020. Note that numbers on probabilistic genotyping software (PGS) use were not formally collected and reported by CTS until recently (DNA Semen 17-5802 is first direct count of PGS in the CTS report summary). Numbers in the PGS column depend on participant reporting or a manual review of summary reports and percentages are based on the number using PGS divided by the number reporting DNA interpretations rather than the total number of participants. Mock evidence samples provided by CTS (Item 3 or Item 4) include single-source blood (B) samples and blood/blood (B/B) or blood/semen (B/S) mixtures. False exclusion or false negative (FN) results involve reporting an exclusion of DNA results from a provided reference sample that was present in the evidence sample. False inclusion or false positive (FP) results involve reporting an inclusion of DNA results from a reference sample that was not present in the evidence sample.

CTS Test Number	Number of Participants	# Using PGS (% participants)	Samples Provided (sample ratio noted)		Results
			Item 3	*Item 4*	
13-581 (DNA Mixture)	128	0	B/S (1:1)	B	2 FN
13-586 (DNA Mixture)	107	0	B	B/S (1:1)	--
14-571 (Forensic Biology)	778	0	B	B	5 FN, 3 FP
14-572 (Forensic Biology)	603	0	B	B/S (1:1)	1 FN
14-573 (Forensic Biology)	357	0	B/B (1:1)	B	3 FP
14-574 (Forensic Biology)	756	0	B/S (1:1)	B	1 FN
14-575 (Forensic Biology)	611	0	B/S (1:1)	B	1 FN
14-576 (Forensic Biology)	334	0	B	B/S (1:1)	3 FN
14-582 (DNA Semen)	149	0	B	B/S (1:1)	--
14-584 (DNA Semen)	169	0	B	B/S (1:1)	5 FN

CTS Test Number	Number of Participants	# Using PGS (% participants)	Samples Provided (sample ratio noted) *Item 3*	*Item 4*	Results
14-581 (DNA Mixture)	130	0	B	B/S (1:1)	--
14-586 (DNA Mixture)	142	0	B/S (1:1)	B	4 FN
15-571 (Forensic Biology)	727	0	B/S (1:1)	B	--
15-572 (Forensic Biology)	631	0	B/B (1:1)	B	--
15-573 (Forensic Biology)	351	0	B	B/S (1:1)	1 FN, 1 FP
15-574 (Forensic Biology)	675	0	B	B	--
15-575 (Forensic Biology)	611	0	B	B/S (1:1)	1 FN
15-576 (Forensic Biology)	320	0	B	B	--
15-582 (DNA Semen)	179	0	B/S (1:1)	B	1 FN
15-584 (DNA Semen)	160	0	B	B/S (1:1)	--
15-581 (DNA Mixture)	145	0	B/S (1:1)	B	3 FN
15-586 (DNA Mixture)	121	0	B/S (1:1)	B	--
16-571 (Forensic Biology)	697	0	B	B	1 FN, 1 FP
16-572 (Forensic Biology)	659	24 (4%)	B/S (1:1)	B	3 FN, 1 FP
16-573 (Forensic Biology)	360	0	B	B	--
16-574 (Forensic Biology)	615	1 (0.2%)	B	B/S	1 FN
16-575 (Forensic Biology)	632	27 (4%)	B/B (1:1)	B	--
16-576 (Forensic Biology)	329	1 (0.3%)	B/S (1:1)	B	1 FP
16-582 (DNA Semen)	174	19 (11%)	B	B/S (1:1)	1 FN
16-584 (DNA Semen)	188	13 (7%)	B	B/S (1:1)	3 FN
16-581 (DNA Mixture)	142	0	B	B/S (1:1)	2 FN
16-586 (DNA Mixture)	144	0	B/B (1:1)	B/S (1:1)	3 FN
17-5701 (Forensic Biology)	672	1 (0.1%)	B	B	--
17-5702 (Forensic Biology)	660	29 (4%)	B	B	--
17-5703 (Forensic Biology)	348	2 (0.6%)	B	B/S (1:1)	3 FN
17-5704 (Forensic Biology)	671	13 (2%)	B/S (1:1)	B	--
17-5705 (Forensic Biology)	594	30 (5%)	B/S (1:1)	B	1 FN, 2 FP
17-5706 (Forensic Biology)	327	9 (3%)	B/B (1:1)	B/B (1:1)	1 FN, 1 FP
17-5802 (DNA Semen)	187	21 (11%)	B	B/S (1:1)	--
17-5804 (DNA Semen)	194	1 (0.5%)	B/S (1:1)	B	1 FN
17-5801 (DNA Mixture)	179	0	B/S (1:1)	B/S (1:1)	1 FN
17-5806 (DNA Mixture)	167	1 (0.6%)	B/S (1:1)	B/B (1:1)	--
18-5701 (Forensic Biology)	683	138 (20%)	B/B (1:1)	B	1 FN, 1 FP
18-5702 (Forensic Biology)	651	168 (26%)	B	B/S (1:1)	1 FN
18-5703 (Forensic Biology)	359	76 (21%)	B	B/S (1:1)	--
18-5704 (Forensic Biology)	672	149 (22%)	B/S (1:1)	B	1 FN
18-5705 (Forensic Biology)	624	193 (31%)	B	B	--
18-5706 (Forensic Biology)	343	97 (28%)	B/B (1:1)	B	--
18-5802 (DNA Semen)	226	46 (20%)	B	B/S (1:1)	--
18-5804 (DNA Semen)	181	22 (12%)	B/S (1:1)	B	1 FN
18-5801 (DNA Mixture)	156	4 (3%)	B	B/S (1:1)	3 FN, 1 FP
18-5806 (DNA Mixture)	178	33 (19%)	B/S (1:1)	B/B (1:1)	--
19-5701 (Forensic Biology)	732	127 (17%)	B	B/S (1:1)	--
19-5702 (Forensic Biology)	739	(35%)*	B	B/B (1:1)	--
19-5703 (Forensic Biology)	366	(30%)*	B	B	--
19-5704 (Forensic Biology)	696	183 (26%)	B	B	1 FN, 1 FP
19-5705 (Forensic Biology)	705	281 (40%)	B/S (1:1)	B	13 FN, 1 FP
19-5706 (Forensic Biology)	333	137 (41%)	B/B (1:1)	B/S (1:1)	--
19-5802 (DNA Semen)	223	46 (21%)	B	B/S (1:1)	--
19-5804 (DNA Semen)	166	22 (13%)	B/S (1:1)	B	3 FN

CTS Test Number	Number of Participants	# Using PGS (% participants)	Samples Provided (sample ratio noted) *Item 3*	*Item 4*	Results
19-5801 (DNA Mixture)	169	38 (22%)	B/S (1:1)	B/B (1:1)	--
19-5806 (DNA Mixture)	171	46 (27%)	B	B/S (1:1)	--
20-5801 (DNA Mixture)	235	42 (18%)	B/B (1:1)	B/S (1:1)	--
20-5701 (Forensic Biology)	671	182 (29%)	B	B	--
20-5702 (Forensic Biology)	734	307 (49%)	B/S (1:1)	B	6 FN
20-5703 (Forensic Biology)	345	156 (49%)	B/B (1:1)	B	--
20-5704 (Forensic Biology)	728	236 (34%)	B	B/S (1:1)	1 FN, 1 FP
20-5802 (DNA Semen)	207	35 (20%)	B/S (1:1)	B	--
20-5804 (DNA Semen)	186	40 (22%)	B/S (1:1)	B	--
TOTAL	**27,602**				**80 FN, 18 FP**

*only percentages of PGS users were provided by CTS

These CTS DNA mixture PTs involve single-source or two-person mixtures created from large quantities of DNA (hundreds to thousands of cells). In other words, the mixtures in the Forensic Biology, DNA Semen, and DNA Mixture PT exams (Table 4.6) are not complex. From the 138 test samples evaluated across these 69 PTs, evidence items (i.e., test samples "Item 3" or "Item 4") were either single-source samples (72 of 138; 52%) or two-person mixtures created with blood and semen (51 of 138; 37%) or blood and blood (15 of 138; 11%) combined in approximately one-to-one (1:1) ratios.

Across these 69 data sets, there were 80 false negatives and 18 false positives reported from 110,408 possible responses[18] (27,602 participants × two evidence items × two reference items). In the past five years, the number of participants using PGS has grown.

Table 4.7. Summary of 14 CTS DNA Interpretation proficiency tests between 2013 and 2020. Evidence profiles are designed from single individuals (single), two-contributor mixtures (2p), or three-contributor mixtures (3p) with the contributor ratios indicated in parentheses. Blue font indicates inclusion of a contributor in the evidence profile that is not a supplied reference profile ("Item 1" or "Item 2"). If four values occur in a column (e.g., # false inclusions in the 15-588 row), then each number represents a summation of participant responses with the comparison (in order of evidence-profile-to-reference-profile) for Item 3 to Item 1, Item 3 to Item 2, Item 4 to Item 1, and Item 4 to Item 2. Results obtained with three-contributor mixtures are highlighted in **bold font**.

Year	CTS Test	Number of Participants	Item 3	Item 4	# False Inclusions	# False Exclusions	# Inconclusives	# No Response
2013	13-589	13	single	2p (4:1)	0	0	0	0
2014	14-588	20	2p (2:1)	single	0	0	0	0
2014	14-589	19	single	2p (2:1)	0	0	0	0
2015	15-588	19	single	2p (3:1)	0,1,0,0	0	0	0
2015	15-589	24	2p (1:4)	single	0	0	0	0
2016	16-588	20	2p (3:1)	2p (1:1)	0	0	1,3,0,3	0

[18] There were also inconclusive responses and no responses that are not reflected in this data analysis. The ability to determine an exact denominator of a test is sometimes limited by how the data are tabulated and summarized by CTS.

Year	CTS Test	Number of Participants	Item 3	Item 4	# False Inclusions	# False Exclusions	# Inconclusives	# No Response
2016	16-589	28	**3p** (2:1:2)	2p (4:1)	0	**1**,0,0,0	**2,4**,0,0	1,0,1,0
2017	17-588	21	**3p** (1:2:1)	2p (1:3)	0	0	**4,2**,1,0	3,0,3,0
2017	17-589	19	2p (1:4)	**3p** (5:1:3)	0	0,0,0,**1**	0,0,**2,4**	0
2018	18-588	25	2p (1:1)	2p (3:1)	0	0	0,0,3,0	0
2018	18-589	36	2p (3:1)	**3p** (6:3:1)	0	0	0,0,**12,12**	0
2019	19-588	28	**3p** (4:1:2)	2p (1:4)	0	0	**1,9**,0,0	0
2019	19-589	38	2p (2:3)	**3p** (5:2:2)	0	0	0,0,**7,9**	0
2020	20-5881	43	**3p** (5:1:3)	2p (4:1)	0	0	**7,9**,0,0	0
	TOTAL	**353**			**1**	**2**	**15,27,25,28 95**	**8**

The DNA Interpretation PTs (Table 4.7), which have been available since 2013 and provide EPGs rather than biological samples, yield a slightly expanded factor space with five (18%) DNA profiles coming from a single-source sample, 16 (57%) containing mixtures with two contributors ("2p"), and seven (25%) involving three contributors ("3p") out of 28 evidence items in the data set examined.

The 14 CTS DNA Interpretation PTs gathered 1412 responses (353 participants × two evidence items × two reference items). These responses include one false inclusion (0.07%), two false exclusions (0.14%), 95 inconclusive results (6.7%), and eight no responses (0.57%). Curiously, the single false inclusion came from a reference Item 2 to a single contributor evidence profile (Item 3, which was not a provided reference profile and was incorrectly classified as a two-contributor mixture by the submitter).

4.3.5. Interlaboratory Studies

Interlaboratory comparison studies, which are sometimes referred to as collaborative exercises or round-robin studies, provide a community-focused approach to demonstrate that multiple laboratories can generate comparable measurements and interpretation when provided with the same samples or DNA profiles.

There have been at least 18 interlaboratory studies involving DNA mixture interpretation (see Table 1 in Butler et al. 2018a as well as Bright et al. 2019a). These studies have been organized by the National Institute of Standards and Technology (NIST), the Defense Forensic Science Center (DFSC), the Spanish-Portuguese Working Group of the International Society for Forensic Genetics (GHEP-ISFG), the European Forensic Genetics Network of Excellence (EuroForGen-NoE), the Netherlands Forensic Institute (NFI), developers of the PGS system STRmix, the UK Forensic Science Regulator, and the UK Association of Forensic Science Providers (AFSP). Some studies provided samples to explore both measurement and interpretation aspects of the process and other studies

provided only DNA profile EPGs to examine interpretation variability across participants (Table 4.8). A few of the studies have explored performance across forensic DNA laboratories with low-level, high-contributor mixtures.

Table 4.8. Summary of factor space coverage with 18 interlaboratory studies involving DNA mixture interpretation. Abbreviations: 2p = two-person mixture; 3p = three-person mixture; 4p = four-person mixture; 5p = five-person mixture; AT = analytical threshold; N/A = not applicable; N.E.S. = not explicitly stated; NOC = number of contributors; pg = picograms; ss = single-source; S&S = Schleicher & Schuell; Unk. = unknown; Year = year study was conducted.

Year	Reference & Study Name	PGS System (Version)	Format of Sample/Data Provided	# Laboratories (Data Sets)	# Sam -ples	# with NOC	Total DNA Amount (pg)	Mixture Ratio Range
1997	Duewer et al. (2001) NIST Mixed Stain Study #1	N/A	Buffy coat cells on S&S 903 paper	22 (37)	11	6-ss 4-2p 1-3p	30,000 to 50,000 30,000 to 50,000 30,000 to 50,000	N/A ≈1:1 ≈1:1:1
1999	Kline et al. (1999); Duewer et al. (2001) NIST Mixed Stain Study #2	N/A	Blood & semen stains on cotton cloth; DNA extracts	45 (70)	11	4-ss 6-2p 1-3p	≈1 µg per source, or ≈1 to 3 million pg for each stain; 500 to 5,000 pg/µL for DNA extracts	3:1 2:1:1
2001	Kline et al. (2003); Duewer et al. (2004) NIST Mixed Stain Study #3	N/A	DNA extracts	74 (117)	6	1-ss 5-2p 1-3p	1,000 to 4,000 pg/µL	3:1 to 10:1 4:2:1
2005	Butler et al. (2018a) NIST MIX05	N/A	EPG data (.fsa files) from 6 STR kits	69 (75)	4	4-2p	N.E.S. (≈1,000 to 1,500)	1:1 to 7:1
2010	Crespillo et al. (2014) GHEP-MIX01	N/A	EPG data (.fsa files) from 2 STR kits	32 (32)	4	4-2p	N.E.S.	1:1 to 10:1
2011	Crespillo et al. (2014) GHEP-MIX02	N/A	EPG data (.fsa files) from 1 STR kit	24 (24)	2	1-2p 1-3p	N.E.S.	5:1 2:1:1
2012	Crespillo et al. (2014) GHEP-MIX03	N/A	EPG data (.fsa files) from 2 STR kits	17 (17)	3	2-2p 1-3p	N.E.S.	5:1 to 10:1 7:3:1
2013	Prieto et al. (2014) EuroForGen Mixture Study	LRmix by all labs	EPG data (csv format) with case scenarios; population allele frequencies	18 (20); 18 (22)	2	2-2p	N.E.S.	N.E.S.
2013	Butler et al. (2018a) NIST MIX13	LabRetriever or TrueAllele used by 3 labs	EPG data (.fsa files) from 2 STR kits with case scenarios	108 (163)	5	2-2p 2-3p 1-4p	N.E.S. (≈300 to 2,000)	1:1 to 3:1 6:1:1; 7:2:1 1:1:1:1

Year	Reference & Study Name	PGS System (Version)	Format of Sample/Data Provided	# Laboratories (Data Sets)	# Sam -ples	# with NOC	Total DNA Amount (pg)	Mixture Ratio Range
2014	Barber et al. (2015) UK Forensic Regulator	LRmix, likeLTD used by 2 labs	4 DNA mixtures and 1 EPG (.fsa format) with case scenarios	8 (18)	5	2-2p 3-3p	N.E.S.	2:1 to 4:1 6:4:1 to 7:1.5:1
2014-2015	Aranda et al. (2015) *talk* DFSC Mixture Study	N.E.S.	N.E.S.	55 (185)	6	4-2p 2-3p	N.E.S.	2:1 to 3.5:1 1:1:1; 4:1:1
2014	Cooper et al. (2015) STRmix	STRmix (2.0?) by all labs	Identifiler profiles from 3 casework samples (ground truth not known)	12 (20)	3	Unk.	N.E.S.	Unk.
2014	Toscanini et al. (2016) GHEP-ISFG Basic	N/A	Stain from 2:1 volume ratio mixture of saliva and blood	72	1	1-2p	N.E.S.	≈2:1
2014	Toscanini et al. (2016) GHEP-ISFG Advanced	N/A	Stain from 4:1 volume ratio mixture of saliva and semen	52	1	1-2p	N.E.S.	≈4:1
2015	Barrio et al. (2018) GHEP-ISFG MIX06	LRmix Studio used by 15 labs	EPG data (PDF) for NGM kit loci pre-analyzed with AT = 50 RFU	25	1[§]	1-3p	N.E.S.	7:3:1
2016	Benschop et al. (2017a) NFI-organized inter- and intra-laboratory exercise	LRmix Studio (v2.0.1) used by 1 lab on some samples	EPG data (PDF) with 4 replicates for NGM kit loci pre-analyzed with AT = 50 RFU; provided in Set A or Set B	3 (26)	5 in each of 2 sets	2-2p 4-3p 2-4p 2-5p	180 24 27 186 360 240 1750	5:1 1:1 1:1:1 25:5:1 10:1:1 5:1:1:1 2:2:1:1:1
2018	Thomson (2018) *talk* UK AFSP	5 STRmix, 1 LiRa, 1 LRmix/ EuroForMix	Re-used DNA mixtures from Barber et al. (2015)	7 (28)	4	2-2p 3-3p	N.E.S.	2:1 to 4:1 6:4:1 to 7:1.5:1
2018	Bright et al. (2019a) STRmix collaborative exercise	STRmix (v2.4 and v2.5)	2 PROVEDIt EPG profiles (.hid files) or text files with STR allele, peak height, and size information; *key known variables were fixed*	42 (174)	2	1-3p 1-4p	750 105	4:4:1 4:1:1:1

[§] in the Barrio et al. 2018 study, a second sample with two males mixed 3:1 was also provided with Y-chromosome data

4.4. Discussion

Demonstrating reliability requires that the provider provide empirical data that is accessible to users of the information for independent assessments of reliability. Agreed-upon criteria from the user are also needed to establish an acceptable degree of reliability. The factor space for DNA mixture interpretation is vast and increases significantly with more contributors (Lynch & Cotton 2018). It is therefore practically impossible to demonstrate reliability across the full extent of any factor space. The focus here is on what empirical data are available so that each party can make their own judgment. Section 4.3 describes ranges (but not specific combinations) of factor space coverage for published validation studies (Tables 4.2, 4.3, and 4.4), internal validation summaries of several PGS systems (Table 4.5), proficiency test data (Tables 4.6 and 4.7), and interlaboratory studies (Table 4.8).

Based on an examination of publicly available information reviewed during the time frame of this study, there is not enough information for the authors of this report to independently assess the degree of reliability of DNA mixture interpretation at any one point in the factor space. This is particularly true without an established and accepted criteria for reliability with complex mixtures involving contributors containing low quantities of DNA template (e.g., Benschop et al. 2015a) or where there is a high degree of allele overlap among contributors (e.g., Bright et al. 2018, Lin et al. 2020).

A bracketing approach (discussed in Section 4.4.5) may provide a pragmatic way to infer reliability for DNA mixtures in a region of the factor space, *but will still require an element of trust in the DNA interpretation system used* since the entire factor space may not be covered with previously collected validation data. Yet even with a bracketing approach where there is not validation data defining every portion of the factor space, a user must trust in the DNA interpretation system enough to extrapolate assessment of reliability across gaps in the factor space covered.

Results from PGS systems do appear to demonstrate *trends* that LR values decrease with less information; either with lower quantities of DNA template (e.g., Perlin & Sinelnikov 2009, Bright et al. 2016) or with greater allele sharing (e.g., "the greater the allele sharing, the less the power there is to discriminate a true contributor from a non-contributor" as noted by Bright et al. 2018). However, such "sanity checks" with observed trends in LR values do not demonstrate the reliability of a specific LR number.

Many of the published PGS studies or available internal validation summaries include graphs of log(LR) values plotted against total input DNA or the average peak height (APH) per known contributor as described in various publications (e.g., Taylor 2014, Moretti et al. 2017). However, to independently assess the degree of reliability of PGS models, metadata associated with specific sample results and the corresponding specific log(LR) value datapoints are needed. Data of this nature are not generally shared in publications or validation summaries. A notable exception includes LR data points for 102 two-person mixtures included in a supplemental file to a published journal article (Rodriguez et al. 2019).

Likelihood ratio results from PGS systems may be reliable, or consistently accurate, in some portions of the DNA mixture interpretation factor space. However, LR results cannot be *externally and independently demonstrated to be reliable* without access to underlying performance data. To establish and support clear reliability boundaries (i.e., a certain number of contributors, a particular quantity of DNA, a specific degree of allele sharing among contributors), data need to be available to users of the information (e.g., DNA analyst or stakeholders using their results) and acceptable levels of reliability must be decided upon by the user.

4.4.1. PCAST Sources and Statements on DNA Mixture Interpretation

Of the 2100 references[19] compiled in conjunction with the September 2016 PCAST Report, there were 294 publications listed in the DNA section. In the PCAST discussion of complex mixtures (PCAST 2016, pp. 75-83), the authors cited eight articles on PGS (Bille et al. 2014, Bright et al. 2014, Taylor 2014, Greenspoon et al. 2015, Perlin et al. 2015, Taylor et al. 2015, Taylor & Buckleton 2015, Bright et al. 2016). After examining these PGS references, the PCAST authors share their judgments (but not their specific criteria for reliability):

> "…*current studies have adequately explored only a limited range of mixture types* (with respect to number of contributors, ratio of minor contributors, and total amount of DNA). The two most widely used methods (STRmix and TrueAllele) appear to be reliable within a certain range, based on the available evidence and the inherent difficulty of the problem. Specifically, these methods appear to be reliable for three-person mixtures in which the minor[20] constitutes at least 20 percent of the intact DNA in the mixture and in which the DNA amount exceeds the minimum level required for the method.[21] *For more complex mixtures (e.g., more contributors or lower proportions), there is relatively little published evidence*… When further studies are published, it will likely be possible to extend the range in which scientific validity has been established to include more challenging samples" (PCAST 2016, pp. 80-81, emphasis added).

Since specific judgment criteria used by PCAST are not stated in their report, it is unclear on what basis PCAST claims that PGS "methods appear to be reliable." We, the authors of this NIST report, emphasize that publicly available data from validation studies, whether or not this information has been published in a peer-reviewed journal, enable a user (e.g., the DNA analyst when the provider is the PGS developer or the court when the analyst is providing their results) to scrutinize the underlying data and supporting details for what is currently possible in research settings (what PCAST terms "scientific or foundational validity") and what is actually happening in casework settings (what PCAST calls "validity as applied").

A follow-on Addendum to the PCAST Report published four months later states:

> "PCAST found that empirical testing of [PGS] had largely been limited to a narrow range of parameters (number and ratio of contributors)… The path forward is straightforward. The validity of specific [probabilistic genotyping] software should be validated by testing a diverse collection of samples within well-defined ranges." (PCAST 2017, pp. 8-9).

[19] https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_references.pdf
[20] Changed to "person of interest" in a January 2017 Addendum to the PCAST Report (see p.8 of https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf)
[21] We note that this PCAST statement does not say anything about the quantity of DNA from the minor contributor(s).

In a footnote to their statement "there has been little empirical validation," the PCAST Addendum concludes:

> "The few studies that have explored 4- or 5-person mixtures often involve mixtures that are derived from only a few sets of people (in some cases, only one). Because *the nature of overlap among alleles is a key issue*, it is critical to examine mixtures from various different sets of people. In addition, the studies involve few mixtures in which a sample is present at an extremely low ratio. *By expanding these empirical studies, it should be possible to test validity and reliability across a broader range*" (PCAST 2017, footnote #11, emphasis added).

Thus, the PCAST Report (PCAST 2016) and its Addendum (PCAST 2017) emphasize a need to have casework with factor space coverage represented in the empirical studies that are performed and shared for independent review as the field adopts PGS methods to assist with DNA mixture interpretation. PCAST specifically mentions the benefits of testing "mixtures from various different sets of people" to explore PGS performance in terms of the degree of allele overlap from contributors and "mixtures in which a sample is present at an extremely low ratio."

As noted above, when their analysis was performed in 2016, PCAST provided their opinion that "current studies *have adequately explored only a limited range of mixture types* (with respect to number of contributors, ratio of minor contributors, and total amount of DNA)" (PCAST 2016, emphasis added).

Now, with the perspective of an additional five years of reflection, what publicly available data exist? Locating and understanding this information have been an important part of this DNA mixture interpretation foundation review.

4.4.2. Comments on Validation Experiments

Validation studies performed in a research or practitioner laboratory provide information to stakeholders to make assessments regarding the degree of reliability for a particular method. Validation studies are designed to generate sufficient data such that the laboratory decision maker (e.g., DNA Technical Leader) can evaluate and decide whether a method is reliable for their application. Guidance documents on validation in forensic science typically focus on types of tests to perform in gathering the data rather than ways to assess the data or the number of samples needed to demonstrate a particular level of performance.

As described earlier in Section 4.1.6, a determination of whether the amount and type of data available is satisfactory or sufficient to the user of the information is something that must be decided by the user of the information (e.g., the DNA analyst), not the provider (e.g., the software developer). It is not helpful for the provider to describe a method as "validated" without providing context around the method's use and access to data to support claims of validity and reliability. Instead, it might be more appropriate to state "the following developmental validation studies have been conducted and here is the complete collection of results obtained, which can be examined by users to make reliability judgments." Internal

validation studies provide an opportunity for the user (e.g., DNA analyst) to understand performance of a method in their forensic laboratory environment rather than trusting the provider's (e.g., the software developer) claim that everything works fine.

An important focus of STR typing kit developmental validation studies involves *measurement capabilities* to demonstrate consistent and accurate allele calling of single-source samples using sizing precision studies, concordance to previous results, and reproducibility among multiple test sites. Results from these types of studies have demonstrated a strong foundation in sizing precision and STR allele designation using allelic ladders and internal size standards with capillary electrophoresis measurements (e.g., Larazuk et al. 1998, Butler et al. 2004). *Demonstrating a method's measurement capabilities is very different from showing reliability of interpretation.*

A common metric for assessing mixture measurement capabilities during STR typing kit developmental validation studies is the ability to detect non-overlapping alleles in minor contributors. For example, one study states: "Alleles unique to the minor contributor were counted and presented as a percentage of the total number of unique alleles expected (percent unique alleles called)" (Oostdik et al. 2014). Earlier developmental validation studies, such as with PowerPlex 16 (Krenke et al. 2002), found that differences in capillary electrophoresis instrument sensitivity and variation in analytical thresholds could have an important impact on the ability to detect minor contributor alleles. After comparing results from 15 contributing laboratories, all laboratories could only identify every minor allele in the prepared mixtures between mixture ratios of 2:1 and 1:2. They could detection ~50% minor alleles at a 9:1 ratio and ~17% at a 19:1 ratio (Krenke et al. 2002). Instrument and assay sensitivity have improved in the past two decades so it is expected that lower-level minor contributors are detectable now across multiple laboratories. This aspect has not been specifically explored in published STR typing kit developmental validation studies or DNA mixture interpretation interlaboratory studies.

4.4.3. Available PGS Validation Studies

A number of articles on PGS (e.g., see Coble & Bright 2019 for a review) and other aspects of DNA mixture interpretation have been published in peer-reviewed journals since the release of the PCAST Report in September 2016. This includes a multi-laboratory response by the developer and users of one of the PGS systems (Bright et al. 2018). In addition, publicly available internal validation summaries were located online as part of this review (see Tables 3.2 and 4.5).

In total, 60 published articles on PGS and associated validation studies from the peer-reviewed literature (Table 4.3) and 11 publicly available internal validation summaries (Table 4.5) were inspected to find the factor space coverage of samples examined with various PGS systems in the published or publicly available studies[22]. Factor space coverage incorporates the number of contributors, total DNA quantity, and mixture ratio ranges.

[22] This information in Table 4.3 and Table 4.5 comes from 31 studies using STRmix, nine studies using TrueAllele, six studies using LRmix or LRmix Studio, three studies using likeLTD, two studies from DNAxs, and one study each from FST, EuroForMix, CEESIt, ArmedXpert, DNA Insight, LoCIM tool, CaseSolver, and SmartRank. In addition, there are 12 studies comparing multiple PGS systems that are also discussed in Table 4.4. A variety of STR typing kits were also used in combination with these various PGS systems.

However, the complete information is not always readily accessible or is not explicitly stated (N.E.S.) in the referenced public source. For example, many internal validation studies described in Table 4.5 do not clearly state the number of samples tested, making it difficult to assess the extent of the studies. The lack of availability of underlying data prevents independent assessments of reliability.

4.4.3.1. Degree of Allele Sharing

An important missing element from many validation studies is the degree of allele sharing that has been tested. Specific STR profiles for mixture contributors are rarely shared in publications. A 2019 article explicitly states: "Profiles used in the validation are covered by privacy rules and cannot be published" (Bleka et al. 2019). Likewise, sample genotypes are typically unavailable in forensic DNA laboratory validation summaries, perhaps due to similar privacy concerns around releasing genotype information of individuals used in these studies.

While privacy concerns may prevent researchers and laboratories from explicitly sharing mixture contributor genotypes, it is useful to convey the assessed degree of allele sharing in experiments performed. Most of the articles listed in Table 4.3 do not address the degree of allele sharing in the tested mixture samples. One exception is a study performed by the Netherlands Forensic Institute where tested samples were designated as possessing high, low, and random allele sharing without revealing the specific genotypes (Benschop et al. 2019a). Another article mentions allele sharing, pointing out a neutral approach to sample selection: "No attempt was made to maximize or minimize the amount of allele sharing between donors" (Schuerman et al. 2020).

If validation studies are conducted using mixtures that do not explore the complexity induced by allele sharing, the user may inadvertently extrapolate validation results and apply methods beyond the limits of the validation studies conducted.

4.4.3.2. Publicly Available PGS Internal Validation Summaries

Within the 11 publicly available internal validation studies summarized in Table 4.5, ten studies involve various versions of STRmix and different STR typing kits and one study assesses TrueAllele and PowerPlex 16. All of these validation summaries report exploring single-source samples as well as two-person and three-person mixtures with contributor ratios ranging up to 25 times the quantity of the smallest contributor for two-person mixtures and up to 10 times the quantity of the smallest contributor for three-person mixtures. Ten of these 11 studies examined four-person mixtures involving contributor ratios spanning 17:1:1:1 to 10:10:5:1 to 4:3:2:1 to 1:1:1:1. Many studies were conducted with total DNA quantities in the range of 500 pg to 1000 pg although minor contributor quantities were sometimes in the range of single-cell analysis (6 pg) where significant allele drop-out would be expected.

Two of the 11 studies in Table 4.5 describe the examination of five-person mixtures, including 12 samples reported by the San Diego Police Department Crime Laboratory and 20 samples

reported by the Washington DC Department of Forensic Sciences. Information on DNA quantities examined, mixture ratios studied, and degree of allele sharing in these five-person mixture samples was not explicitly stated in the referenced public sources. Additional data exploring five-person mixtures (and other mixtures examined) may exist within these 11 laboratories; however, as previously described, this report considers only publicly available data.

Although more validation studies (see Tables 4.3 and 4.5) have been performed since the 2016 PCAST Report was released almost five years ago, in their present form, publicly available internal validation *summaries* often do not provide sufficient information to assess factor space coverage. Further, these summaries typically do not provide data points (e.g., LR values) and associated information (see Box 4.1) necessary to assess the degree of reliability and performance under potential case scenarios.

KEY TAKEAWAY #4.4: Additional PGS validation studies have been published since the 2016 PCAST Report. However, publicly available information continues to lack sufficient details needed to independently assess reliability of specific LR values produced in PGS systems for complex DNA mixture interpretation. Even when a comparable reliability can be assessed (results for a two-person mixed sample are generally expected to be more reliable than those for a four-person mixed sample, for example), there is no threshold or criteria established to determine what is an acceptable level of reliability.

4.4.4. Comments on Available Data

Historically, forensic laboratories have not publicly shared internal validation data for review by those outside their laboratories. For some stakeholders, freedom of information or court-ordered discovery requests can enable access to specific data or information. However, these requests also do not typically make the underlying data *publicly available* for independent scientific assessment.

One explanation for this lack of public data is simply that there has been no expectation to provide it. Choosing not to make public the data underlying decisions that are made in laboratory protocols is generally without consequence, while giving public access carries a risk of increased scrutiny. A recent call for a more collaborative approach to validation studies (Wickenheiser & Farrell 2020) may encourage more open community data sharing. Science progresses best when it can be critically assessed by other scientists, which is, of course, an important purpose of peer-review publication. This point is highlighted in the National Academy of Sciences' publication *On Being a Scientist: A Guide to Responsible Conduct in Research* (NAS 2009).

Potential reasons why forensic laboratories choose not to make their internal validation data publicly available include: (1) the information from a study itself may not be publishable[23]

[23] The willingness of journals to publish validation studies is a separate issue from the willingness of laboratories to make data available on their website for anyone to download or at least sharing full data sets with credible parties in a timely manner when requested.

due to lack of novelty (e.g., Buckleton 2009), (2) genotype data may include information from donors who did not consent to public sharing of their DNA profiles (e.g., Manabe et al. 2017), and (3) sharing foundational data is not required by current accreditation or guidance documents. Table 4.9 summarizes issues with available information from the data sources examined in this scientific foundation review.

Table 4.9. Issues with available information for the data sources examined in this study.

Data Sources	Issues with Available Information	Recommendations
Published Developmental Validation of STR Typing Kits (see Table 4.2)	• typically a single two-person mixture is evaluated with various mixture ratios to explore limits of detection for non-overlapping alleles in minor contributors • studies focus on the range of reliability for generating STR profiles with single-source samples using sensitivity, reproducibility, concordance, heterozygote balance, and stutter product ratios • robustness is also examined for STR typing kit components and factors, such as PCR master mix composition, PCR cycle number, differing annealing temperatures, primer concentrations, and species specificity	Recognize that these studies cover only a small portion of the factor space; they are useful for demonstrating reliability and robustness with single-source samples; however, these studies cannot be used to assess the degree of reliability for complex DNA mixture interpretation
PGS Validation Publications (see Tables 4.3 and 4.4)	• a lack of uniformity and data details makes comparing information across studies difficult • the following are not consistently provided: contributor genotypes or degree of allele sharing, EPGs of mixtures, ground truth information on the number of contributors (see Box 4.1)	Adopt a community-wide uniform approach to publishing information (e.g., Bright et al. 2019a, Rodriguez et al. 2019) to enable independent assessment of PGS performance (see Box 4.1)
Internal Validation Data and Summaries (see Table 4.5)	• few forensic laboratories currently provide publicly available internal validation data or summaries • contributor genotypes or degree of allele sharing is rarely provided	Adopt a community-wide uniform approach to sharing internal validation information and data to enable independent assessment of PGS performance (see Box 4.1)
Proficiency Tests (see Tables 4.6 and 4.7)	• mixture PTs consist mainly of simple mixtures with high-quality and quantity DNA and some PTs only utilize single-source samples (e.g., Hundl et al. 2020)	Require more challenging PT samples (e.g., UKFSR 2020)containing low-level, degraded DNA and mixtures with more than two contributors
Interlaboratory Studies (see Table 4.8)	• most previous studies are not relevant to PGS methods in use today	Future studies would benefit from data gathered independent of PGS developers[a]

[a] In October 2020, the National Institute of Justice funded Noblis and Bode Technology to study interlaboratory variation in interpretation of DNA mixtures (see https://nij.ojp.gov/funding/awards/2020-r2-cx-0049).

KEY TAKEAWAY #4.5: Current proficiency tests are focused on single-source samples and simple two-person mixtures with large quantities of DNA. To appropriately assess the ability of analysts to interpret complex DNA mixtures, proficiency tests should evolve to address mixtures with low-template components or more than two contributors – samples of the type often seen in modern casework.

KEY TAKEAWAY #4.6: Different analysts and different laboratories will have different approaches to interpreting the same DNA mixture. This introduces variability and uncertainty in DNA mixture interpretation. Improvements across the entire community are expected with an increased understanding of the causes of variability among laboratories and analysts.

4.4.5. Bracketing Approach

It is unrealistic to obtain and examine the volume of samples needed in order to provide complete coverage of the potential factor space with DNA mixture interpretation. Therefore, a practical solution is to map regions of the factor space. To investigate case-specific reliability of the laboratory's measurement and interpretation process, an analyst can use ground truth from known samples *similar* to the casework sample of interest and study the results. A *bracketing approach*[24], which considers results from samples that are more complex and less complex than the casework sample of interest, is a sensible way of understanding case-specific reliability of the system. Indeed, publicly available information from validation studies, PT results, and interlaboratory studies only cover a portion of the possible factor space (Tables 4.2 to 4.8) – suggesting that a bracketing approach may be needed to inform method performance with specific casework samples.

Particular attention should be paid to validation data for DNA mixture interpretations that are expected to have a high degree of uncertainty, for example, when a contributor of interest has contributed very low DNA template quantities, or there are large amounts of allele sharing, or many contributors in the sample. While access to internal validation summary reports provide the ability to see trends in results and the types of experiments that have been performed, only access to individual data points and accompanying metadata (i.e., information about the data) can enable a full independent review.

On the question "Are currently used PGS systems reliable?" the answer is "It depends." It depends on the region of the factor space for the case sample of interest and coverage with available ground truth data for assessing reliability.

KEY TAKEAWAY #4.7: The degree of reliability of a PGS system when interpreting a DNA mixture can be judged based on validation studies using known samples that are similar in complexity to the sample in the case. To enable users of results to assess the degree of reliability in the case of interest, it would be helpful to include these validation performance results in the case file and report.

[24] This concept was originally proposed by Steven Lund of the NIST Statistical Engineering Division and presented to the Resource Group at a meeting in April 2018.

4.4.6. Comments on Likelihood Ratio Values

The process of interpreting DNA mixtures is guided by principles of the underlying biology as well as statistical representations of the empirically observed relationship between genotypes and EPGs, all of which are often combined and codified in the form of models. Their fitness for any given purpose is informed by results of validation studies involving test runs with ground-truth known data and covering the space of anticipated application scenarios.

Writing in 2018, the ISFG DNA Commission stated:

> "… there are no true likelihood ratios, just like there are no true models… Depending on our assumptions, our knowledge and the results we want to assess, different models will be adopted, hence *different values for the LR will be obtained*. It is therefore important to outline in our statements what factors impact evaluation (propositions, information, assumptions, data, and choice of model)" (Gill et al. 2018, emphasis added).

Different experts using different assumptions, different statistical models, and different inference procedures may arrive at different LR values. Information regarding the extent to which their LR values agree or disagree is typically not available. There appears to be a general misconception that LR assessments made by different experts will be close enough to one another to not materially affect the outcome of a case. Although they may be close enough in many instances, this is not known for any particular case and it is not advisable to take this for granted.

In addition, there are a number of different LR values that can be generated by a PGS system, such as a highest posterior density (HPD) LR to adjust for sampling uncertainty, a unified LR to account for both related and unrelated individuals under the defense proposition, a population stratified LR to incorporate relative proportions of different subpopulations, a variable number of contributors (varNOC) LR estimation, or various combinations of these LR adjustments (Kelly et al. 2020). Appreciating the assumptions and information provided by each of these numbers is important to communicating what a specific LR value reflects (see Table 2.4).

The degree of reliability or trustworthiness of a given PGS method in a given case is dependent upon the number of instances where that method has been tested with samples that are judged to be of similar complexity as the casework sample, the performance of the method among those instances, and how the characteristics (e.g., number of contributors, DNA amounts, level of degradation) of the ground truth known samples compared to those of the sample in the case at hand. More validation samples and denser coverage of the space of application scenarios provide better estimates of casework-relevant reliability metrics and error rate estimates.

The desired performance for a DNA mixture interpretation PGS model is often described in terms of trends. For example, authors of the STRmix developmental validation study write:

> "the log(LR) for known contributors (Hp true) should be high and should *trend* to 0 as less information is present within the profile. Information includes the amount of DNA from the contributor of interest, conditioning profiles (for example, the victim's profile on intimate samples), PCR replicates, and decreasing number of contributors… The LR should *trend* upwards to neutral [for known non-contributors] as less information is present within the profile" (Bright et al. 2016, emphasis added).

Beyond producing LR values that follow expected trends (which are an important starting point), it is valuable to also consider the question: Does a particular PGS system provide LR values that appropriately represent the strength of the evidence? This is a much harder question to answer and requires more data than required for simply illustrating trends.

Since repeatability and reproducibility are components of reliability, it is fair to ask to what extent the LR values offered by different experts using different databases and different models differ from one another. If the accuracy and reliability of a specific LR assignment is important to a case, then understanding what level of reproducibility there is between laboratories or between forensic scientists will help assess reliability. Whereas each laboratory or expert may feel justified in considering their assessments to be reliable, the recipients of such assessments in a given case need guidance on what to do in situations where variation among different LR assessments could impact the outcome of a trial. In particular, because there are no standards to compare to and no traceability considerations as there are for measurements, judgments of reliability by decision makers or triers of fact will be helped by comparing LR assessments from multiple systems and made by multiple experts (Gill et al. 2015).

Likelihood ratios must satisfy an internal consistency requirement (called the property of being well-calibrated or "calibration accuracy," for short) which can be empirically tested (Ramos & Gonzalez-Rodriguez 2013, Meuwly et al. 2017, Hannig et al. 2019). The scientific validity of any particular PGS system used in casework can be assessed, at least partly, by investigating (1) repeatability, (2) reproducibility, (3) calibration accuracy, and (4) the efficiency or discriminating power. Such an exercise will help identify the better-performing PGS systems for consideration in casework applications.

The accuracy of the LR assessment in any specific casework situation cannot be determined. However, results of LR assessments across a collection of casework-similar, ground-truth known, scenarios can assist in informing the receiver of the LR assessment as to how much weight should be given to the LR assessment in the case at hand.

The specific propositions selected impact the LR values obtained (see Table 2.4). This fact should encourage more effort to standardize development of propositions as it has been noted: "The truth lies in the propositions: either the prosecution proposition is true or the [defense] proposition is true" (Gittelson et al. 2018). The implicit assumption in this statement is that the propositions are exhaustive. Otherwise, there is the possibility that neither the prosecution proposition nor the defense proposition is true. Ground truth information can only tell us whether H_1 (Hp) is true or H_2 (Hd) is true, but it cannot tell us what the LR value should be. Studies can, however, estimate the percentage of time the LR values are on the wrong side of 0 when using log(LR)) and providing adventitious

exclusionary or inclusionary support (see Riman et al. 2021). Sometimes, data may be favorable to H_1 even when H_2 is true. This happens not just due to adventitious matches but also due to limitations of models.

As forensic laboratories share their validation summaries *and data used for making decisions* to enable future independent review of their work, the field has the opportunity to be strengthened. Tables with sample details and LR values have been made available as supplemental files in some publications (e.g., Bright et al. 2019a, Rodriguez et al. 2019). When only aggregate graphs are provided in publications (e.g., Taylor 2014) or validation summaries without specific metadata for the data points displayed, there is no ability to correlate the data and samples used to generate them. Aggregate graphs can also make it challenging for users of data to understand what aspect of the factor space is being covered in the experiments being reported (e.g., see the number of N.E.S. [not explicitly stated] fields in Table 4.5 examining publicly available internal validation summaries).

KEY TAKEAWAY #4.8: We encourage a separate scientific foundation review on the topic of likelihood ratios in forensic science and how LRs are calculated, understood, and communicated.

4.5. Thoughts on a Path Forward

The discussion section of this chapter (Section 4.4) comments on limitations in currently available data from PGS systems used for DNA mixture interpretation. This section describes a path forward in terms of desired data when conducting independent scientific assessments for LR values assigned by PGS systems and ways that these data might be evaluated to provide increased confidence in these results. Interested readers may also wish to consult slides from a September 2020 validation workshop[25] covering discrimination power and LR accuracy calibration. This workshop covers use of receiver operating characteristics (ROC) plots and illustration of calibration.

4.5.1. Desired Data to Benefit Independent Scientific Assessments

Not only is available information limited as described above, sometimes helpful, or even essential, information is missing. This makes it impossible to know what has actually been examined in a particular study. Note the "N.E.S." designations throughout Table 4.3 and Table 4.5 highlighting where important information is not explicitly stated in the referenced publication. Thus, the community would benefit from a more uniform approach to both sharing information generally and sharing needed information to enable independent scientific assessments of PGS and other DNA mixture interpretation studies performed.

The value of having a standard set of information to share when describing validation data can be seen with an approach taken by the digital PCR (dPCR) community, where "Minimum Information for Publication of Quantitative Digital PCR Experiments" has been adopted and recently updated (dMIQE Group 2020). This group notes:

[25] See https://strbase.nist.gov/pub_pres/ISHI2020-ValidationWorkshop-Butler_Iyer-Slides.pdf

> "To assist independent corroboration of conclusions, comprehensive disclosure of all relevant experimental details is required. To support the community and reflect the growing use of dPCR, we present an update to dMIQE, dMIQE2020, including a simplified dMIQE table format to assist researchers in providing key experimental information and understanding of the associated experimental process. Adoption of dMIQE2020 by the scientific community will assist in standardizing experimental protocols, maximize efficient utilization of resources, and further enhance the impact of this powerful technology" (dMIQE Group 2020).

The dPCR community has found it beneficial to supply a checklist of essential information that can be used by authors, reviewers, and editors when research articles are submitted for publication. This checklist includes details on specimens (types, numbers, sampling, storage), nucleic acid extraction (description of methods, volume used, number of replicates), dPCR protocol (instrument and model, primer and probe concentrations, template treatment, complete thermocycling parameters), assay validation (analytical specificity, analytical sensitivity, testing for inhibitors), and data analysis (description of dPCR experimental design, comprehensive details on negative and positive controls, repeatability, reproducibility, number of partitions measured, partition volume, statistical methods used for analysis, data transparency). For data transparency, raw data from dPCR experiments may be included as supplemental files.

In a spreadsheet that must be completed when a dPCR manuscript is submitted for publication, authors indicate "yes" or "no" for each item on the dMIQE2020 list. When "yes" is selected, a comment box in the spreadsheet can be used to describe the location of the required information (e.g., in a specific supplemental table to the manuscript). When "no" is selected, the comment box is used to outline rationale for the omission, such as why a particular item may not apply depending on the experiment(s) performed.

Adoption of a similar approach would benefit the forensic DNA community with future DNA mixture interpretation assessments to avoid omission of essential information in publications. Similar guidelines for minimum information on PGS validation experiments could be developed by SWGDAM[26] or the OSAC Human Forensic Biology Subcommittee[27].

Box 4.1 includes desired information for reliability assessments of LR values assigned in PGS systems that can enable a quantitative assessment of these LR results. Availability of this information should enable assessment of discrimination power and LR calibration accuracy for associated method(s).

[26] https://www.swgdam.org/
[27] https://www.nist.gov/osac/human-forensic-biology-subcommittee

Box 4.1. Desired Information for Reliability Assessments of LR Values in PGS Systems

The following information should help an independent reviewer assess reliability of a DNA measurement and interpretation (end to end) system. With this information, reliability assessments could include (1) assessment of discrimination ability, (2) LR value calibration accuracy in PGS systems, and (3) some exploration of regions of the factor space where LR values assigned by a PGS system are more reliable versus less reliable. If such data are available for different PGS systems, then a performance comparison may be possible (e.g., You & Balding 2019).

1. Sample Number or Unique Identifier
2. Number of Contributors (NOC)
3. Target DNA Template Amounts
4. Degradation Status of DNA Template(s)
5. NOC used for Analysis (Apparent NOC)
6. H_1 true? (Yes/No)
7. Person of Interest (POI) position in the mixture (if H_1 is true)
8. Reported Log_{10}(LR)
9. Mixture EPG results*
10. POI profile*
11. Known contributor A profile* and any additional known contributors
12. Noncontributor profile (if H_1 is not true): is this profile simulated or determined from an actual sample?
13. Analytical threshold used for analysis
14. PGS parameters and settings

* If privacy of the profile genotypes is a concern, then alleles could be used in an algebraic format as described previously (Gill et al. 1998). For example, the letters A, B, C, D, etc. can be used in place of actual alleles at the various loci.

Within the digital PCR community, dMIQE requirements have (1) enabled authors to design, perform, and report experiments with greater scientific integrity, (2) facilitated replication of experiments described in published studies where these guidelines are followed, and (3) provided critical information that allows reviewers, editors, and the wider scientific community to measure the technical quality of submitted manuscripts against an established standard (dMIQE Group 2020).

A similar approach to the dMIQE data reporting requirements with studies involving PGS systems would benefit the forensic DNA community – both practitioners and users of their data. In addition, sharing more details on validation experiments could provide community-wide cost savings using a collaborative validation approach (Wickenheiser & Farrell 2020).

4.5.2. Performance Testing with Case-Similar Data

Generally speaking, models and interpretation methodologies developed using known DNA samples may be expected to perform satisfactorily (fit for purpose) when applied in new but similar scenarios. However, their suitability for application in scenarios not represented adequately within the available empirical data used during model training is questionable. For example, models developed using known samples involving at most two contributors may perform well in other two-person mixture scenarios but may perform poorly when applied in situations involving three or more contributors.

PGS models developed using samples covering a specific region of the factor space may work well for similar situations but may or may not work satisfactorily when applied to data that are unlike any of the scenarios considered in the training set; for example, using mixture data with at most five donors involving sufficient DNA amounts from each donor to reduce the possibility of stochastic effects (e.g., 100 pg or more). Mapping the factor space coverage of PGS testing (e.g., Table 4.3 and Table 4.5) can assist in understanding the limits of application scenarios for any given interpretation strategy. Identification of those scenarios where the performance of a specific method is judged to be inadequate will assist in establishing operational limits for the types of samples that may be reliably interpreted and also point to areas where the measurements or models require improvements.

Alternatively, it may be the case that demonstrating, based on a large number of ground-truth known samples, a method performs well in scenarios more complex than the case at hand (e.g., test cases with more contributors, less DNA template, or more degradation) inspires confidence that the method performs well in scenarios like the case at hand, even when there are few (or no) ground-truth-known samples with closely matching characteristics.

As described in Section 4.4.5, the "bracketing approach" is a pragmatic solution considering the vast number of different mixture scenarios that might be encountered in casework[28]. Running thousands of validation experiments to cover all potential factor space for complex DNA mixtures is not practical. Additionally, this approach provides a potential guideline for identifying the limits among a given body of validation experiments. That is, casework samples are considered outside the limits of that body of validation experiments if there does not exist a collection of ground-truth-known analyses among scenarios as difficult as or more difficult than the casework sample that convincingly support the performance of the considered method.

A single binary (i.e., yes/no) statement of reliability, based on aggregate performance across many types of samples and many different PGS systems, does not provide the information needed to judge the reliability of the measurement and interpretation in a particular case of interest. Rather what is needed in the context of a specific case is information concerning the performance of these methods when applied in casework-similar scenarios.

[28] Note that one need not consider all validation samples more difficult than the case at hand when evaluating performance. For example, if a casework sample had two contributors each with an estimated 100 pg, one might consider the method's performance among validation experiments conducted with three contributors each with 100 pg and additional validation experiments conducted with two contributors each with 50 pg but exclude validation experiments conducted with contributors each with 10 pg.

4.5.3. Summary

Statistical tools are available for examining discrimination efficiency, especially for comparing two or more PGS systems. Receiver operating characteristics (ROC) plots are a commonly used tool for this purpose and have been used in evaluation of PGS systems previously (e.g., Bleka et al. 2016b, You & Balding 2019). Tools for examining calibration accuracy of LR assignments (e.g., Ramos et al. 2013, Hannig et al. 2019) are less widely known to forensic DNA analysts.

Though component-level reliabilities eventually determine system reliability, it is the system reliability that is of direct interest in applications. Journal articles discussing reliability of PGS systems often address only the reliability of specific components and, unless careful attention is given to details regarding which of the reliability-influencing factors were varied in the study, there is a danger of inadvertently viewing results from narrowly-focused studies as applicable to system reliability.

There are many sources of uncertainty to consider when examining DNA mixture interpretation. Presence of multiple sources of uncertainty, by itself, does not decrease reliability of strength-of-evidence assessments. If the sources of uncertainty are acknowledged and correctly modeled, the resulting LR statements are expected to be well-calibrated. If all (or almost all, in practice) of the discriminating (between H_1 and H_2) information present in the sample has been used in the LR assessment, then the PGS system is expected to have good discrimination power. Regardless of sources of uncertainty and complexity of the samples, reliability of a PGS system boils down to checking its calibration accuracy and discriminating power at every conceivable scenario described by the factor space. A limitation to any reliability assessment is going to be the amount of casework-similar empirical data that is available for comparison in each specific case.

In the end, the reliability of LR values produced by a PGS system means little if relevance of the DNA evidence has not been established first (see Chapter 5 in this report).

5. Chapter 5: Context and Relevance Related to DNA Mixture Interpretation

This chapter considers foundational issues regarding the relevance of DNA test results in criminal investigations, particularly when small quantities of DNA are examined. We review the literature on mechanisms of DNA transfer, factors that affect the variability of transfer and persistence, and the potential transfer of contaminating DNA at any stage in an investigation. These studies show it is possible to handle an item without transferring any detectable DNA to that item, that DNA may have been deposited before the crime and therefore may not be relevant to the crime, and that DNA might be present due to indirect (secondary or tertiary) transfer. A common theme from the DNA transfer literature is that association of a reference sample from a person of interest with a crime scene sample cannot automatically be used to infer involvement with the crime. We also review the literature on case types dealing with transfer and methods of interpretation. We consider the implications of the reviewed studies and outline strategies for dealing with questions of DNA transfer. The suggested strategies are (1) to minimize contamination at all stages, not just in the laboratory; (2) to consider evidence in context, because the same findings will have different significance in different circumstances; (3) to ask and answer appropriate questions and work to ensure that stakeholders do not use the answer to a source (or sub-source) proposition to address activity or offence propositions; (4) to use the Case Assessment and Interpretation model to identify the most probative samples and the hierarchy of propositions to identify the appropriate questions to be addressed; and (5) to separate investigation from evaluation, realizing that a sub-source likelihood ratio (LR), which is very useful to identify a suspect, will need to be further evaluated for use in court.

5.1. Introduction

Every contact leaves a trace. This phrase, often associated with the early French forensic scientist Edmond Locard, explains why investigators often seek support for two items having been in contact. However, what Locard actually said was:

> "The truth is that none can act with the intensity induced by criminal activities without leaving multiple traces of his path" (cited in Roux et al. 2015).

With this, we can see that the aphorism, *every contact leaves a trace*, is an oversimplification. Locard's statement implies at least two things. First, the trace is not only associated with the fact of contact, but also with an activity of greater or lesser intensity. Second, multiple traces of the activity can be expected, and therefore it would be inadequate to consider only a single trace in isolation.

Furthermore, to the extent that every contact does leave a trace, we need a way to separate the relevant traces—those associated with the commission of the crime—from the irrelevant ones. In earlier times, separating the relevant from the irrelevant presented less of a challenge because relatively large amounts of DNA were needed to produce a profile. With samples containing a large amount of DNA (e.g., a bloodstain the size of a coin), common sense was often sufficient for determining relevance. For example, with a visible blood or semen stain, the cell type could be determined, and the activity that caused a sample to be deposited could often be inferred, even by nonexperts.

That situation changed with the advent of methods that can detect very small quantities of DNA. The 1997 *Nature* publication "DNA Fingerprints from Fingerprints" (van Oorschot & Jones 1997) demonstrated that DNA could be recovered from touched samples, which are invisible and may not have an easily identifiable cell type. In addition, DNA can transfer readily under some circumstances (e.g., Szkuta et al. 2017b) and can persist for fairly long periods of time (e.g., van Oorschot et al. 2014a). Our summary of the above papers is that the relevance of a DNA sample to the crime is often difficult to discern.

Forensic science typically involves investigating multiple pieces of evidence in an effort to shed light on a past event that has taken place at a particular moment in time. Figure 5.1 illustrates the opportunities for transfer of DNA at various stages before, during, and after a crime event. These multiple transfers mean that DNA found at a crime scene may be irrelevant to the crime, and, furthermore, that the DNA present is often in the form of a DNA mixture, which further complicates the process of interpretation.

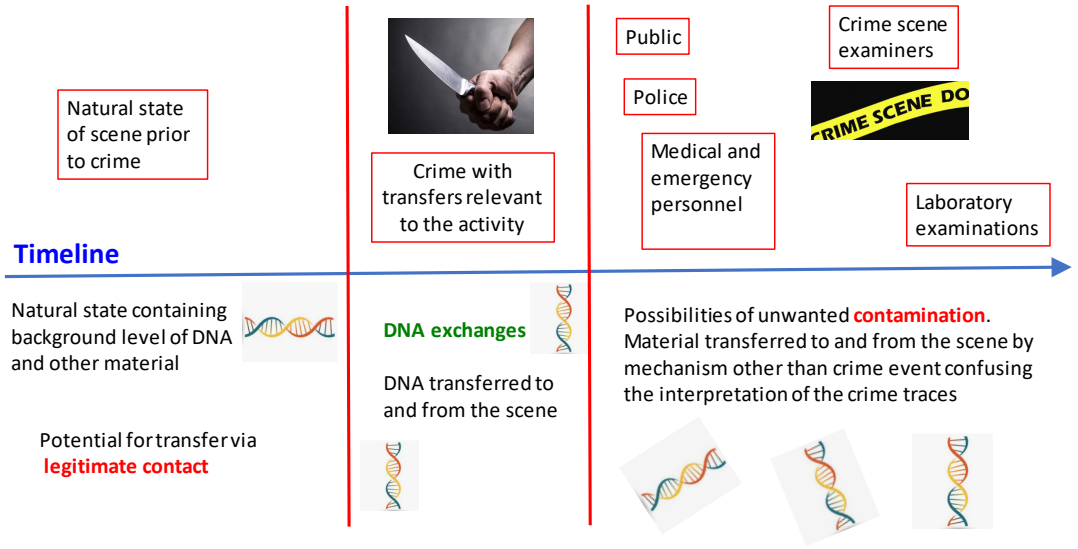

Figure. 5.1. Timeline illustrating the potential for transfer via legitimate contact before the crime activity DNA exchange and the possibility of contamination after the crime event (adapted from Gill 2002).

To properly assess the relevance of a DNA sample to a crime event, it is necessary to understand the factors that affect the transfer of DNA and how long it persists in different circumstances. This chapter reviews the literature on this subject.

KEY TAKEAWAY #5.1: DNA can be transferred from one surface or person to another, and this can potentially happen multiple times. Therefore, the DNA present on an evidence item may be unrelated (irrelevant) to the crime being investigated.

5.2. Data Sources Used

The information in this chapter is based on peer-reviewed literature, most of which was found via multiple searches of the PubMed database.[29] A search for "trace DNA" conducted on October 4, 2018, found 4085 papers. Most of the references from this search were not related to forensic DNA applications. Those relevant to small quantities of DNA for use in criminal investigations were retained. Further PubMed searches for "transfer, mixture DNA" in October 2018 located 270 articles, which were checked for relevance. We also found additional studies cited in the reference lists from three review articles that preceded our study (Wickenheiser 2002, Meakin & Jamieson 2013, Gill et al. 2015) and several additional reviews that were published during the course of our study (Taylor et al. 2018, Burrill et al. 2019, van Oorschot et al. 2019, Gosch & Courts 2019).

We divided the topics presented in the collected literature into several subject areas, as shown in Table 5.1.

Table 5.1. Subject areas examined as part of this review.

Subject Area	Number of Articles Reviewed [a]	Comments
Mechanisms of DNA transfer	16	Studies on how DNA transfers
Structured experiments to examine key variables affecting DNA transfer	40	Includes overlap with persistence
Studies on DNA transfer that mimic casework scenarios	19	Relevant to transfer and not covered in structured studies or casework section
Studies on contamination	26	Mainly studies to identify sources
Interpretation and evaluation	28	Papers particularly relevant to the issues outlined in this chapter
Casework reports	21	Useful information collated or taken from individual cases

[a] We categorized the articles we reviewed according to the main message of the paper, but there is overlap, particularly among transfer, contamination, and casework. Additional sources, such as textbooks or classic references, are cited throughout the text and listed at the end of the chapter.

[29] https://www.ncbi.nlm.nih.gov/pubmed

5.2.1. Obstacles to Comparing Data Across Studies

The existing studies are difficult to compare with each other for various reasons. For instance, different laboratories use different analytical parameters, which result in different strengths of evidence. In addition, over the past 20 years, sensitivity of detection has increased, and aspects of how DNA profiles are produced have changed (see Chapter 2 and Appendix 1). These changes mean that a study from 2000 is not directly comparable with one from 2019.

For example, researchers may measure the efficiency of transfer based on the percentage of alleles detected, but there are subtle differences in this approach depending on how homozygous and shared alleles are counted. In other studies, only unique alleles are used to assess transfer efficiency. Sometimes this is not an issue because, unlike in a crime scene scenario, the contributors' profiles are known in a controlled research study. However, if a study records the criteria used in casework to assess their findings (e.g., Breathnach et al. 2016), a different set of criteria in another laboratory may make it difficult to compare results across studies.

DNA transfer studies have also increased in complexity. Many now consider multiple transfers and, as in real casework, consider profiles from very small quantities of DNA. This has prompted use of probabilistic genotyping software (PGS) LR assignments rather than allele counting, which adds to the difficulty in comparing results across studies.

5.3. Reviewing the Data

The contents of the reviewed publications were taken as an overall view of the current state of knowledge. A great deal remains unknown about mechanisms of primary DNA transfer, about the factors that affect secondary and higher-order transfers (transfer via one or more intermediaries, which can be animate or inanimate), and persistence. Strategies to improve research by gathering more systematic data have recently been suggested (Gosch & Courts 2019).

5.3.1. Mechanisms of DNA Transfer

Although there is widespread acceptance in the literature and in practice that DNA transfers, there is relatively little research on the actual mechanisms of transfer.

The term *touch DNA* is frequently used, but there is a lack of clarity about the underlying processes that allow recovery of DNA when an item is handled. The most common view is that DNA originates from skin cells shed during the action of touching (Hanson et al. 2011). There is disagreement on this view, however, because the outer skin cells have no nuclei and therefore are not expected to contain nuclear DNA. There are alternative theories, but the number of studies as seen in the following paragraphs is limited.

Attempts to identify cell types via RNA analysis have been carried out in conjunction with nuclear DNA studies. A group of 22 collaborating laboratories carried out simultaneous extraction of RNA and DNA in order to identify the tissue source of the DNA and had some success with skin markers (Haas et al. 2015). Five messenger RNA (mRNA) markers were

identified that demonstrated a high degree of specificity for skin. The use of these markers has enabled the detection and identification of skin using as little as approximately 5 pg to 25 pg of input total RNA from skin and, significantly, in swabs of human skin and various touched objects (Hanson et al. 2012). These researchers acknowledge that if touch DNA consisted of naked DNA in body secretions such as sweat or sebaceous fluid, skin-specific mRNA markers may be present at a concentration too low to be detected.

Several mRNA markers were used to determine whether different epidermal layers could help identify the type of activity, such as a firm grip or a casual touch, that gave rise to a transfer (Bhoelai et al. 2013). The study did not establish any relationship with the type of contact.

Because of the possibility that DNA may be transferred either in sweat or sebaceous fluid, there is a question as to whether touch-related DNA profiles come from extranuclear DNA rather than nuclear DNA in shed skin cells (Quinones & Daniel 2012, Zoppis et al. 2014). Testing of sweat collected from volunteers yielded an average of 11.5 ng of DNA from 1 mL cell-free sweat samples. This observation prompted the proposition that DNA transferred through the act of touching consists of cell-free nucleic acids of length suited for STR analysis (Quinones & Daniel 2012). Another study suggested that DNA fragments on touched objects may originate from the epidermal cells of the cornified layer that are constantly sloughed off and leave the skin surface with sweat (Kita et al. 2008).

A morphological study using microscopy and immunology reported the following: "When swabs from touch samples were analyzed, using imaging and flow cytometry, 84–100% of DNA detected was extracellular" (Stanciu et al. 2015). These experiments involved volunteers who held objects, with some having been asked to wash their hands prior to handling the objects. Hand washing resulted in a decrease in the amount of extracellular DNA but did not have a significant impact on the number of epidermal cells detected. The flow cytometry experiments showed two distinct fractions—fully differentiated keratinocytes (i.e., corneocytes) and cellular debris/fragments. Buccal cells were not observed, indicating saliva was not a significant source of the DNA found on subjects' hands (Stanciu et al. 2015).

It has been postulated that DNA in touch samples is transferred in the sebaceous fluid (Zoppis et al. 2014). These studies found that the ability to shed sebaceous fluid had a major influence on secondary transfer, which supports the view that dividing participants into good and bad shedders (see section 5.3.2.1) is too simplistic. Instead, the ability to shed sebaceous fluid will vary with age, hormonal condition, skin diseases, and the part of the skin that touched an object (e.g., Kamphausen et al. 2012). The relative tendency of fingertips or palms to produce DNA was examined with the view that the tips were the better source (Oleiwi et al. 2015). This study supports the claim that palms have relatively fewer sebaceous pores (Zoppis et al. 2014).

Some work has focused on the potential loss of DNA during extraction, with the possibility that touch samples may benefit from improved extraction methods (Vandewoestyne et al. 2013). It has been noted that a better understanding of the mechanism for DNA transfer will "increase our confidence in assigning a weight to DNA evidence obtained in such circumstances" (Quinones & Daniel 2012).

Researchers studying glass slides touched by donors have commented:

> "The underlying science of touch DNA recovered from criminal casework is directly related to the basic biology and genetics of normal skin regeneration and programmed cell death (apoptosis) and lends an understanding of the inherent variability in DNA recovery from handled items" (Hazell-Smithen et al. 2014).

This perspective is supported by an alternative method of sample collection involving searching surfaces for clumps of cells (Hanson & Ballantyne 2013, Farash et al. 2015, Farash et al. 2018). The approach of physically separating cells on a surface (see Chapter 6) has the advantage of being able to generate single-source DNA profiles and thus avoid the complex mixtures that arise when swabbing a surface containing cellular deposits from multiple individuals.

While the number of studies is low, it seems that the sensitivity in DNA testing today is sufficient to generate a profile from cornified layer cells (Kita et al. 2008) that still contain DNA. The cornified layer and apoptosis may account for the possibility of additional alleles from degraded cells. Almost every transfer study discussed in this chapter has unexpected additional alleles that would support the possibility of cell-free DNA being present.

Figure 5.2 illustrates potential sources of DNA from touch evidence sample deposits taken from a recent comprehensive review on the topic, where the authors state: "Our current understanding of the cellular content of touch deposits and the origin of the potential trace DNA therein is extremely limited" (Burrill et al. 2019).

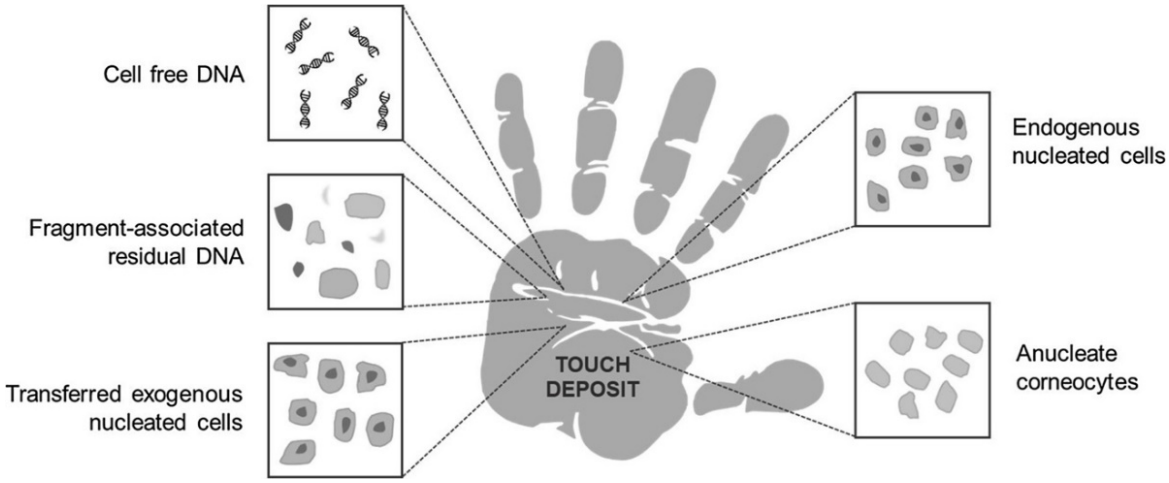

Figure 5.2. Concept map of potential sources of DNA deposited by touch/handling. It is currently well established that individuals may leave behind detectable DNA when they handle items, but the anatomical origin of that DNA remains unsolved. It is possible that the DNA typically recovered from handled items in forensic scenarios comes from nucleated cells from hands, anucleate cells from hands, nucleated cells transferred onto hands from elsewhere, residual cell fragments (including free nuclei) from hands, or from outside a cellular architecture in sweat on hands or residual transferred body fluids. Reproduced with permission from Burrill et al. (2019).

The most recent work from these researchers "raise questions about shed corneocyte DNA content previously assumed to be negligible" (Burrill et al. 2020).

5.3.2. Structured Experiments to Examine Key Variables Affecting DNA Transfer

Several studies have been conducted to assess factors that affect transfer and persistence of DNA. This transfer may occur with blood or saliva or small quantities of DNA of unknown cell type. Available studies can be divided into two broad categories: (1) systematic studies that examine variables affecting transfer and persistence of DNA, and (2) studies carried out to address specific case-like situations.

Table 5.2 provides details on structured experiments that examined key variables for transfer and persistence of DNA. These publications record a number of variables. The purpose of each study and key findings have been summarized. Comparison of findings across these studies is difficult because the criteria used and the methods used to measure transfer have evolved over time (e.g., different STR kits and PCR conditions).

A number of studies covered the following four topics, which are discussed in more detail below. The first topic involves **shedder status**, in which experiments are conducted to assess whether an individual sheds low or high amounts of DNA. The second topic involves **substrate effects**, in which experiments examine how DNA transfer is affected by the surface where the sample is deposited. The third topic involves **persistence studies**, which examine the length of time DNA can be detected on a surface following deposition. The fourth topic involves studies concerning **non-self-DNA on individuals**, in which experiments are conducted looking for DNA not associated with the individual who touched an item.

Other variables that affected DNA transfer in these studies included moisture (Goray et al. 2010a, Lehmann et al. 2013, Verdon et al. 2013), pressure (Tobias et al. 2017), and friction (Verdon et al. 2013).

Table 5.2. Studies involving structured experiments to examine key variables for transfer and persistence of DNA.

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
1	van Oorschot and Jones (1997) DNA fingerprints from fingerprints	Various tests with 1 to 4 repeats Profiles: 2 ng to 150 ng DNA	Can a profile be generated from items participants touch?	Profiles generated 13/13; secondary transfer noted
2	Lowe et al. (2002) The propensity of individuals to deposit DNA and secondary transfer of low-level DNA from individuals to inert surfaces	8 participants, 3 time intervals repeated 5 times; 22 participants, one time interval repeated 3 times; 2 pairs, 3 time intervals, 5 replicates % profiles obtained	Study secondary transfer of DNA when body fluid is not known	Secondary transfer is possible; participants differ in their propensity to deposit DNA; time since handwashing is a key variable

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
3	Phipps & Petricevic (2007) The tendency of individuals to transfer DNA to handled items	60 participants with 5 volunteers chosen to test good shedder/ bad shedder theory; tested over four days Number of alleles obtained with relative proportion of full profiles, partial profiles, and no results	Check transfer of DNA and repeat Lowe et al. (2002) study	Handwashing is not a key factor as an individual cannot be relied upon to shed a consistent amount of DNA over time; results indicated that it may be more difficult than expected to classify individuals as *good* or *bad* shedders
4	Farmen et al. (2008) Assessment of individual shedder status and implications for secondary DNA transfer	9 participants tested with palms swabbed at two time intervals; handshakes followed more swabbing and holding a beaker Number of matching alleles	Assess shedder status and check effect on secondary transfer	Shedder categorization confirmed with a good shedder picked up on other participants' hands and objects; transfer noted on all occasions in this study
5	Goray et al. (2010a) Secondary DNA transfer of biological substances under varying test conditions	DNA 5 µL/mL, blood, saliva on wool, cotton, and plastic using passive, pressure, and friction; each combination replicated four times % DNA transferred	Factors affecting secondary transfer; deposit including moisture level, the primary and secondary substrate, and type of contact	Initial deposit of DNA was 20 times greater when deposited onto porous cotton surface than onto a smooth and hard plastic surface, with less in reverse; nature of substrate and moisture were significant; other biological materials were the same
6	Goray et al. (2010b) Investigation of secondary DNA transfer of skin cells under controlled test conditions	One donor produced DNA skin cells; 6 times for each variable; 1 and 2 substrate, passive, pressure, and friction % DNA transferred; initial amounts of DNA needed to transfer to generate good profile (1 ng at that time) measured; results varied with conditions from 385 ng to 2 ng	Study of factors affecting secondary transfer of skin cells	Freshness of deposit not a factor; friction increased rate of transfer; skin cells deposited onto nonporous substrate transfer more readily but further transfers facilitated more by porous substrate. Nonporous to porous with friction most effective
7	Daly et al. (2012) The transfer of touch DNA from hands to glass, fabric, and wood	300 participants, 50/50 male /female held in their fist for 60 s; no distinction made between dominant or no-dominant hand Gene scanner and gene mapper 50 relative fluorescence units (RFU) for heterozygous and 200 for homozygous	Check the variation onto glass, wood, and cloth.	9% for glass samples, 23% for fabric, and 36% for wood; NO difference between males and females; 22% classified as shedders; secondary transfer inferred by no. of alleles

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
8	Lehmann et al. (2013) Following the transfer of DNA: How far can it go?	4 replicates of six transfers of wet and dry blood and touch DNA on cotton and glass Average % DNA transferred as Goray (2010a)	Measure the detectability of DNA following multiple transfers	Wet blood detected up to 4 transfers on cotton and 6 on glass; dry blood detected up to 2 on cotton and 6 on glass; DNA detected on first transfer on cotton and second on glass
9	Verdon et al. (2013) The influence of substrate on DNA transfer and extraction efficiency	6 fabrics as substrates, three nonporous substrates; wet and dry blood; passive and friction; 4 replicates % transfer DNA	Influence of nine substrate types on DNA transfer involving blood	High transfer when primary substrate nonporous and secondary porous; extraction most efficient from nonporous; friction and wet give best transfer
10	Poetsch et al. (2013) Influence of an individual's age on the amount and interpretability of DNA left on touched items	213 individuals at different stages of life Total DNA amount and allele counts	Effect of age on transfer	Amount of DNA of children and older participants could be distinguished
11	van Oorschot et al. (2014a) DNA transfer: The role of temperature and drying time	4 replicates of four temperatures in 13 time conditions % DNA transfer flaking blood from nonporous surfaces may affect yields	Time to dry biological fluids and effect on transfer	Exponential decay rates regardless of temperature; blood dries fairly quickly; transfer of DNA very dependent on dryness of sample, so timing since deposit needs to be considered
12	van Oorschot et al. (2014b) Persistence of DNA deposited by the original user on objects after subsequent use by a second person	54 pens and 88 nylon/polyester elastic bands "used" by one donor and given to second users; 46 items solely used by one individual given to second user Relative % contribution of each participant using relative RFU contributions at each locus; where alleles were shared, RFU portion determined using RFU of other alleles at that locus	Check the persistence of DNA following prior use by an individual	% contribution of first user decreases in a linear manner with time; depends on substrate; hard porous surface loses first person's DNA quicker than soft porous item; unknown source alleles detected
13	Gršković et al. (2014) Impact of donor age, gender, and handling time on the DNA concentration left on different surfaces	60 participants touched 9 items; 540 samples Amounts only; no profiling carried out	Test correlation between donor age, gender, and handling time and trace DNA amount recovered on paper, plastic, and plastic-coated metal surfaces	Item texture, donor age, and gender influence trace DNA concentration; independent of handling time

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
14	Davies et al. (2015) Assessing primary, secondary, and tertiary DNA transfer using the Promega ESI-17 Fast PCR chemistry	Couples gripping plastic tubes, directly or following handshakes for 30 s % of unambiguous alleles actually detected compared to those available for detection; summing of the peak heights of all of the detected unambiguous alleles matching the expected donor, divided by the number of alleles expected	Measure the levels of DNA transfer from direct, secondary, and tertiary transfer	Variable nature of primary transfer; occasional secondary transfer greater than primary; even in primary transfer, nondonor alleles were detected; suggestion that there was a limit for template?
15	Lehmann et al. (2015) Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA?	DNA, wet and dried blood used as substrate on glass and cotton; one donor as background on first set of six and different donors as background on second set of six; target DNA added to first substrate before transfers; replicated by 4 % DNA as noted by unique alleles	Transfer and detection in the presence of background	Presence of background DNA influenced the transfer of DNA differently depending on the combination of biological material and surface type; detection decreased after multiple contacts due to decreased DNA and complexity of mixtures
16	Fonneløp et al. (2015a) Secondary and subsequent DNA transfer during criminal investigation	3 donors deemed to be good shedders; 30 transfer chains; 11 repeats for wood, 9 for plastic, and 10 for metal Quantity of total DNA in ng and % DNA transferred as assessed by number of alleles above 200 RFUs	Primary transfer to wood, plastic, and metal and secondary transfer via nitrile gloves; onto fabric and paper	DNA can be transferred onto a third substrate via nitrile gloves in 5 out of 30 transfer chains
17	Fonneløp et al. (2015b) Persistence and secondary transfer of DNA from previous users of equipment	4 participants: 2 male, 2 female Alleles present to include; person could not be excluded or contributors cannot be detected	Study of persistence of DNA from previous user to new user's hand	Initial user alleles detectable up to 8 days after receiving the equipment
18	Goray & van Oorschot (2015) The complexities of DNA transfer during a social setting	Three participants repeated five times STRmix, to record exclusion, not excluded, and no. persons in the mixture	Study transfers with group having a drink together	DNA can be detected without actual contact between individuals; DNA of unknown source can be transferred from hands

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
19	Oldoni et al. (2015) Exploring the relative DNA contribution of first and second object's users on mock touch DNA mixtures	Fourteen volunteers acting as first or second handlers of 5 plastic, 2 metal, 1 fabric and inside and outside of nitrile gloves giving 231 mixtures Relative peak height (50 RFU) of the two contributors with markers showing no allele sharing; average profile contribution was calculated over several samples paired	To gain knowledge on the relative contribution of DNA left behind by different users over time	Second handler contribution increased from 21% to 73% between 5 and 120 min; unexpected full profiles detected in 7 simulations suggesting indirect transfer
20	Meakin et al. (2015) The deposition and persistence of indirectly transferred DNA on regularly used knives	4 volunteers paired, experiment in triplicate repeated for 5 weeks at 1 hour, 1 day, and 1 week time intervals % profiles on the basis of unique alleles, RFU 100; total amount of DNA	Whether transferred DNA could be detected on regularly used items	DNA of person who shook hands with knife handler; regular user could be detected in 10:1 ratio, but alleles were detected for up to 1 week; unexpected alleles also detected, suggesting indirect transfer
21	Montpetit & O'Donnell (2015) An optimized procedure for obtaining DNA from fired and unfired ammunition	Ten volunteers carried half their ammunition for 2 days before loading weapons, and the other half was loaded directly; each shooter loaded half of their cartridges into a magazine and tested unfired cartridges; other half were fired and analyzed Quantities of DNA and reportable alleles recorded and interpretable profiles as judged by fixed criteria	Study to optimize collection and profiling of DNA from fired and unfired ammunition	Less than 50 pg on 78% (607 of 800), 27% (229 of 785); 40% had mixtures or indication that more than loader's genotype detected; available information is human handling at manufacture stage less than 1%
22	Oldoni et al. (2016) Shedding light on the relative DNA contribution of two persons handling the same object	Fourteen persons acting in pairs as first and second user handled a range of everyday items in three time simulations Alleles over 50 RFU counted so long as they appeared in 2 amps; % contribution calculated.	To understand the relative proportion of DNA deposited by different persons through time	Contribution from second user increased in time and became the major profile in many instances after 120 min; indirectly transferred DNA in 8/234 cases; a full profile in one case; evidence of shedder status; porous and nonporous effects

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
23	Samie et al. (2016) Stabbing simulations and DNA transfer	4 donors, 16 experiments, and 64 traces 30 RFU; allelic count and STRmix;70% more than 6 loci considered full profile	Study transfer of DNA from handler and check if handlers would transfer DNA from persons closely connected to them	DNA of person handling the knife in 83% of cases; person nearby not detected; 2, 3 and 4 person mixtures
24	Cale et al. (2016) Could secondary DNA transfer falsely place someone at the scene of the crime?	12 participants using 24 knives Quantity of DNA and allelic counts 50 RFU	Detection of interpretable secondary DNA profiles	After 2 min handshake, secondary DNA transfer was detected in 85% of the samples; in five samples, secondary contributor was major or only contributor
25	Goray et al. (2016) Shedder status—An analysis of self- and non-self-DNA in multiple handprints deposited by the same individuals over time	240 handprints from 10 individuals; self and nonself DNA determined Deposits varied 0.05 to 5 ng; total DNA; total alleles per locus; STRmix using depositor and staff elimination database; evaluation of mixture proportions	Determine if individuals deposit consistent quantities of their own DNA as well as variability	Some individuals shed more readily than others, but there is a lot of variation; nonself, usually as minor component in 79% of samples; depositor excluded from deposit in 7 samples; good shedders had less nonself DNA; total amount of DNA independent of ratio of self to nonself
26	Buckingham et al. (2016) The origin of unknown source DNA from touched objects	4 participants; seven tests % unique alleles and unique alleles of other participants; total adjusted peak height used to get % contribution DNA	Test whether the last person to handle an item can be detected in the DNA profile produced from that item	Nonself DNA common on a person's hands; material deposited and retrieved from an object is dependent on who touches what, how, and when; evidence of the prevalence and complexity of nonself DNA in its deposit and transfer
27	Helmus et al. (2016) DNA transfer—a never-ending story; a study on scenarios involving a second person as carrier	3 pairs, each participant acted as donor, giving 6 implementations per scenario of participants repeated twice Allele counting at each locus >50 RFU; classified as complete if each allele present without additional peaks or if 5 or more regardless of additional deemed to be partial; <5 alleles regarded as no profile	Study of second person as a carrier	DNA transfers from donor to cotton to plastic or cotton via second person 40% of 180 samples; cotton much more receptive than plastic; effect of gloves not as strong as expected

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
28	Manoli et al. (2016) Sex-specific age association with primary DNA transfer	128 individuals, experiment in triplicate, 768 swabs % alleles	Effect of age and sex on transfer of DNA; also test if shedder status remained constant in 1 and 2 transfers	DNA not always transferred; claim primary and secondary can be distinguished; 77% of participants changed shedder status through the trials; young males more likely to transfer than older males; actual results suggest females poorer shedders but not claimed by authors
29	Lacerenza et al. (2016) A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers (PHF)	Samples collected from 30 males and 30 females Peak height, 50 RFU; 16 tissue markers for mucosa; saliva; semen; vaginal mucosa; menstrual secretions; and skin	Study to explore source of transferred DNA using DNA/RNA; levels of foreign material on hand surfaces of the general population	Nonskin cellular material observed in 15% of PHF; amount of DNA from these samples higher than skin cells only; donor alleles 75% in males and 60% in females; 30% females had mixtures with a component of 20% or more and 8% males had such mixtures
30	van den Berge et al. (2016) Prevalence of human cell material: DNA and RNA profiling of public and private objects and after-activity scenarios	549 samples, four categories: public (105); private; transfer-related; and washing machine samples RNA and DNA co-extracted; in-house multiplex used for RNA; known genotypes used with in-house software to assess contribution to mixtures; maximum allelic counts used to determine the minimum number of contributors	Gain understanding of cell material on surfaces contributing to background traces; DNA mRNA on various items	High DNA not related to increased number of contributors; major DNA on an individual may not be owner; in activity situations, perpetrator not always the major
31	Voskoboinik et al. (2017) Laundry in a washing machine as a mediator of secondary and tertiary DNA transfer	Eight new unworn socks - various cotton blends washed with typical laundry of four households - various washing conditions; six new unworn socks and a T-shirt laundered without additional items; 15 washing machine drums swabbed Amount of DNA and allele calls; 60 RFU detection threshold, 200 RFU stochastic threshold	Check the possibility of secondary and tertiary DNA transfer during laundry washing of worn and unworn garments in household and public washing machines	Secondary transfer detected in 22% of cases; tertiary transfer experiments indicated that the possibility of DNA transfer between separate washing cycles via the deposition of biological material in a washing or drying machine's drum is unlikely

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
32	Fonneløp et al. (2017) The implications of shedder status and background DNA on direct and secondary transfer in an attack scenario	20 participants, 60 experiments with test tubes; 17 simulated attacks with four samples from each Quantity of DNA; mixture interpretation according to International Society for Forensic Genetics (ISFG) guidelines; three-person mixture considered if major profile	Shedder status and effect of background DNA; simulated attacks	No aerosol transfer from talking; DNA transferred in attacks (16/17); background DNA from the environment can be confused with crime samples (1/148)
33	Szkuta et al. (2017a) Transfer and persistence of DNA on the hands and the influence of activities performed	Volunteers paired on 12 occasions; each of 24 participants acted as depositor or known contributor LR for POI using STRmix; up to 4 participants analyzed with caution LR of 100 billion reported rather than exact number	Whether nonself DNA transferred via handshake could be detected on surfaces and what effect activities had	Depositor of handprint main depositor; minor contributions from handshaker decreasing with the number of handshakes; main depositor excluded on several occasions; concept of "parking," i.e., retransfer of DNA on used items
34	Meakin et al. (2017) Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly and indirectly transferred DNA on regularly used knives	4 volunteers carrying out experiments on three separate weeks at 1 hour, 1 day; and 1-week intervals; 36 knives for examination in total Total DNA amount; peak heights and % unique alleles as well as RMP and LR using LRmix 2.0	To study directly and indirectly transferred DNA on regularly used knives; extension of 2015 study	When dealing with items already having a DNA load, it may be possible to use intrinsic qualities of profiles to distinguish between directly and indirectly transferred DNA
35	Ruan et al. (2018) Investigation of DNA transfer onto clothing during regular daily activities	50 participants supplied shirts, various areas sampled worn for 7–9 h and sampled again; 38 participants received 10 × 10 swatches to add to their laundry STRmix used to examine profiles produced Y-allele at the amelogenin locus in PowerPlex® 21 System	Check the transfer of DNA to clothing during regular activity; test the effect of laundering	The adventitious transfer of trace DNA means that the DNA recovered in forensic casework may not always have evidentiary relevance; freshly laundered clothes had interpretable mixtures from which uploadable foreign DNA profiles could be determined; in some cases, the donor of the clothing was not even the predominant DNA profile in the sample

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
36	Pfeifer & Wiegand (2017) Persistence of touch DNA on burglary-related tools	Three types of tools with and without gloves; 234 samples in total Completeness of profiles based on unique alleles; casework approach to reporting for German database entries, 4/8 of some markers together with 7/13 of another set; statistical comparisons conducted using GraphPad Prism	Explore the persistence of DNA on tool handles when more than one person touched them; different types of tools tested with and without gloves; experiments carried out to get data to address activity propositions in case of mixed profile on a screwdriver	Owner detected in 47% of cases before burglary and in 1/30 cases after mock burglary and never as major; more moderate action gives possible match to first or second user; 30% tools from households have reportable profile of owner; 57% have mixture that cannot be resolved; amounts varied in manner that did not help; one case of second user even though wearing gloves; nature of contact, substrate, and user characteristics variables
37	Bowman et al. (2018) Detection of offender DNA following skin-to-skin contact with a victim	Nine pairs tested three times; some changes resulting in total of 15 females and seven males; 266 samples collected; 72, 94, and 100 from time points 0 h, 3 h, and 24 h; skin and clothing sampled Unique alleles recorded and STRmix used for mixture interpretation	Test value of collecting DNA samples in mock assault situations	Support for H_p for 56% and 77% for medium and heavy pressure used in assault; amount of DNA falls off rapidly on skin but detectable on clothes up to 24 h; high amount of nonself alleles detected in control areas; information on shedder varying with time
38	Poetsch et al. (2018) Impact of several wearers on the persistence of DNA on clothes	4 females and 2 males wearing sweatbands for times from 10 min to days; each combination of times done with 6 different pair/trios of individuals, giving a total of 204 samples Amount of DNA and allele peaks interpreted when greater than or equal to 300 RFUs for single; allele counting at each locus >50 RFU; classified as complete if each allele present without additional peaks or if 5 or more regardless of additional deemed to be partial; <5 alleles regarded as no profile	Test how long DNA persists on an item used in daily routine and how long a piece of clothing must be worn to definitively leave detectable DNA behind	After 10 min, at least a partial profile of the second/third wearer of a piece of clothing could be demonstrated; even after the sweatband was worn for 3 days by the second wearer, the complete profile of the first wearer was still detectable in 42% of these samples

No.	Reference and Title	Size of Study and Measurement Criteria	Purpose of Study	Key Results
39	Helmus et al. (2018) Persistence of DNA on clothes after exposure to water for different time periods—a study on bathtub, pond, and river	Five participants; epithelial cells and separately a drop of blood added to clothes left in bathtub, pond, and small river for varying periods up to six months Allele peaks >50 RFU; complete profile if all alleles detected even if additional peaks; partial if more than half loci; and regarded as no profile if less than half of the evaluable loci in every allele of the individual in question was found	This study was conducted to attempt a general statement about the conditions under which sufficient DNA remains can be expected for molecular genetic analysis	Complete STR profiles could be detected even after immersion in water, dependent on conditions; longest time recorded was full profile after 2 weeks in a pond in winter
40	Tobias et al. (2017) The effect of pressure on DNA deposition by touch	2 participants, 36 samples Quantity of DNA and % profiles based on alleles	Test whether pressure affects the amount and the quality of DNA transferred by touch	Increase in pressure resulted in an increase in DNA from both donor and unknown sources; no difference between participants at 4 kPa but variation noted at 21 kPa and 37 kPa

5.3.2.1. Shedder Status

Shedder status refers to the greater or lesser tendency of an individual to shed DNA (Lowe et al. 2002). This is an important variable affecting transfer. There is a consensus that some people are better shedders than others, but there is less agreement about whether individual variation over time is of comparable magnitude. Different studies use different criteria to classify participants as good or bad shedders. Therefore, even though there is agreement that people vary, there is no universal scheme for classification.

The first article describing touch DNA results (van Oorschot & Jones 1997) noted variable amounts of DNA recovered from objects touched by different individuals, though these individuals were not formally classified as *good shedders* or *poor shedders* at that time. One of the first studies to examine shedder status specifically concluded:

> "While a good DNA shedder may leave behind a full DNA profile immediately after hand washing, poor DNA shedders may only do so when their hands have not been washed for a period of 6 hours" (Lowe et al. 2002).

Some studies have raised doubts about the ability to classify individuals as good or bad shedders (Phipps & Petricevic 2007), while other studies have confirmed that these categories can be useful (Farmen et al. 2008, Goray et al. 2016, Kanokwongnuwut et al. 2018).

Many recent studies have reported that,

> "while there is substantial variation in the quantities deposited by individuals on different occasions, some clear trends were evident with some individuals consistently depositing significantly more or less DNA than others" (Goray et al. 2016).

Another study carried out with 128 individuals found that shedder status varied with individuals over time in 77% of cases (Manoli et al. 2016). When age was studied, children and older participants could be distinguished (Poetsch et al. 2013, Grškovic et al. 2014). Younger males were more likely to shed than older males, though this effect was not noted in females (Manoli et al. 2016).

One study found that the amount of DNA transferred was not correlated with how long an item was handled (Gršković et al. 2014). Other studies investigating activities found time between activities had an impact. For example, when the deposition of a handprint was delayed, the activities performed by the individual had a substantial effect on the resultant detection of the contributing profile. In addition, multiple contacts with the same items increased the likelihood that the known contributor's DNA would be retained and subsequently detected due to the parking and retransfer of DNA on used items (Szkuta et al. 2017b). Moisture was found to increase the amount of transfer (Goray et al. 2010a, Lehmann et al. 2013, Verdon et al. 2013).

Some studies examining secondary transfer have found that contributions from particular donors dominate, with this finding being explained by a shedder effect (Fonneløp et al. 2017, Buckingham et al. 2016, van Oorschot et al. 2014a). Other studies exploring the contribution of two and more people to the surface of an object proposed shedder status as a major factor explaining the variability in percentage contributions (Oldoni et al. 2015, Oldoni et al. 2016, Meakin et al. 2015, Goray et al. 2016).

It is more useful to think of shedder status as existing on a continuum—as opposed to there being good or bad shedders—as these studies do support the idea that some individuals routinely shed more DNA than others. The most convincing example in the literature thus far is a longitudinal study of contamination in an operational biology laboratory over a period of time (Taylor et al. 2016d). In this study, DNA linked to one individual was greater in quantity and more widely distributed than DNA from a coworker with similar duties working nearby (Taylor et al. 2016d).

It may be that a definite answer to the question of shedder status will not be possible until we gain a better understanding of the mechanisms of DNA transfer, as discussed earlier. However, the degree to which an individual sheds DNA is a variable that needs to be kept in mind when considering the relevance of DNA in a mixture or in any situation where there is the question of how or when the DNA was deposited.

5.3.2.2. Substrate Effects

The material onto which DNA transfers (i.e., the substrate) has an effect on how easily DNA will transfer or be retained. Researchers have examined the effect of moisture and substrate

for transfer of skin cells and noted that skin cells are deposited more readily onto porous substrates, such as cotton. However, secondary and higher-order transfers of skin cells are facilitated more by non-porous substrates, such as plastic. The most effective transfer chain was from non-porous to porous substrates with the use of friction (Goray et al. 2010b).

A study of 300 participants holding glass, cloth, and wood found the likelihood of obtaining a DNA profile was approximately 9% for glass samples (average recovery of ≈0.50 ng or ≈85 cells), 23% for fabric (average recovery of ≈1.2 ng or ≈200 cells), and 36% for wood (average recovery of ≈5.8 ng or ≈975 cells) (Daly et al. 2012). If this particular study, which was conducted with the STR kit SGM Plus using 28 cycles, was repeated with the higher-sensitivity DNA tests being routinely used today (e.g., the STR kit GlobalFiler with 29 cycles or PowerPlex Fusion with 30 cycles), then the DNA amounts detected via transfer would be expected to increase, while the relative suitability of surface types would probably remain the same.

Another study involving nine different substrates also found that the amount of DNA transfer was highly dependent on the porous or non-porous nature of a surface (Verdon et al. 2013). The finding that transfer was highest when the primary substrate was non-porous and the secondary substrate was porous is in keeping with our everyday experience of how materials behave. When transfers onto wood, plastic, and metal were considered in another study, nitrile gloves were found to be good vectors for additional transfers onto fabric and paper (Fonneløp et al. 2015a). More DNA transferred onto the wood and plastic than onto the metal initially, but proportionally more was transferred from the metal onto the gloves. DNA was transferred in highest concentration to plastic and plastic-coated metal, and least onto paper in a different study (Gršković et al. 2014).

Substrate effects were again noted when controlled experiments were carried out to check the persistence of DNA from a prior handler following handling by a second person:

> "The retrieval of the profile of the initial user of the object is dependent on the type of substrate and on how the object was used. When considering a hard, non-porous object, the first user's contribution to the profile drops approximately 50% immediately upon use by a second person and drops to approximately 15% after 90 minutes. When considering a worn object made of soft porous material, the first wearer's profile remained higher than that of a second wearer during the first 10 hours of wear by the second wearer, and still accounted for approximately 12% after 96 hours" (van Oorschot et al. 2014a).

Other researchers, when exploring the impact of a second user following a first user or habitual user, studied a range of materials, and though they reported that the second user became the major DNA contributor for all substrates after 120 min, they did note "extreme values" for both non-porous plastic bracelets and porous nurse caps (Oldoni et al. 2015). Items of clothing, i.e., porous material, were used in two other studies that broadly sought to test whether wearer DNA could be identified (Breathnach et al. 2016, Magee et al. 2018).

A recent review explored the underlying mechanisms of metal-DNA interactions. It acknowledges how ionization and electron affinity of metals impact the degree of interaction with DNA as a negatively charged molecule. The proposal is that this bonding is responsible for the difficulty in recovering DNA from certain metal surfaces and it shows that understanding these metal-DNA interactions are fundamental to improving the chances of getting interpretable profiles from trace samples (Bonsu et al. 2020).

5.3.2.3. Persistence Studies

For a DNA association to be relevant in a particular case, the DNA must have been deposited at the time the crime occurred (see Figure 5.1). If any cells or DNA molecules were left prior to the crime and persist, then this *non relevant DNA* could contribute to the crime scene evidence (e.g., possibly creating a mixture) and potentially influence the relevance of the final result. Therefore, it is important to understand the factors that affect the persistence of DNA.

DNA persistence has rarely been studied in isolation. One study, using the Profiler Plus kit with 28 cycles, detected DNA out of doors that had been deposited up to two weeks before (Raymond et al. 2009a). The sensitivity of the technology has increased since that time, so it is possible that today, profiles would be detectable for a longer period of time. However, similar studies have not yet been undertaken with newer STR kits and CE instruments. In a study considering the persistence of primary and secondary transfer from previous users of equipment, alleles of the previous user were detected for up to eight days (Fonneløp et al. 2015b). In a study of buildup of DNA contamination from staff members in a semi controlled laboratory environment, DNA profiles were detected long after deposition, and in fact could be detected months later, rather than merely days or weeks (Taylor et al. 2016d).

A study on the detection of offender DNA following a simulated assault involving skin-to-skin contact showed a rapid decrease in detection of the offender's DNA on the skin, though DNA profiles could still be detected up to 24 hours post assault (Bowman et al. 2018). DNA could also be detected on clothing worn over the assault area up to 24 hours later, and the authors suggested that sampling from clothing worn over the assaulted area may be an additional or better avenue for the recovery of offender DNA post assault, when there has been a significant time between assault and sampling (Bowman et al. 2018).

As will be discussed in a later section on digital penetration, there have been a number of persistence studies dealing with fingernails.

Information on persistence can also be gained from studies on the effect of a second user when the persistence of the first user is studied. The DNA of the initial user decreases with time, though in a study involving knives used by a person following a handshake, DNA from the handshaker was detectable on a knife handle for at least a week albeit as a partial profile (Meakin et al. 2015).

5.3.2.4. Non-Self-DNA on Individuals

Many of the studies summarized in Table 5.2 detected alleles or profiles that could not be accounted for by DNA from the individuals participating in the study. For example, foreign alleles were detected approximately 50% of the time, with 31% consisting of one to three alleles and 9% containing six or more (Manoli et al. 2016).

Such alleles from unknown sources have received more emphasis in recent studies because of increases in DNA test sensitivity. The authors of one study, which sought to look at DNA transfers in a social setting rather than in structured experiments, reported that,

> "simple minor everyday interactions involving only a few items in some instances lead to detectable DNA being transferred among individuals and objects without them having contacted each other through secondary and further transfer. Transfer was also observed to be bi-directional. Furthermore, DNA of unknown source on hands or objects can be transferred and interfere with the interpretation of profiles generated from targeted touched surfaces" (Goray et al. 2015).

In another study, non-self-DNA was detected on 79% of hands (Goray et al. 2016). Results from this study showed that in most situations, participants were majority contributors or the only source of the DNA deposited. An average of 74% of detected DNA derived from self, while the other 26% appeared to be non-self-DNA. In instances involving participants that the researchers classified as *poor shedders*, non-self-DNA rather than self-DNA was transferred. This was found to be the case in seven samples, 2.9% of the time (Goray et al. 2016).

A study about a new collection and extraction procedure for obtaining DNA from ammunition also provided an example of detection on non-self-DNA (Montpetit & O'Donnell 2015). In this study, 10 volunteers handled various fired or unfired rounds of ammunition, which were then swabbed for DNA. With 97% of interpretable results, the volunteer that handled or loaded the ammunition was detected. However, non-self-DNA was detected unexpectedly: the DNA profile from a child of one of the volunteers was recovered from ammunition where there was no opportunity for the child to touch the ammunition directly (Montpetit & O'Donnell 2015).

In a number of studies, the major profile was not always associated with the last person to handle an item (Cale et al. 2016, Buckingham et al. 2016, Goray et al. 2016). This may result from background DNA or from the handler depositing non-self-DNA.

5.3.3. Studies on DNA Transfer that Mimic Casework Scenarios

5.3.3.1. Caution with Using DNA in Domestic Settings

Given that DNA transfers readily, investigating crimes in domestic settings can be challenging. Numerous researchers have conducted experiments on transfer during clothes washing/laundering. This is important because moisture was noted as one of the factors affecting secondary transfer of biological materials and DNA (Goray et al. 2010a, Goray et al. 2010b). The potential for transfer of spermatozoa in washing machines has been accepted by forensic biologists for some time (Kafarowski et al. 1996). More recent studies have also

found transfer of DNA rather than spermatozoa during washing (Brayley-Morris et al. 2015, Noël et al. 2016). Together, these washing studies suggest that finding DNA from one member of a household on another needs to be interpreted with caution. DNA from family members was detected on children's underwear even in instances where semen was not placed on the samples (Noël et al. 2016). In another study, DNA from blood of a household member was detected on laundered items, but DNA from saliva or epithelial abrasions was not detected (Kamphausen et al. 2015). A 2018 study reported that it is not uncommon for foreign DNA to transfer onto an individual's clothing during laundering and included a note of caution in relation to the investigation of crime in domestic situations (Ruan et al. 2018).

5.3.3.2. Mixtures in Sexual Assault Cases

In the early days of DNA profiling, most mixtures were from sexual assault cases where epithelial cells from the female victim were mixed with sperm and epithelial cells of the perpetrator. Although such samples can involve allele overlap and other complicating factors, sperm and epithelial cells are relatively easy to separate because sperm cells are more resistant to extraction, which allows the DNA from the two types of cells to be extracted without mixing. It is important to note that sexual assault samples may contain epithelial cells from the perpetrator (from seminal fluid, skin contact, saliva) which will be co-extracted with female epithelial cells; however, male epithelial cells are typically in the minority on swabs taken from the female victim and may not result in detectable alleles. Differential extraction (Gill et al. 1985) continues to be an important method in these types of cases.

5.3.3.3. Sexual Intercourse versus Social Contact

There are various other situations in sexual assaults where mixtures of unknown cell types are encountered. Researchers have tended to design specific experiments to address these issues, as seen below. Although the sample numbers in the experiments are limited, they do provide better information than uncalibrated experience in the absence of ground truth.

In some cases in which DNA is recovered, the trier of fact needs to assess whether the DNA transfer occurred during a sexual assault or during simple social contact. A series of experiments measured the amount of female DNA transferred to male undergarments and genitals following sexual intercourse and following non-intimate social contact that was designed to maximize transfer (Jones et al. 2016). In the experiments performed, it was not possible to replicate the high levels of DNA transferred from sexual intercourse by non-intimate contact (Jones et al. 2016). Although this study was confined to one couple carrying out the sexual intercourse experiments, the findings are in keeping with the effects of moisture on transfer seen in earlier transfer experiments (Lehmann et al. 2015).

A retrospective survey of sexual assault cases noted *positive findings* consisting of epithelial cells recovered from the penis highlighting the advantage of collecting such samples in sexual assault cases (Fonneløp et al. 2019). When such samples are examined and a female victim claims vaginal penetration, the defendant may offer an alternative explanation of secondary transfer of victim's cells to his penis. Fourteen couples were recruited to test the

hypotheses that female DNA was more likely to be detected following intercourse than social contact. The authors report the possibility of using their data to make a statistical model to distinguish "between samples taken after intercourse and samples taken after secondary transfer by skin contact" (Bouzga et al. 2020).

5.3.3.4. Digital Penetration

Recent studies of digital penetration used information from Y-STR markers on vaginal swabs (McDonald et al. 2015). Conversely, earlier work focused on the possibility of getting DNA matching the female from under the fingernails.

> "Full female profiles were obtained from all swabs collected at 0 and 6 hours after digital penetration, indicating that female DNA was always transferred and persisted in the short term. Furthermore, full female profiles were produced from three-quarters of samples collected after 12 hours whilst mixed profiles were produced in the majority of samples taken after 18 hours. The analysis of several variables indicated that hand washing had a significant effect on the persistence of female DNA profiles" (Flanagan & McAlister 2011).

An earlier study of fingernails at autopsy stage did not record foreign profiles in the majority of cases (Cerri et al. 2009).

In a study involving

> "deliberate scratching of another individual (n = 30), 33% of individuals had a foreign DNA profile beneath their fingernails from which the person they scratched could not be excluded as a source; however, when sampling occurred 6 hours after the scratching event, only 7% retained the foreign DNA" (Matte et al. 2012).

In controlled experiments with females scratching males to simulate assaults, 95% (38 out of 40) of fingernail samples collected immediately and 60% (24 out of 40) of those collected five hours later were "suitable for comparison" (Iuvaro et al. 2018). Analyses of fingernail samples in criminal cases were also studied (Bozzo et al. 2015).

Clothing is also submitted in cases of alleged digital penetration. In an experiment designed to better target sampling, a mannequin was used to determine how much DNA was transferred by volunteers to parts of underwear (Ramos et al. 2020).

5.3.3.5. Wearer versus Toucher

In the past, it may have been common to use the DNA profile obtained on a garment as a proxy for the DNA profile of the person who wore the garment (e.g., Casey et al. 2016). However, the issue of increased sensitivity is again relevant. A recent study showed that the wearer profile was detected in all interpretable profiles, and it was the major profile 50% of the time (Magee et al. 2018). However, the definition of *interpretable* varies across laboratories (e.g., Benschop et al. 2017a). Therefore, information obtained from many of

these DNA transfer studies will only be valuable in a particular case when carried out under similar conditions and interpretation criteria.

An inter-laboratory study considered upper garments following being worn by individuals who embraced (contact), went on an outing together (close proximity) or spent a day in another person's environment (physical absence). The wearer was typically but not always, observed as the major contributor to the profiles obtained. The authors of the study noted: "DNA from the activity partner was observed on several areas of the garment following the embrace and after temporarily occupying another person's space. No DNA from the activity partner was acquired by the garments during the outing even though both participants were in close proximity" (Szkuta et al. 2020).

5.3.4. Studies on Contamination

Contamination is a type of DNA transfer. However, it is typically considered as a special case of transfer and is investigated separately from the types of DNA transfer studies discussed above. Many studies focus on contamination and on suitable methods to avoid it. A list of such studies is presented in Table 5.3.

Table 5.3. Studies where measuring or investigating potential sources of contamination is the main focus.

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
1	Rutty et al. (2003) The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime	Series of experiments were undertaken to determine the extent to which an investigator could contribute to any DNA contamination of a scene of crime under different simulated activities; effectiveness of protective clothing checked	18 experiments with one participant	In total, 413 alleles were identified in the 18 experiments, and 34 were not attributable to the subject and therefore considered to be contamination; vigorous activity, even when wearing protective garments, can cause contamination of a crime scene	Need for ongoing checks on the effectiveness of protective clothing
2	van Oorschot et al. (2005) Beware of the possibility of fingerprinting techniques transferring DNA	Check the potential of fingerprint brushes to transfer DNA	13 brushes used to powder surface containing saliva before powdering clean plates; DNA contaminated brushes used to powder 6 plastic sheets in another experiment	Transfer occurred when brushed over a biologically stained area or fresh print	Need to ensure fingerprint brushes are not transferring DNA

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
3	Durdle et al. (2009) The transfer of human DNA by *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae)	*Lucilia cuprina* were fed either human blood or human semen *ad libitum* and their artifacts were analyzed for human DNA content	Samples containing 1, 10, 30, and 50 artifacts	Blowfly artifacts can be a source of DNA at crime scenes, in addition to being a potential contaminant; data suggest the amount of DNA in artifacts can be dependent on the meal type	Depending on the environment, be conscious of ability of living things, other than humans, to transfer DNA
4	Preuße-Prange et al. (2009) The problem of DNA contamination in forensic case work—How to get rid of unwanted DNA?	Tested the efficiency of different (chemical and physical) procedures for DNA removal with focus on the commonly recommended ultraviolet (UV) irradiation	Saliva and pure DNA applied to glass slides for 9 time periods from 5 min to 24 h and exposed to UV sources at 8 and 48 cm	Pure DNA reduced more effectively than saliva	UV irradiation can only reduce the contamination but does not eliminate it completely;- importance of contamination avoidance prior to analysis
5	Daniel & van Oorschot (2011) An investigation of the presence of DNA on unused laboratory gloves	A preliminary investigation of three brands of laboratory gloves was undertaken to determine the levels of human DNA present on unused gloves from closed and open boxes	In total, 56 gloves were examined from six to seven closed boxes of three different brands	5 gloves from four of seven boxes of one brand had up to 20 alleles	Use certified DNA-free gloves
6	Digréus et al. (2011) Contamination monitoring in the forensic DNA laboratory and a simple graphical model for unbiased EPG classification	Devising a classification scheme for monitoring contamination events	25 EPGs compared with classification made by two reporting officers	Scheme operational	Potential for monitoring across laboratories
7	Durdle et al. (2011) The change in human DNA content over time in the artefacts of the blowfly *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae)	Check whether human DNA that can be profiled from blowfly changes with time	41, 43, and 22 samples tested for blood, semen, and saliva fed to blowflies	Blood and semen data showed that the amount of human DNA that could be extracted increased over the first 400 days but had decreased to one-month levels by 750 days; no changes in saliva over 60 days in the amount of human DNA that could be extracted	Issue for cases held in storage

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
8	Goray et al. (2012b) DNA transfer within forensic exhibit packaging: Potential for DNA loss and relocation	Investigation of how much DNA is "lost" from an exhibit due to its transfer to the inside of the packaging containing the exhibit, and transfer from one area of an exhibit to another	Multiple variables of substrate and packaging	Demonstrated that DNA could be transferred from the deposit area to either other parts of the item or to the bag itself and usually to both	% total DNA and number of alleles
9	Szkuta et al. (2013) The potential transfer of trace DNA via high-risk vectors during exhibit examination	Check level of DNA potentially transferred between high-risk vectors (scissors, forceps, gloves) and exhibits during the examination process in both light and heavy contamination/contact scenarios	24 swatches stained with 25 µL of blood used as source for multiple uses by three vectors	DNA transfer was observed for all vectors in both heavy- and light-contact scenarios; sufficient alleles to identify the origin except in case of forceps, where only a small number of alleles were transferred under light conditions	Tools and equipment should be cleaned or replaced immediately if they come into contact with substrate containing blood
10	Neuhuber et al. (2009) Female criminals—It's not always the offender!	Systematic search for errors in the investigative process following the contamination of multiple cases in 1993 and 2009 by female DNA	In 34 out of 191 swabs, peaks were found at 4 or more loci of the SGM+-kit; these 34 swabs corresponded to 2 manufacturers	Noted that cotton swabs that had been sterilized with radiation were often contaminated	Manufacturing process, as well as the products themselves used in collection of DNA trace evidence, should be reevaluated with the emphasis on preventing contamination
11	Henry et al. (2015) A survey of environmental DNA in South Australia Police facilities	Survey of police areas where items are sometimes examined prior to submission to laboratories, 18 facilities across South Australia	20 various items sampled; number of times sampled varied from 1 to 29	50% had DNA, 4% originated from 1 person, 9% from 2 people, 19% from 3 people, and 18% from 4 or more people; 20% weak profile; 30% no profile	Need procedures to reduce environmental DNA in examination rooms
12	Kovács & Pádár (2015) Misinterpretation of sample contamination in a Hungarian case report	Case report of DNA from soft tissue from bone sent to two laboratories for identification with conflicting results, which were due to mix up	One bone sent to two laboratories	Results of a case study	The risk of contamination mu st never be ignored in forensic examination, and the evaluation of minor/major components of a mixed profile can lead to a wrong interpretation

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
13	Margiotta et al. (2015) Risk of DNA transfer by gloves in forensic casework	All the gloves used in one day by four operators were analyzed; For every glove evaluated, the presence of contamination DNA from the operator or from other samples was detected	16 pairs of gloves used by 4 operators; 5 negative controls from used and unused boxes	12.5% no alleles; 10% operator-related alleles; 12.5% alleles referable to the operator and to the test sample; 50% a mixture of alleles of the test sample and unknown subjects; 15% alleles of unknown subjects different from the operator	Operators must change gloves every time after touching items or surfaces, prior to touching the exhibit
14	van Oorschot et al. (2015) Considerations relating to the components of a laboratory DNA contamination minimisation monitoring (DCMM) program	Advice on what an environmental monitoring program should include	Discussion paper rather than experimental study	Discussion paper rather than experimental study	Information available on what needs to be considered for environmental monitoring
15	Szkuta et al. (2015a) DNA transfer by examination tools—a risk for forensic casework?	Check if DNA and blood transferred to DNA-free surfaces via scissors, forceps, and gloves	Twenty sets of vectors, multiple donors, and four replicates per transfer set; transfer sets each contained blood and touch DNA	DNA-containing material can be transferred from exhibit to exhibit by scissors, forceps, and gloves	Encourage awareness amongst staff of the potential sources of contamination within the laboratory and during examination
16	Szkuta et al. (2015b) Residual DNA on examination tools following use	Check the proportion of DNA that remains on the high-risk vectors following contact with the substrate.	Transfer experiment as Szkuta et al. 2015a	While DNA-containing material is picked up by DNA-free vectors and transferred from exhibit to exhibit, sufficient DNA remains on these vectors, which can potentially result in further transfer and contamination through subsequent contact	See Szkuta et al. 2015a

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
17	Fonneløp et al. (2016) Contamination during criminal investigation: Detecting police contamination and secondary DNA transfer from evidence bags	Check level of contamination in police facilities and check scene-of-crime officers' profiles against casework from 2009 to 2015 A pilot study to assess whether DNA from the outside package of an exhibit could be transferred to a DNA sample was also carried out	Areas divided into high-, medium-, and low-risk areas and three gloves checked after checking case-created scenarios	Environmental DNA was detected in various samples from hot spots; furthermore, 16 incidences of previously undetected police-staff contamination were found; in 6 cases, the police officers with a matching DNA profile reported that they had not been involved with the case	Important to ensure that "best-practice" procedures are upgraded, and appropriate training is provided in order to ensure that police are aware of the increased contamination risks; specific recommendations listed below
18	Bolivar et al. (2016) Assessing the risk of secondary transfer via fingerprint brush contamination using enhanced-sensitivity DNA analysis methods	Check whether fingerprint brushes transfer DNA from fingerprint when using traditional profiling and low-template profiling methods	Six samples, six substrate controls, and six brush controls were collected from each of the three sets of latent and contaminant donors for a total of 18 samples, 18 substrate controls, and 18 brush controls	Although LCN improves the recovery of the DNA profile from the latent print evidence, it also increases the chance of detection of extraneous DNA, such as that transferred by fingerprint brush contamination	Improper procedures may lead to false exclusions or false associations between evidence and crime scene; therefore, procedures for examining latent print evidence should be carefully examined, especially when higher-sensitivity DNA analysis methods are utilized
19	Taylor et al. (2016d) Observations of DNA transfer within an operational forensic biology laboratory	Investigation of the extent to which individuals at Forensic Science SA (FSSA) deposit their DNA on objects throughout the floor of the building where DNA examinations take place by examining monitoring and contamination events as well as specific sampling	138 samples were taken from areas across the floor	Evidence that some individuals shed DNA more readily than others over time; last person to handle an item not necessarily detected; primary transfer accounted for 9/14 contamination events	Questions of how and when did the DNA get there more challenging than statistical calculations; more studies needed to avoid more uninformative responses such as is possible

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
20	Neuhuber et al. (2017) Police officer's DNA on crime scene samples—Indirect transfer as a source of contamination and its database-assisted detection in Austria	Systematic investigation of contamination events	Between the years 2000 and 2016, 347 contamination incidents were detected in approximately 46,000 trace samples (0.75%)	The DNA profiles were screened for contamination incidents by combining a manual check with database-assisted profile comparisons using the national Police Elimination Database (PED) as well as the profile comparison tool of the GeneMapperID-X software	The potential and importance of reference databases containing DNA profiles of police officers and examiners for the detection of contaminated crime scene samples is demonstrated
21	Pickrahn et al. (2017) Contamination incidents in the pre-analytical phase of forensic DNA analysis in Austria—Statistics of 17 years	Continuation of work from Neuhuber et al. (2017)	347 contamination incidents in 17 years	The usefulness of reference profile databases that contain DNA profiles of police officers to detect contamination incidents of trace material	With improved detection methods, it also becomes apparent that indirect transfer of biological material is a serious issue
22	Szkuta et al. (2017b) DNA decontamination of fingerprint brushes	Assessment of the contamination risk of reused fingerprint brushes through the transfer of dried saliva and skin deposits from and to glass plates; assessment of ability to eradicate DNA from brushes	7 new and used squirrel and fiberglass fingerprint brushes used in simulated casework scenarios using glass plates with saliva, single and multiple handprints as substrates; repeated 6–12 times on each substrate and 3 deposits on secondary surface following washings	No profiles observed on new fiberglass brushes, but yields of ≤1 ng on squirrel brushes containing alleles to imply 3 to 4 people; detectability dependent on secondary surface and on biological nature of material being transferred; squirrel brushes easy to clean effectively but fiberglass brushes became tangled and matted	A protocol needed to ensure brushes not used as vectors for transfer of DNA within and between crime scenes

No.	Reference and Title	Purpose of Study	Size of Study	Key Results	Implication
23	Basset & Castella (2018) Lessons learned from a study of DNA contamination from police services and forensic laboratories in Switzerland	National inventory of contaminations to better understand their origin and to make recommendations in order to decrease their occurrence	Mean of 11.5 (9.6 to 13.4) contaminations per year per 1000 profiles sent to the Swiss DNA database	86% of these contaminations originated from police officers, whereas only 11% were from genetic laboratories; direct contact between the stain and the contaminant person occurred in only 51% of the laboratory contaminations, whereas this number increased to 91% for police collaborators	Improving sampling practices at the scene could be beneficial to reduce contaminations
24	Helmus et al. (2019) Unintentional effects of cleaning a crime scene—When the sponge becomes an accomplice in DNA transfer	The aim of this study was to investigate whether DNA traces could be distributed by cleaning an object	Blood, saliva, and epithelial cells from 5 individuals; samples deposited onto two surface types and cleaned with wet sponge; 218 samples initially and 384 in a different experimental setup	It is not only possible but rather probable to distribute DNA from one place to another by cleaning the surface of an object as long as the DNA source is blood or saliva. Regarding DNA from epithelial cells, a transfer of enough DNA for a complete profile by wiping is unlikely	Disposable materials best for cleaning surfaces contaminated with biological fluids
25	Goray et al. (2019) DNA transfer: DNA acquired by gloves during casework examinations	The aim of this study was to investigate DNA transfer during actual casework examinations even when wearing gloves	96 gloves from the examination of 11 exhibits carried out by 5 examiners	Gloves used during examination can collect DNA from the exhibits; for instance, during trace sampling, such losses to the gloves can result in the reduction of DNA available, impacting the quality of the evidentiary profile; furthermore, DNA collected on the gloves could be redeposited on other parts of the exhibit	Profiles were interpreted and statistically evaluated using continuous probabilistic software STRmix (version 2.06) This software weights genotype combinations and allows comparison to persons of interest (POI) and the staff elimination database, expressed as likelihood ratios

The differences between DNA profiles produced by low-template and standard STR multiplex analysis were discussed when high-sensitivity approaches were introduced (Gill et al. 2000). At that time, 100 pg, or about 16 cells, was the minimum amount of DNA that would be analyzed. Duplicate analyses were recommended, and, interestingly, it was noted that laboratory contamination in the form of random alleles could not be eliminated.

The appearance of random additional alleles was previously encountered when profiling wildlife samples from bear (Taberlet et al. 1996). In that situation, the authors explained that the alleles must have arisen as an artifact during PCR because the alleles detected had not been encountered in that laboratory before and therefore could not have been the result of contamination.

Many of the studies on transfer and persistence in which ground truth is known note the presence of alleles not associated with subjects of the study. These alleles are generally attributed to contamination. Such contamination could add to the difficulties of mixture deconvolution when dealing with casework.

The studies on contamination in Table 5.3 illustrate the various ways that contamination can occur during the crime scene examination prior to receipt into the laboratory. The studies give information on possible vectors and other risks that could give rise to such false inclusions. The possibility of contamination from an innocent person's profile is discussed, and the value of elimination databases is supported (Pickrahn et al. 2017, Fonneløp et al. 2016). Miscarriages of justice have arisen because of contamination either before the laboratory or in the laboratory (e.g., Gill 2014, Gill 2016, Gill 2019a).

Contamination is often considered in the context of laboratory handling. The early application of low-template DNA outlined precautions needed in the laboratory (Gill 2001). The main concern at the time was that contamination by stray alleles would cause false exclusions. These precautions included the need to carry out PCR amplification in a separate contained laboratory, that personnel wear disposable laboratory coats and face masks, that staff and police elimination databases be used, and that duplicate tests be performed when possible. A study was conducted on the risk of contamination via routine implements such as scissors and forceps (Szkuta et al. 2015a). Results obtained demonstrated not only that DNA transfers from exhibit to exhibit, but also that DNA persisted on the tools, making future transfers possible. This can give rise to the possibility of false inclusions as well as exclusions (Szkuta et al. 2015a).

Three studies examined the possibility of nitrile gloves acting as vectors (Fonneløp et al. 2015a, Szkuta et al. 2015a, Goray et al. 2019). The results illustrate the need for frequent and appropriate changing of gloves to avoid moving DNA from object to object. The finding of sufficient levels of DNA capable of providing STR alleles on unused gloves is an additional cause of concern (Daniel & van Oorschot 2011, Margiotta et al. 2015). A study on DNA acquired by gloves during casework found:

> "In many instances, the case associated person of interest was observed within the profile generated. So too were profiles of the examiner or other staff

> members, predominantly from the first and last gloves used during the examination, which were associated with removing the exhibit from its packaging and repackaging it." (Goray et al. 2019)

Fonneløp et al. 2016 considered the possibility of contamination prior to receipt by a laboratory. Environmental DNA was detected in samples from various hot spots. It was demonstrated that DNA from the outside of bags could contaminate an exhibit during examination (Fonneløp et al. 2016). Fingerprint brushes also were the subject of a study as potential vectors for transfer of DNA. The additional concern in the case of brushes was that some new brushes had considerable detectable DNA (Szkuta et al. 2017b). The transfer of human DNA by blowfly *Lucilia cuprina* has also been reported (Durdle et al. 2009).

Contamination avoidance is a well-known concept in DNA laboratories (e.g., Butler 2012, p. 18). The UK Forensic Science Regulator (UKFSR) has issued guidance on avoiding contamination in the DNA laboratory (UKFSR 2015), during sexual assault forensic medical exams (UKFSR 2016a), and at the crime scene (UKFSR 2016b). The Scientific Working Group on DNA Analysis Methods (SWGDAM) has also published guidelines on contamination prevention and detection (SWGDAM 2017b).

van Oorschot et al. (2015) discussed a program for monitoring and minimizing laboratory DNA contamination in the context of key performance indicators (KPIs) and the cost of such a program. Periodic sampling of work areas, blind proficiency testing of individuals, practitioner self-assessment of compliance, general compliance with audits, and practitioner observation and assessment were recommended. This approach called for root cause analysis when contamination was detected.

The Netherlands Forensic Institute (NFI) identified contamination as a particularly important quality concern. They published a study reporting on errors in casework during the period 2008 to 2012 (Kloosterman et al. 2014). NFI observed an increase in the number of cases of contamination over that time period. This increase was explained by an increase in the number of analyses, a more sensitive analytical system, an increase in the number of persons in the elimination databases (which allowed for more contamination to be recognized), and an increase in the requests for "touch DNA" evidence (Kloosterman et al. 2014). The NFI study distinguished between cases where there are multiple samples of DNA and those where the findings consist of a single low-level DNA sample. The authors noted that there are signals that would prompt a scientist to consider possible contamination, but only if the scientist is alerted to this possibility through tools like an elimination database (Kloosterman et al. 2014).

An article highlighted the possibility of DNA contamination in mortuaries and suggested that time and money may be wasted searching for profiles matching deceased individuals who may be already buried or cremated (Rutty 2000).

A recent report of contamination incidents in Austria over a 17 year period also highlighted the need for elimination databases (Pickrahn et al. 2017). The infamous Phantom of Heilbroun case involving contamination of swabs by the manufacturer (Neuhuber et al. 2009, Butler 2012, p. 79) may have prompted the study of potential contamination by police

officers collecting evidence at crime scenes (Nuehuber et al. 2017). Such contamination, which causes false positive results and can potentially mislead investigations, is an ongoing challenge for forensic laboratories and a constant reminder of the ease with which DNA transfers. A recent publication in this area presented lessons learned from a study of DNA contamination of police services and forensic laboratories in Switzerland (Basset & Castella 2018). An international documentary standard was published in 2016 to help address potential contamination in reagents and products used to collect and process DNA samples (ISO 18385:2016).

Given that DNA can transfer readily, precautions are needed both before and after evidence is submitted to a laboratory. Fonneløp et al. 2016 noted 16 instances of previously unknown police-staff contamination and called for a national elimination database or elimination protocol in Norway. The difficulty of identifying contamination if elimination databases are not in place is implicit in the following statement:

> "This and the previous source will be difficult to identify, since currently most morticians, pathologists, and even the police officers and their allied workers do not have their DNA profiles in the database for exclusion purposes" (Rutty 2000).

5.3.5. Studies Involving Casework Scenarios

As with any community of practice, some insight can be gained from a review of casework. Many groups have collated the type of samples from which successful results were obtained (Castella & Mangin 2008, Dang et al. 2012, Djuric et al. 2008, Dziak et al. 2018, Mapes et al. 2016, van Oorschot 2012). Other groups have considered particular evidence or sample types, such as adult necks (Graham & Rutty 2008), sandals (Ferreira et al. 2013), zip-lock bags in drug cases (Hellerud et al. 2008), and ammunition (Montpetit & O'Donnell 2015).

Several studies investigated the use of low amounts of DNA in various property crimes and proposed considering factors in a wider context (Forr et al. 2018). Some transfer studies attempted to mirror casework (Raymond et al. 2008a, 2008b, 2009a, 2009b; Fonneløp et al. 2017), while others sought to assess outcomes in mock scenarios (Benschop et al. 2012, Goray et al. 2012a). Finally, case context and interpretation issues in specific case examples were explored by several authors (McKenna 2013, Jackson 2013, Jackson & Biedermann 2019).

5.3.6. Literature on How to Evaluate DNA Relevance in Context

The scientific publications examined in this chapter cover properties of low-template DNA and provide insights into how those properties affect transfer and persistence. In addition, several publications describe approaches to interpretation that explicitly consider relevance of the DNA to the crime. Publications that cover this last topic are listed in Table 5.4.

Table 5.4. Summary of topics and some associated references explored in the following section.

Topics	References
Insight on the impact of low-template DNA	Taberlet et al. 1996, Gill et al. 2000, Gill 2001, Gill 2002, Gill & Buckleton 2010b, Gill et al. 2015, Benschop et al. 2015a
Case assessment and interpretation model (CAI) and the hierarchy of propositions	Cook et al. 1998a, Cook et al. 1998b, Evett et al. 2000a, Evett et al. 2000b, Evett et al. 2002, Jackson et al. 2006
Theoretical frameworks for assessing transfer evidence	Biedermann & Taroni 2012, Champod 2013, Taylor et al. 2018, Taylor et al. 2017d, Taroni et al. 2013, Taylor et al. 2019, Samie et al. 2020
Addressing propositions	Biedermann et al. 2016, Hicks et al. 2015, Gittelson et al. 2016, Kokshoorn et al. 2017, Taylor et al. 2017d
Distinction between investigation (police) and evaluation (court) uses of DNA	ENFSI 2015, Gill et al. 2018, UKFSR 2018a

The strategies and approaches presented in the publications listed in Table 5.4 are further discussed in the sections below.

5.4. Discussion

5.4.1. Implications of What We Know

If *every contact leaves a trace*, or, given the more correct interpretation, if every contact *might* leave a trace, one must separate relevant DNA from irrelevant DNA. There may be insufficient information in many areas. In addition, the studies reviewed herein show that the amount and likelihood of DNA transfer vary widely under different conditions. However, the possibility of transfer cannot be ignored when interpreting DNA evidence. If it is ignored, DNA findings, when considered in isolation, have the potential to be misleading.

From an overall reading of the literature described above, it is possible to outline several ways in which DNA transfer might mislead an investigation. These include the following:

- **It is possible to handle an item without transferring any detectable DNA to it.** The absence of detectable DNA was noted in 11% of experiments by Manoli et al. (2016) and in 2.9% by Goray et al. (2016). In addition, Meakin et al. (2017) noted that full profiles were not always detected. The shedder status of the donor as well as effectiveness of extraction and analytical methods are all relevant here and have been

extensively discussed in the literature (Lowe et al. 2002, Farmen et al. 2008, Taylor et al. 2016d, Taylor et al. 2017d, Taylor et al. 2018).

- **Genetic material may have been deposited before or after the crime and therefore may not be relevant to it.** This can happen because the person has legitimate access to the scene or item, or because the DNA was transferred in some other way (Raymond et al. 2009a, Goray et al. 2015). Studies examining persistence of original user or wearer following another user show that the substrate as well as the shedder status of the first and second user affect the findings (Fonneløp et al. 2015b, Oldoni et al. 2015, Oldoni et al. 2016, Meakin et al. 2015, Meakin et al. 2017, Pfeifer & Wiegand 2017).

- **Detected DNA might be present due to indirect (secondary or tertiary) transfer, whether by a person or an object.** These transfers can occur before or after the commission of a crime due to innocent activity in the area. They might also be the result of contamination during evidence collection, transport, and other stages of the investigation or during the laboratory submission, storage, and examination processes.

The above three points apply to any low-level profile and therefore also apply to profiles containing mixtures. While the traditional view is to focus on the major contributor to a mixture based on the assumption that the profile belongs to the last person to handle an item, some studies have shown this is not always the case (e.g., Cale et al. 2016, Buckingham et al. 2016, Goray et al. 2016).

Based on these factors, it is clear that even if a very high value of likelihood ratio (LR) is obtained, the DNA might not be relevant.

The highly sensitive DNA methods that have become common in recent years increase the likelihood of detecting irrelevant DNA. Peter Gill, in a review article covering the previous 20 years of development in the field, claims that all laboratories today are testing for low-template DNA (Gill et al. 2015).). Although the definition may be considered trivial, the method of interpretation is important: “the lower the amount of DNA present in a sample, the greater the chance that it may not be associated with a crime-event” (Gill et al. 2015).

Relevance was identified as an issue when low-template DNA work was first introduced: “Inevitably, there is a direct relationship between the quantity of DNA present and the relevance of the evidence” (Gill 2001). The authors of a study seeking to establish the limits for DNA mixtures using small amounts of DNA concluded:

> “The relevance of the evidence, rather than the DNA typing methodology or statistical model, may be the limiting factor for obtaining useful results for forensic casework and court going purposes.” (Benschop et al. 2015a)

The full implications of these observations have not yet infiltrated the routine practice of DNA testing in many criminal investigations. Instead, weight-of-evidence statistics (e.g., sub-source likelihood ratios; see Taylor et al. 2018) are often produced in forensic

laboratories as stand-alone findings, perhaps with a brief disclaimer in the accompanying report that mentions the possibility of transfer but does not treat this issue sufficiently. The studies in this chapter suggest that this area would benefit from more attention during routine practice to avoid potentially misleading findings.

The following section discusses strategies to help ensure that LRs are considered in context and to mitigate the risk that DNA transfer might mislead an investigation.

> **KEY TAKEAWAY #5.2: Highly sensitive DNA methods increase the likelihood of detecting irrelevant DNA. When assessing evidence that involves very small quantities of DNA, it is especially important to consider relevance.**

5.4.2. Strategies for Mitigating the Risk of Misleading DNA Results

5.4.2.1. Minimize Contamination

Forensic scientists have known since the advent of forensic DNA methods that DNA can transfer readily. This is evidenced by the systems that laboratories have had in place since then to avoid contamination (e.g., Butler 2012, p. 18). However, the use of highly sensitive methods increases the probability of detecting small amounts of contaminating DNA. "Along with increased sensitivity comes the prospect of detecting contaminating DNA, complicating the interpretation of profiles" (Szkuta et al. 2013).

The contamination avoidance strategies in forensic laboratories that have long been in place are more important than ever. Furthermore, as evidenced by the studies outlined in Table 5.3, contamination can happen during a scene investigation. Therefore, contamination avoidance procedures must be in place during all stages of an investigation, from the crime scene through the production of the profile. These studies also highlight the need for elimination databases (e.g., Basset & Castella 2018, Basset & Castella 2019) to avoid wasting resources following up on profiles that arise from the examination and also as a way of reducing complexity in mixtures.

Contamination can be seen to take various forms and can consist of stray alleles arising from unknown sources or profiles or alleles from persons handling the items, or it can result from inappropriate handling in the laboratory or transfer from one surface to another, which can be a particular risk when dealing with heavily blood-stained items.

Table 5.5. Examples of routes where contamination of DNA can occur as illustrated in the UK Regulator's guidance on DNA Anti-Contamination–Forensic Medical Examination in Sexual Assault Referral Centers and Custodial Facilities (UKFSR 2016a).

Direct transfer						
Sample	to	Environment/item				
Environment/item	to	Sample				
Consumable	to	Sample				

Person	to	Environment/item				
Indirect transfer—secondary transfer						
Environment/item	to	Examinee	to	Sample		
Environment/item	to	Consumable	to	Sample		
Environment/item	to	Practitioner	to	Sample		
Environment/item	to	Environment/item	to	Sample		
Person	to	Examinee	to	Sample		
Person	to	Environment/item	to	Sample		
Sample 1	to	Environment/item	to	Sample2		
Indirect transfer—tertiary transfer						
Person	to	Environment/item	to	Consumable	to	Sample
Person	to	Environment/item	to	Examinee	to	Sample
Environment/item	to	Environment/item	to	Examinee	to	Sample
Environment/item	to	Environment/item	to	Practitioner	to	Sample
Sample 1	to	Environment/item	to	Examinee	to	Sample 2

KEY TAKEAWAY #5.3: Highly sensitive methods increase the likelihood of detecting contaminating DNA that might affect an investigation. Contamination avoidance procedures should be robust both at the crime scene and in the laboratory.

5.4.2.2. Consider Evidence in Context

It is a principle of forensic science that results only have meaning in context (e.g., Evett & Weir 1998, Cook et al. 1998a, Cook et al. 1998b). The trend, however, is for the forensic scientist to have limited access to information about the case. This trend is driven in part by efforts to avoid confirmation bias. These efforts risk isolating the forensic scientist from contextual information that may be crucial when assessing relevance. It is possible to facilitate both approaches by sequential unmasking of information (Butler 2015a, pp. 461–464).

One way of considering evidence in context is to view the case as a whole rather than simply evaluating a single sample in isolation. As noted at the beginning of the chapter, Locard spoke of a criminal "leaving *multiple* traces of his path…" (emphasis added). This observation should serve as a caution against expecting a single association to *solve* the crime. This is in keeping with views expressed by others (e.g., Gill 2014, Sense about Science 2017) that DNA should not be used as the sole evidence in a criminal case, and that it is inappropriate to assume that DNA always has greater value than other types of evidence. A recent publication outlines a method for combining different types of evidence (de Koeijer et al. 2020).

A miscarriage of justice that occurred in Australia demonstrated these points. In this case, DNA was the only evidence in an alleged rape, but that DNA was later shown to have resulted from cross-contamination in a sexual assault examination room. The judge who later reviewed the circumstances that led to the conviction stated:

> "In the present case, the obviously unreserved acceptance of the reliability of the DNA evidence appears to have so confined thought that it enabled all involved to leap over a veritable mountain of improbabilities and unexplained aspects that, objectively considered, could be seen to block the path to conviction" (Vincent 2010).

This review cautioned that DNA

> "must be carefully used and placed into proper perspective and understood that a calculation of statistical likelihood provides a dangerous basis for conviction, if it is upon that alone that proof beyond reasonable [doubt] rests" (Vincent 2010).

While the Australian case involved cross-contamination of evidence, the warning from the judge about misusing a statistical likelihood applies to any case that may involve DNA transfer. The LR, as typically used when interpreting DNA mixtures, is based only upon the analytical properties of the DNA. It does not provide information about other important aspects of the evidence, such as the quantity of DNA or the whether the cell type is known. Therefore, a large blood stain might produce a very similar LR to a swab from a light switch, yet the two have would very different meanings in the context of a case (e.g., Taroni et al. 2013). While an LR value is an expression of the strength of evidence under a pair of propositions, the result should be considered in context (i.e., the result represents the evidence for what?).

5.4.2.3. Ask and Answer the Right Questions

Keith Inman and Norah Rudin have written: "One of the greatest unrecognized contributions that a criminalist can provide [to a case] is framing the correct question" (Inman & Rudin 2001). The fact that this quote is taken from their section "Reasoning from Traces of Past Events" supports the view that the function of forensic science is to shed light on a past event. In this context, it is important to carefully consider what questions are being addressed.

The trier of fact needs to know the answers to multiple questions, many of which the forensic scientist cannot address. Who, What, When, Where, How, and Why all need to be answered at the criminal trial. The LR as typically used in DNA mixture interpretation addresses the *who* question, but it does not address the questions of *when* and *how* the DNA was deposited. This presents a risk that the trier of fact might use an answer to a relatively easy question to answer the more difficult questions. A recent review article, describing this phenomenon as an attribute substitution, stated: "If someone doesn't know the answer to a difficult question, they will substitute an easier question (even if subconsciously) and answer that instead" (Eldridge 2019). This tendency highlights the need to be clear about what questions are being addressed with any particular interpretive method.

KEY TAKEAWAY #5.4: DNA statistical results such as a sub-source likelihood ratio do not provide information about how or when DNA was transferred, or whether it is relevant to a case. Therefore, using the likelihood ratio as a standalone number without context can be misleading.

5.4.2.4. Use Case Assessment and Interpretation

The references cited in Table 5.4 include a paper that introduces a framework for ensuring that case context is considered when evaluating evidence (Cook et al. 1998a). Case Assessment and Interpretation (CAI), which has come to be known as evaluative reporting, provides a systematic way to produce "an assessment of the strength to be attached to the findings in the context of alleged circumstances" (ENFSI 2015).

CAI requires the forensic scientist to document their expectations in a given scenario before examining the evidence. For example, a violent assault involving significant bloodshed would typically be expected to yield multiple transfers to the assailant rather than trace amounts of DNA of no known cell type. Documenting expectations in this way can help avoid being "findings-led" (i.e., trying to make the findings fit the case). Without an assessment before examinations, the scientist can be accused of drawing the target after the shot is fired, also referred to as the Texas sharpshooter fallacy (Thompson 2009).

CAI serves as the basis of several guidelines developed over the last 10 years (AFSP 2009, ENFSI 2015, ANZPAA 2017). The principles of CAI include:

- The findings are assessed in the context of the case, because they have no intrinsic value in isolation.
- At least two propositions are considered when using the LR. The assessments are dependent on the propositions addressed.
- In order to avoid what is commonly referred to as "transposing the conditional," (Thompson & Schumann 1987), the scientist reports on the findings, not the propositions.

5.4.2.5. The Hierarchy of Propositions

The researchers who formulated the CAI framework outlined a hierarchy of propositions, with each level addressing different questions (Cook et al. 1998b). This helped to clarify the questions addressed during evidence evaluation (Cook et al. 1998b, Evett et al. 2000a). The propositions at the lower end of the hierarchy—source, sub-source, and sub-sub-source—are defined in Figure 5.3. These levels only address questions about the source of the DNA. An example of a source-level proposition might be that the DNA mixture contains DNA from the POI and the victim. These source- or sub-source-level propositions are based on the genotypes or alleles present in the evidence, but they do not address in any way *how* the DNA was deposited.

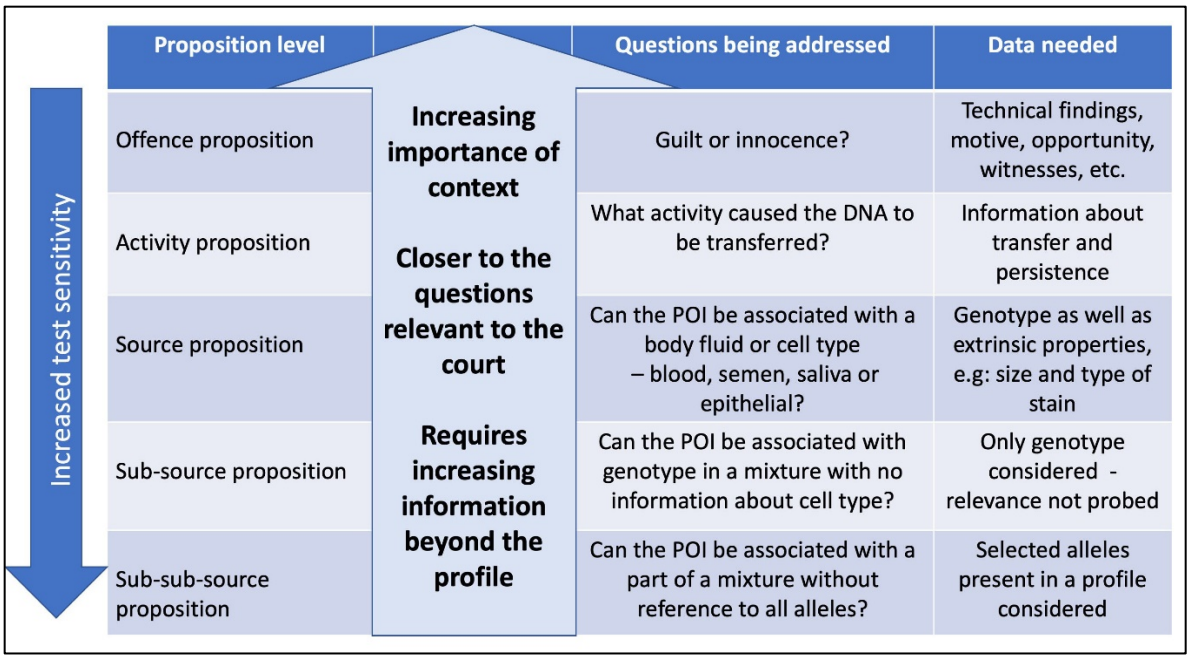

Figure 5.3. The hierarchy of propositions (adapted from Taylor et al. 2018).

Above the source-level propositions are activity propositions, which address questions about *how* the DNA came to be present in a mixture. An activity proposition might be, for instance, that DNA collected during a sexual assault examination was deposited during sexual activity, or that DNA found on the handle of a knife was deposited during the act of stabbing a victim. Activity-level propositions more directly address issues of interest to the court (Jackson 2013, Taylor et al. 2018), and they almost always involve greater uncertainty than source-level propositions.

Finally, offense-level propositions address questions of guilt or innocence. These questions are generally addressed by the courts rather than by forensic scientists.

It is vital that users of forensic science information understand the differences between levels in the hierarchy and that they do not use the LR for one level to address a question at a higher level. It has been noted:

> "Due attention must be paid to the position in the hierarchy of propositions that can be considered. This information must be effectively conveyed to the court to avoid the risk that an evaluation at one level is translated uncritically and without modification to evaluation at a higher level. We cannot over-emphasize the importance of this. A DNA match may inform decisions about the source of the DNA, but decisions about an activity, say sexual intercourse versus social contacts, involve additional considerations beyond the DNA profile." (Buckleton et al. 2014)

Peter Gill also discussed the risks of conflating source and activity propositions in his book *Misleading DNA Evidence: Reasons for Miscarriages of Justice* (Gill 2014). This book introduces the concept of an "association fallacy," where "a probability is transposed from one level of the framework of propositions to higher level." Several miscarriages of justice

have been shown to result from misleading DNA evidence due to this fallacy (Gill 2014, Gill 2016, Gill 2019a). This risk is increased by the fact that the vast majority of criminal cases in the United States are settled through plea bargaining (Gramlich 2019). Suspects and attorneys may overestimate the value of the DNA findings and accept a plea possibly even when innocent.

5.4.2.6. Activity Propositions

There is evidence that activity-level questions have been receiving greater attention in court in recent times (Taylor et al. 2018). The CAI approach involves formulating activity-level propositions in order to calculate an activity-level LR. For example, in the case of a stabbing, the prosecution hypothesis might be that the DNA was transferred to the handle of a knife during the activity of stabbing, while the defense hypothesis might be that the DNA was deposited due to contamination or secondary transfer. There are many references in the literature to the suitability of this approach but little in the way of prescriptive assistance. Bayesian networks have been suggested as a method with which to identify those variables that are most likely to impact the activity-level LR (Taylor et al. 2017d, Biederman & Taroni 2012, Taylor et al. 2019). Depending on the questions being addressed, the sub-source LR may not be relevant. This is true when trying to differentiate the expected findings in light of the potential of primary or secondary transfer, for example.

Simulation and modelling are used to assess the impact of variables on LRs based on activity propositions. The results show that regardless of the DNA outcome, the most impacting variable is the "DNA match probability when the defence alleged that the person of interest (POI) had nothing to do with the incident". When secondary transfer is alleged, the DNA match probability has less impact and variables associated with the donor are important. Extraction, sampling quantity of DNA on hands and background are the variables to be considered. The authors provide a tool to assess the impact of varying the latter two parameters (Samie et al. 2020).

The LRs produced from activity propositions are generally much lower in numerical value than those produced from source propositions. An early paper illustrated this observation, showing an activity level LR of the order of 1000, in contrast to what the authors describe as an infinite LR in favor of a sub-source level proposition (Evett et al. 2002). Some have argued that, given that activity propositions produce more conservative assessments of weight of evidence and are more relevant to the issues of the court, their use is more appropriate (Biedermann et al. 2016b, Kokshoorn et al. 2017, Taylor et al. 2018, Szkuta et al. 2018).

In addition, it is possible to get value from the CAI approach after the production of a statistic by having another scientist carry out an assessment and assign probabilities for transfer, errors, contamination, etc., and then evaluate the findings in light of the previously analyzed results.

5.4.2.7. The Value of CAI-based Reasoning

In its fully realized form, CAI involves formulating activity-level propositions, assigning probabilities to those propositions, and calculating an activity-level LR. Assigning those probabilities requires an understanding of DNA transfer and persistence. For instance, the probability that a person transferred DNA onto the handle of a knife during a stabbing would be affected by the material that the knife handle is made of (wood versus plastic), the shedder status of the person of interest, and the history of the knife. There may be insufficient empirical data to assign probabilities based on these factors. Some researchers have argued that, in that case, it would be appropriate to assign "subjective probabilities" (Biedermann et al. 2016a, ENFSI 2015), while others have argued that this would not be appropriate (Meakin & Jamieson 2013). In either case, documenting expectations and identifying propositions as required by CAI are useful ways to consider context, avoid being findings-led, and ensure that the findings address appropriate questions.

The efficiency and throughput of DNA laboratories may work against these organizations taking on these issues and ignoring relevance for various reasons, including the fact that they are not always aware of case context.

Confining the report to a sub-source LR and answering questions about relevance if and when they arise in court are not balanced efforts and are therefore likely to be biased to one side or the other depending on the circumstances. Discussion about the lack of suitability of this approach is well argued in Biedermann et al. (2016b).

Balance, transparency, logic, and robustness were suggested as four requirements for reporting of scientific findings (ENFSI 2015). Confining the requirements to robustness in isolation is not sufficient to ensure that the court is fully informed.

CAI was originally formulated to help assess the tests that would be most probative. With laboratories under ever-increasing pressure to conduct more tests, this type of analysis would help ensure that laboratory resources are used most effectively. The Resource Group (see Chapter 1) strongly supported the notion that decisions about what evidence items to test should be made by forensic experts rather than policy-makers. CAI provides an ideal framework for making these types of decisions but requires that these experts are familiar with the transfer and persistence of DNA and their laboratory's ability to detect such transfers.

5.4.2.8. Separate Investigation from Evaluation

There are two phases in assessing evidence in a criminal case. During the investigative phase, the goal is to narrow the lines of inquiry and produce a suspect. During this phase, questions of relevance may be set aside while the police might identify other evidence that might provide context. During the subsequent evaluation phase, the scientist would evaluate the evidence by formulating competing propositions that are based on the surrounding case circumstances.

The DNA Commission of the International Society for Forensic Genetics (ISFG) distinguishes between investigative and evaluative modes when using LRs (Gill et al. 2018). The UKFSR does as well (FRS-G-222; UKFSR 2018). Both sets of guidelines anticipate a scientist delivering results in an iterative manner. The challenges and advantages of this approach have been outlined previously (Buckleton et al. 2014). Separating the investigation and evaluation phases has a major impact on the propositions used in LR calculations. The investigator produces information or explanations for findings at a scene. The investigative mode is most appropriate when it is not possible to formulate a pair of propositions or when there is insufficient conditioning information (ENFSI 2015).

The ISFG DNA Commission states:

> "The scientist works in an investigative mode if there is no person of interest in the case. If a suspect is identified, then generally the scientist switches to evaluative mode with respect to this suspect and needs to assign the value of their results in the context of the case. If there is new information (in particular from the person of interest), the scientist will need to re-evaluate the results. It is thus important that reports contain a caveat relating to this aspect" (Gill et al. 2018).

At source level, an evaluation might consider including relatives in the propositions. It also might affect conditioning on particular genotypes if, for instance, the evidence includes the victim's DNA, as often happens in cases of sexual assault. At activity level, wider issues such as opportunities for transfer, persistence, and shedder status should also be considered.

These nuances in different uses of DNA and the effect of different propositions are well reflected in the literature. Nonetheless, in practice, the focus is on the number—that is, the LR. Authors of a recent article on formulating propositions stated:

> In [their] experience, "this may be referred to as 'the number' by prosecutor and defense attorney. This practice breaks the connection between the LR and the propositions, and this is regrettable. Discussion in court very likely evolves to activity level, yet there is no direct relationship between the LR for sub-source level propositions and one for activity level propositions." (Gittelson et al. 2016)

The "number" (LR value) is like seeing the highlight of an advertisement without reading the small print and considering the propositions behind the number. Kwong recognized this for DNA in a *Harvard Law Review* article:

> "Yet despite the perception of DNA evidence as definitive proof, when DNA evidence involves complex mixtures of multiple individual's DNA, science is not as simple as it appears on television." (Kwong 2017).

The evaluation stage is an opportunity to use the risk-mitigating strategies outlined previously, to review the findings in light of the case context, to assess the possibility of contamination or error, and to formulate activity propositions. It may also be necessary to do additional sampling, seek information about other genotypes in the mixture, or conduct *ad*

hoc transfer experiments that apply to the particulars of the case. This has been referred to as "sense making" by Paul Roberts (Roberts & Stockdale 2018).

The Deputy Commissioner for Crime of the Victoria Police in Australia has commented:

> "DNA matching [is] very valuable to police for intelligence and evidentiary purposes, but, when used as evidence, *[has] to be seen as one part of a circumstantial case and not as the entirety of it*" (Vincent 2010, emphasis added).

KEY TAKEAWAY #5.5: The fact that DNA transfers easily between objects does not negate the value of DNA evidence. However, the value of DNA evidence depends on the circumstances of the case.

5.4.2.9. Examples to Illustrate Use of Mitigating Strategies

Two hypothetical case scenarios are considered as an illustration of the importance of context. Each scenario contains the same finding of a knife on which a three-person mixture is detected. A reference sample from the person-of-interest (i.e., the individual suspected of stabbing) is also provided, which is associated with the mixture profile found on the knife.

Case A
This case involves a stabbing in a private home. During a burglary attempt, the burglar is disturbed by the house owner and grabs a knife from the kitchen, which he uses to stab the house owner. In this case, the relevance of DNA results on the knife handle may be obvious.

Case B
This case involves a knife fight in a hotel kitchen. During this fight, a chef is thought to be stabbed by a coworker. The knife is later recovered in a nearby alley. In this case, the relevance of DNA results on the knife handle may not be obvious.

Expectations and Risks: There is a lower risk of using a sub-source LR result in isolation with Case A than with Case B. In Case A, the burglar had no previous access to the house, and therefore finding an association with the knife would be probative. In Case B, the same finding needs more investigation before the relevance of a DNA result can be assessed. If a suspect in the hotel kitchen stabbing case had prior access to the knife as part of his or her job, then there is some expectation that a profile matching him or her would be detected on the knife handle before the stabbing occurred.

Considering Possible Contamination: To reduce risk that a profile arose from contamination, additional scene samples could be taken, particularly from areas expected to be handled by the assailant. If that same DNA appears in multiple evidence items, contamination would be less likely (Jackson 2013, NRC 1996).

Ask and Answer the Right Questions: In both cases, the real question being sought from the DNA finding is whether the POI transferred his or her DNA to the knife handle while

stabbing the victim. In other words, the important question is at the activity level. In Case A, if the POI's profile can be associated with the knife, then one might infer that the transfer happened during the stabbing (though the possibility of contamination must be considered). Therefore, the elevation of sub-source questions to activity questions (i.e., moving from sub-source LR to activity LR) is low risk. However, in Case B, the elevation of sub-source to activity level is higher risk. Reporting an association between the POI and the knife, where there is the possibility of the POI's profile being present prior to the crime, cannot be taken to indicate that it was transferred there at the time of the stabbing. For example, DNA transfer studies tell us that the last person to handle an item may not be the major profile in a mixture. In this instance, we also have a situation where there are at least three contributors to the mixture, so further investigation is necessary. Triers of fact should be made aware that the LR value addressing a sub-source level question is not sufficient evidence that the POI transferred his or her DNA to the knife at the time of the stabbing.

Using the CAI Approach: For Case B, a preassessment of the case might prompt questions as to when the knife in question was last used and a decision on whether a sub-source LR would be helpful. Also, additional samples may be requested to get a fuller picture of the shedding characteristics of the POI to help assess whether DNA from regular use would be expected. This would depend on when the knife was last used and assumptions about how long it was handled during the knife fight. There may not be sufficient data available, in which case the findings would be neutral. At a minimum, the risk of misleading information based on sub-source LRs alone must be emphasized to the trier of fact.

5.4.3. Growing Awareness of DNA Transfer and Persistence

Interest in DNA transfer and persistence studies has grown over the last 20 years. A recent review noted a growth from five papers published in 2000 to 35 articles on the topic in 2015 (Kokshoorn et al. 2018). In spite of an increase in the number of published studies on DNA transfer, the results of these studies have not been combined to deal with broad questions about transfer mechanisms (Taylor et al. 2017d, Gosch & Courts 2019). Rather, information from published studies can be seen as a way of gaining sufficient knowledge to address the questions being raised in court about how DNA is deposited. A logical framework in which questions of transfer mechanism can be approached probabilistically has been published, together with identification of the gaps that need to be addressed (Taylor et al. 2017d).

One of the reasons there is so much variation in the results of the transfer studies is that results can vary across laboratories, as interlaboratory studies show (Steensma et al. 2017, Szkuta et al. 2020). Therefore, any laboratory planning to assist the court by offering probabilities based on these studies will need to adjust for their own level of sensitivity. For example, if the laboratory has a higher level of sensitivity than a particular study, their likelihood of detecting transfer may be higher than the study would suggest.

KEY TAKEAWAY #5.6: There is a growing body of knowledge about DNA transfer and persistence, but significant knowledge gaps remain.

5.5. Summary

One of the foundational principles of forensic DNA analysis is that DNA transfers and persists (see Chapter 2). This is what makes it possible to investigate crimes using DNA in the first place. However, this also means that the relevance of DNA to a crime cannot be taken for granted and needs to be assessed, because when DNA transferred and whether it transferred directly or indirectly affect its relevance to the crime. This is the obvious overall implication from the studies presented in the earlier part of this chapter.

Furthermore, a sub-source LR value (or other statistic) produced by mixture interpretation methods considers only the rarity of the profiles. It does not say anything about whether the DNA is relevant to the crime and may well contain genotypes not relevant to the crime. Therefore, it is important that the LR not be used in isolation. Instead, one must consider the LR within the larger context of the case and ensure that stakeholders do not use the sub-source "number" alone as an indication of the contribution of DNA to the case.

6. Chapter 6: New Technologies: Potential and Limitations

New technologies are often investigated to assess whether they can provide solutions to existing problems in the forensic community. The adoption and implementation of these technologies depends upon a cost/benefit analysis within forensic laboratories. An appreciation of fundamental challenges with DNA mixture interpretation can provide an impetus to consider whether new approaches can bring desired improvements. The ability to analyze short tandem repeat alleles by sequence, in addition to length, promises to bring some new capabilities to forensic DNA laboratories. Next-generation sequencing platforms also enable additional genetic markers to be examined. Microhaplotypes have been pursued for their potential to improve DNA mixture interpretation. Additionally, cell separation techniques offer the potential to separate contributors prior to DNA extraction.

6.1. Technology Development and Drivers

Previous chapters have examined measurement and interpretation issues (Chapter 4) and case context and relevance for DNA mixtures (Chapter 5). This chapter explores the potential and limitations of new technologies to assist with DNA mixture interpretation.

As described in Appendix 1, DNA technologies (and interpretation approaches) have advanced over the past three decades. These advancements have been fueled largely due to ongoing efforts in biotechnology, specifically the commercialization of new instruments and techniques for clinical analysis and large-scale DNA sequencing efforts. Having multiple uses for a single technology allows commercial manufacturers to develop application-specific products with minimal risk. Thus, "piggy-backing" onto these broader advances provides capabilities to the forensic DNA community that would not be available otherwise. A prime example is the capillary electrophoresis (CE) technology that was developed for chemists to separate molecules according to size and charge, but also enabled the sequencing of billions of nucleotides for the Human Genome Project (Lander et al. 2001).

Over the past 20 years, CE technology has been the mainstay in forensic DNA laboratories around the world for separation and detection of short tandem repeat (STR) markers, starting with the ABI 310 Genetic Analyzer and then multi-capillary ABI 3100, 3130, and 3500 systems (Butler 2012, pp. 141-165). Some high-throughput forensic laboratories have also implemented the 3700 or 3730 Genetic Analyzers with 48 or 96 capillaries.

The polymerase chain reaction (PCR) is also used broadly in molecular biology, and forensic applications combine this method with fluorescently labeled primers to enable various configurations of STR typing kits. These kits have evolved both in terms of sensitivity and the number of targeted STR markers – the latter in keeping with increases to DNA database core sets (Gill et al. 2006a, Hares 2012, Hares 2015). Modern CE-based STR kits examine over 20 locations in the human genome from only a few cells (Butler 2012, Butler 2015a). An increase in STR typing kit sensitivity improves detection of proportionally lower-level contributors in DNA mixtures, potentially resulting in a greater number of alleles in a mixed DNA sample. Although collecting more information is generally viewed as positive,

examining additional data can add to the complexity of interpretation and communication of results obtained from an electropherogram (EPG).

Millions of STR profiles, primarily single-source reference samples from convicted offenders or arrestees, now exist in national DNA databases around the world, with substantial resources invested to create these law enforcement databases. With increasing knowledge of the human genome, new genetic markers are being proposed for forensic identification purposes. This is described later in this chapter. However, adoption is challenging due to the existence of large STR profile databases (see Butler 2015b). Before implementing a new technology, the degree of potential improvement needs to be considered in terms of the amount of information gained along with the cost and effort of changing.

The marketplace has played an important role in developing forensic DNA typing technology. The forensic DNA community uses commercial DNA extraction and quantification kits, STR typing kits, CE instruments for detection, and software for analysis and data interpretation (Figure 6.1). The adoption of commercially available options has led to more uniformity of methods employed in laboratories and consistent quality control. However, these same benefits can result in an increased reliance on ready-made solutions. This can result in lost opportunities for innovation.

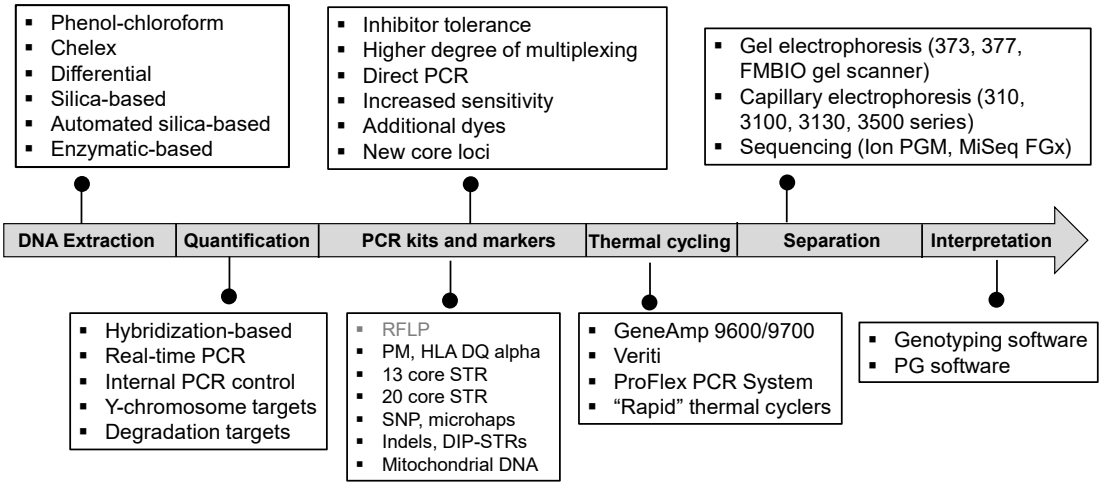

Figure 6.1. Advances and introduction of new technology to support the STR typing workflow.

Commercial suppliers must consider production and sales volume in deciding which products to develop and maintain in the marketplace. Thus, even if new technologies are developed, they may not be implemented in the forensic arena for reasons that can be either technology-based or market-driven. A proposed solution with a new technology may not sufficiently address the problem it is trying to solve to warrant change. A forensic laboratory determines whether the cost (including time and labor) of purchasing, training, performing internal validation experiments, implementing, and maintaining new procedures or equipment is expected to provide a satisfactory solution to an existing problem. We note that forensic laboratories can perform developmental validations for methods established in-house; however, most methods originate in the commercial sector where the vendor performs the

developmental validation. Vendors often collaborate with a forensic laboratory on the developmental validation, but most forensic laboratories are solely performing internal validation studies.

Adopting a new method or technology is not necessarily a linear process. Therefore, understanding the complexity of DNA mixture analysis and the way a new technology may or may not overcome known difficulties is important. Although a formal process for adoption and implementation does not exist, general steps can be considered. Table 6.1 lists considerations in deciding whether to adopt a new technology.

Table 6.1. Steps and considerations for implementing a new technology or method into practice.

Steps	Considerations
Research and Development	• Review work performed by commercial vendors or researchers • Seek input from technical working groups or previous adopters
Evaluation	• Perform informal studies (e.g., beta tests) • Examine early stage publications describing the potential of the new technology performed by researchers or other practitioners
Decision to Move Forward	• Assess the "cost" (e.g., personnel time, new equipment) • Consider available funding for adoption • Weigh the changes and potential impact (e.g., new core loci, change in vendor support) • Critically assess benefits of the new technology to address issues and consider potential limitations
Internal Validation	• Examine published developmental validation studies (typically performed by the vendor) • Conduct internal validation studies • Perform additional supporting experiments as needed
Implementation	• Prepare standard operating procedures (SOPs), conduct training and competency testing, and establish proficiency testing and reporting/testimony guides
Other	• Evaluate if additional documentary or physical standards are needed

6.2. Fundamental Mixture Challenges

In this section, we examine the challenges that are fundamental to DNA mixtures and areas of possible improvement via new technologies.

Sample collection, extraction, and quantitation are the first steps in the DNA measurement and interpretation workflow (see Figure 2.1 in Chapter 2). Improvements in DNA extraction

efficiencies can help ensure maximal recovery of the evidence and, in theory, reduce the potential for stochastic variation observed with lower amounts of DNA (e.g., minor components in a mixture).

A DNA mixture arises when cells from multiple contributors are present in a sample. These cells are physically distinct prior to DNA extraction, but the DNA from those cells commingles and mixes during and after the extraction process (Figure 6.2). Thus, if cells from different contributors to a sample could be physically separated prior to extraction, then cells from each contributor could potentially be analyzed separately as a single-source sample. For example, chemical differences of the cell walls of sperm enable differential extraction to partition a sexual assault victim's epithelial cells from a perpetrator's sperm cells (Gill et al. 1985). However, when cells from multiple contributors are co-extracted, DNA mixtures result.

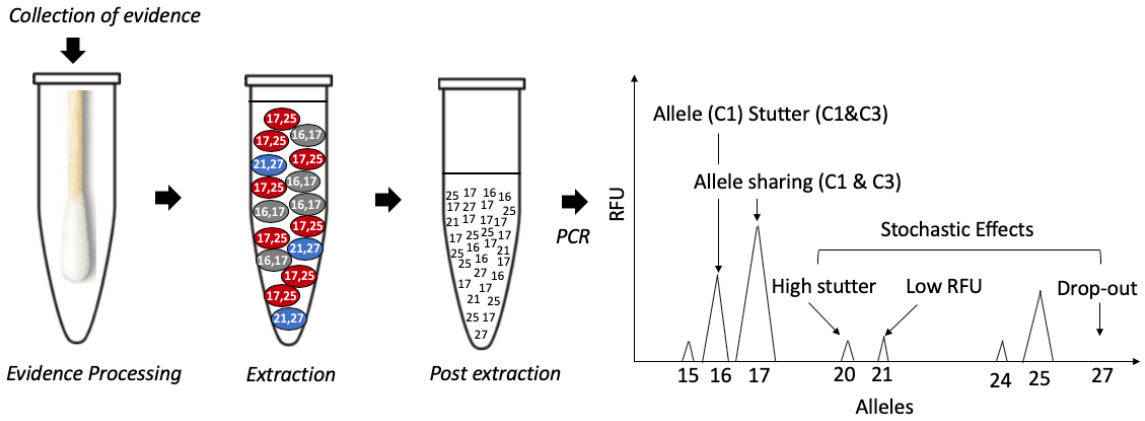

Figure 6.2. General illustration of steps involved in generating a DNA mixture profile and some of the possible factors in interpretation. If an evidentiary swab contains a mixture of cells from three contributors (Contributor 1 (C1) [grey], Contributor 2 (C2) [blue], Contributor 3 (C3) [red]) and the corresponding genotypes at one STR locus as an illustration are (Contributor 1 [16,17], Contributor 2 [21,27], Contributor 3 [17,25]), then allele sharing occurs with the "17" allele. If only a few cells are recovered for one or more of the mixture contributors, then stochastic effects, such as high stutter, heterozygote peak imbalance, and allele drop-out may occur.

From a measurement and interpretation standpoint, several challenges are fundamental to DNA mixture interpretation (see Chapter 2). Briefly, with any PCR system, there will be ***stochastic variation*** when small amounts of DNA are analyzed. Stochastic effects impact the recovery of alleles and genotypes from mixture samples and lead to uncertainty in assigning alleles to genotypes and genotypes to contributor profiles. When STR markers are examined, ***stutter products*** add noise to the system. Stutter products impact uncertainty when alleles from minor contributor(s) overlap with stutter peaks of alleles from major contributor(s). Use of non-repetitive genetic markers (described further in section 6.4.2) can avoid stutter products but may not possess the genetic variation of STRs, which are needed to improve detection of genotypes from multiple contributors. Finally, ***sharing of common alleles*** can mask the presence of contributor alleles and affect the ability to estimate the number of contributors. When combined with stochastic variation and the existence of stutter products, allele sharing increases the complexity of a DNA mixture.

Allele sharing is illustrated in Figure 6.2 with allele 17 of Contributor 1 and Contributor 3. Stutter products (of allele 17) can also overlap an allele of the same length (allele 16). Stochastic effects can lead to high stutter (what appears to be an allele 20) and missing information (drop-out of allele 27). The illustration in Figure 6.2 does not account for further complications in the data caused by DNA degradation, PCR inhibitors, contamination (see Chapter 5), or cell-free DNA that may also be present in collected forensic evidence. STR allele sequencing technologies that rely on PCR amplification will still be subject to these fundamental mixture issues.

6.3. Possible Improvements: Physical Separation of Cells

Physically separating cells from different contributors prior to DNA extraction and STR typing can reduce the need for DNA mixture interpretation (Figure 6.3). This separation is an attractive concept but presents new challenges of working directly with cells prior to DNA extraction.

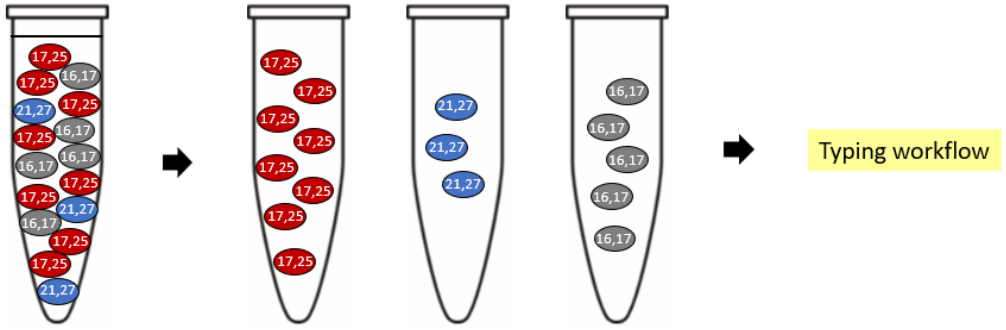

Figure 6.3. Illustration of physical separation and sorting of cells based on properties unique to a contributor's cell -type.

Separating cells from multiple contributors can sometimes be performed with laser-capture microdissection (Ballantyne et al. 2013) or micromanipulation (Farash et al. 2015). Cell separation can also be based on a unique property, such as the binding of a specific antibody to a unique feature on the cell surface (Verdon et al. 2015, Fontana et al. 2017). This type of work has been described using fluorescence assisted cell sorting (FACS) methods and fluorescently labeled antibodies (Verdon et al. 2015, Dean et al. 2015, Stokes et al. 2018). Proof-of-concept research has been conducted, but the work is laborious and usually demonstrated on fresh samples.

In one micro-manipulation approach, 40 discrete "bio-particles" (20 single and 20 clumped cells) were collected under a microscope and subjected to PCR conditions optimized for low-level DNA detection, resulting in recovery of single-source STR profiles in 41% of the 479 tested samples (Farash et al. 2018). Another approach for recovering individual cells is the DEPArray system, which is an image-based, microfluidic digital sorter that can isolate pure cells (Fontana et al. 2017, Williamson et al. 2018). DNA profile recovery can also be improved through separating PCR inhibitors and DNA templates using a digital agarose

droplet microfluidic approach (Geng et al. 2015). Similarly, agarose reactors can also allow for single-cell PCR within an encapsulated droplet (Geng et al. 2014).

One of the challenges of the FACS and microreactor methods is that crime scene evidence is typically composed of dried cells and may also contain cell-free DNA adhering to the outside of cells (Wang et al. 2017). The reconstitution of cells is not always straightforward, and it is important to maintain the integrity of the cell membrane to avoid mixing DNA from multiple cells. Dried cell membranes are more permeable and fragile, which may lead to cell breakage and DNA loss during preparation (Verdon et al. 2015). In addition to demonstrating success with samples subjected to real-world conditions, cell separation workflows would need to be streamlined prior to widespread adoption in the forensic laboratory.

6.4. Possible Improvements: Sequencing

Next-generation sequencing (NGS), also known as massively parallel sequencing (MPS) in the forensic community, has been used for more than a decade to perform high-throughput DNA sequencing for biotechnology discovery purposes (Hert et al. 2008). NGS is widely described as important to the future of forensic DNA testing (Børsting & Morling 2015, Alonso et al. 2017, Alonso et al. 2018). Table 6.2 summarizes potential benefits and issues with the use of new sequencing technologies for DNA mixture interpretation. Compared to existing CE-based methods, NGS provides an additional dimension and more detailed resolution of genetic information, which includes the sequence of targeted PCR amplicons and accompanying stutter products with STR alleles.

Table 6.2. Summary of the application of STR sequencing technologies to DNA mixtures.

Topics	Comments on Capabilities, Limitations, and Unknowns in Comparison to CE Methods
Smaller PCR Amplicons than CE	• Smaller and more consistently sized PCR products across STR loci (without the need to separate by size on an EPG) improve performance, particularly with degraded samples
Larger Multiplexes than CE, Potential Additional Markers	• Additional markers can be analyzed simultaneously to include more autosomal STRs, X and Y chromosome STRs, mitochondrial genome, single nucleotide polymorphisms (SNPs), and microhaplotypes • Additional information could potentially improve estimates for the number of contributors in a DNA mixture • Need to assess whether the observed mixture ratios of contributors are maintained across the examined loci
Targeted PCR similar to CE	• Sensitivity similar to CE methods • Sequencers may tolerate a higher PCR DNA input than CE • Stochastic effects still present with low amounts of DNA

Topics	Comments on Capabilities, Limitations, and Unknowns in Comparison to CE Methods
Different Artifacts from CE	• Fluorescent dye artifacts are not present (e.g., spurious EPG noise peaks, spectral "pull up", or dye blobs) • Sequence-based artifacts may arise (e.g., homopolymers, phasing)
Different Determination of Thresholds	• Analytical thresholds, which discern noise sequences from biological sequences of STR alleles, are based on sequence data rather than CE molecule fluorescence
Sequenced Stutter Products	• Potential exists to discern a stutter product from a minor contributor allele if the allele sequence differs • Examination of the sequence context can allow a more accurate modeling of stutter product amounts • STR markers consisting of multiple repetitive regions may produce multiple stutter products per allele
Additional STR Alleles	• STR sequences may differentiate some identical-by-length STR alleles, separating some mixture components possessing shared alleles, which in turn may assist in an improved estimate of the number of contributors to the mixture • Not all STR loci experience significant gains from sequencing (e.g., TPOX, TH01) • Additional STR alleles requires sequence-based allele frequencies for statistical calculations • Sequenced STR alleles are compatible with current DNA databases using length-based STR information
Interpretation	• To take full advantage of sequencing capabilities, an NGS-based probabilistic genotyping model will be required

In a 2015 review article, the authors state:

> "Sequencing of complex and compound STRs with many alleles of the same size may simplify mixture interpretation, if the contributors have alleles of the same size with different sequence compositions or if the true allele of the minor contributor has a different sequence than the stutter artifact of the major contributor" (Børsting & Morling 2015).

Furthermore, the authors note the difference between detecting alleles and distinguishing alleles from artifacts and noise:

> "It was recently demonstrated that sequences from the minor contributor in 1:100 or 1:50 mixtures were detectable by NGS – something that is not possible with the current PCR-CE technology. In these types of mixtures, the reads from the minor contributor will be difficult to separate from stutters and noise sequences, however, the mere fact that they could be identified opens up for new possibilities in mixture interpretation and it is certainly something that should be explored further" (Børsting & Morling 2015).

Figure 6.4 illustrates the expected results from sequencing of the STR locus that was typed with CE methods and shown in Figure 6.2. Stochastic sampling effects similar to those encountered with CE data will continue to exist with amplified and sequenced low-template samples. For example, high stutter (from C2) and allele drop-out (27 allele of C2) are not addressed through sequencing, and allelic imbalances (not shown) could still impact the genotype determination of a contributor.

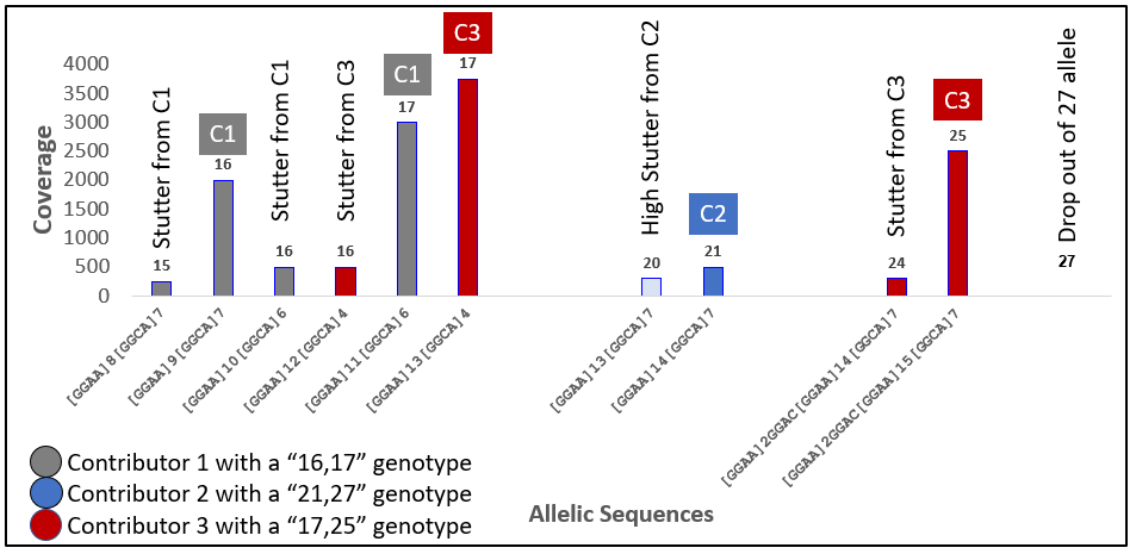

Figure 6.4. Illustration of results in "sequencing space" for the mixture example in Figure 6.2. The allele length and sequence are represented along the horizontal axis while relative sequence abundance (coverage) for the various alleles and stutter products is shown on the vertical axis. The same length "17" alleles from contributor 1 (C1) and contributor 3 (C3) can be resolved from one another. In addition, the stutter products from C1 and C3 can be separated by sequence from the "16" allele of C1.

Note that in Figure 6.4, the "17" allele (from C1 and C3) are distinguishable from one another through sequencing as are their corresponding "N-1" stutter products. In general, the degree of allele sharing is expected to decrease corresponding to an increase of observed alleles by sequencing, along with improved resolution and characterization of stutter artifacts. Each of these sequenced "17" alleles will have an associated sequenced-based allele frequency that would be applied in a statistical calculation, strengthening "matches" compared to a length-based STR analysis. The magnitude of the improvement will depend on the exact scenario and allele combinations, with gains expected primarily from the more complex STR markers, such as D12S391, D2S1338, and D21S11 (e.g., Gettings et al. 2018), as shown in sequenced-based allele frequency publications (summarized in Table 1 of Gettings et al. 2019). STR sequence-based nomenclature formats are under discussion (Parson et al. 2016, Phillips et al. 2018, Gettings et al. 2019), and will need to be determined to facilitate data exchange across laboratories.

6.4.1. NGS Studies of STR Markers with DNA Mixtures

Because sequencing forensic STR markers is relatively new, much of the initial mixture-related work in the literature consists of straightforward mixture *detection* experiments,

rather than deconvolution with an associated statistical weight. These experiments can be thought of as "proof-of-concept" detection of the minor allele in a mixture to determine whether it is comparable to CE-based methods. This is not dissimilar to DNA mixture experiments designed for and performed in an internal validation for CE-based methods (see Chapter 4).

Table 6.3. Examples of factor space covered in two STR sequencing assay evaluations using two-person mixtures of various mixture ratios and input DNA amounts (Fordyce et al. 2015, van der Gaag et al. 2016).

Fordyce et al. 2015			van der Gaag et al. 2016	
Mixture Ratio	**Input DNA**		**Mixture Ratio**	**Input DNA**
1000 : 1	10 ng : 10 pg			
100 : 1	5 ng : 50 pg		99 : 1	5.94 ng : 60 pg
50 : 1	5 ng : 100 pg			
20 : 1	2 ng : 100 pg		19 : 1	1.14 ng : 60 pg
10 : 1	1 ng : 100 pg		9 : 1	540 pg : 60 pg
5 : 1	1 ng : 200 pg		4 : 1	400 pg : 100 pg
2 : 1	1 ng : 500 pg			
1 : 1	500 pg : 500 pg		1 : 1	250 pg : 250 pg

Using the Ion Torrent NGS platform (the Ion PGM) and a 10-plex assay consisting of amelogenin and mostly simple STR loci (CSF1PO, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, and D16S539), a 2015 Danish study examined two-person mixtures with eight mixture ratios (Fordyce et al. 2015), as shown in Table 6.3. Mixtures were easily deconvoluted down to 20:1 for the vWA and D3S1358 STR markers, although some minor contributor alleles were not identified by the associated software and required manual allele calling. This work also described stutter artifacts as a challenge:

> "The main factor hindering mixture deconvolution down to 100:1 was the stutters corresponding to the major contributor alleles. Hence, if stutters could be reduced, perhaps with an optimized PCR and possibly improved software, then it should be possible to deconvolute mixtures down to 100:1" (Fordyce et al. 2015).

In 2016, a group from The Netherlands used the PowerSeq assay to examine 45 mixtures, which consisted of five, two-person mixtures at ratios shown in Table 6.3 (van der Gaag et al. 2016). We note that input DNA was inferred in our analysis and not explicit in the text. For all the mixtures at all 22 PowerSeq STR markers, the authors state each allele for both contributors was detected in the expected ratio. Alleles in overlapping or stutter positions were not included in this analysis. The authors conclude:

> "When analysing alleles with abundance below 5% of the highest allele of the locus, additional PCR/sequence error variants were observed for several loci which can complicate the interpretation of a DNA sample. Therefore, the analysis of minor contributions of 5% or less in a mixture without prior knowledge of the ratio between the different donors, remains difficult for some, but not all loci, using the current experimental and analysis setup for

> this assay. Increasing the sequencing coverage increases the read counts of these artefacts as well and will not help to distinguish them from genuine alleles" (van der Gaag et al. 2016).

Published NGS studies have focused on simple two-person mixture examples in an effort to count the number of minor alleles detected in the mixture (e.g., Jäger et al. 2017). This is often reported for non-overlapping alleles between samples in the mixture and provides a general indicator of the minor allele detection capability. Full minor profiles are commonly detected at about 9:1 ratio range with allele drop-out starting to occur at the 19:1 level and greater (e.g., Alonso et al. 2018), which is essentially equivalent to CE-based methods used currently.

The need for robust thresholds to enable confident allele calling (e.g., Riman et al. 2020) and a systematic framework to account for sequenced stutter artifacts is often recommended. Research in these areas is underway in the community (Zeng et al. 2017, Alonso et al. 2018, Vilsen et al. 2018a, Vilsen et al. 2018b, Riman et al. 2019a) and should enable progress toward the goal of sequence-based interpretation. To date, the research has been largely proof-of-concept, and less effort has been spent on assigning a likelihood ratio or conducting a statistical analysis of results (e.g., Chan Mun Wei et al. 2018). As our understanding of sequence noise and sequence-specific stutter are developed (e.g., Just & Irwin 2018), this information can assist future NGS-specific models for probabilistic genotyping. The ability to *detect* alleles in a mixture is not the same as exploring the *interpretation capabilities* of NGS. These types of studies are still needed to understand the levels of measurement and interpretation errors that might occur.

Additional autosomal STR markers have been evaluated to ascertain their value in mixture detection based on sequence variation. Dozens of new highly polymorphic STRs have been identified (Tan et al. 2017, Novroski et al. 2018). In addition, *in-silico* analysis of two-, three- four-, and five-person mixtures was performed to rank the best STR markers for distinguishing alleles, which improved the estimates of the number of contributors in a mixture (Young et al. 2019).

6.4.2. Alternate Markers

As described previously, the PCR amplification process for detecting STR alleles creates stutter product artifacts that interfere with unambiguous identification of minor contributors in an unbalanced mixture. Single nucleotide polymorphisms (SNPs) have been characterized for forensic use and explored to extend the capabilities of mixture interpretation. An important advantage of STR markers with mixture interpretation is the existence of many possible alleles within a population. This provides a greater chance of distinguishing multiple contributors from one another because of non-overlapping alleles compared to bi-allelic SNPs (Butler et al. 2007). SNPs and other alternative marker systems will be incompatible with existing STR databases. However, SNPs are amenable to array-based detection methods, which may be less expensive than STR sequencing for databasing single-source samples. While generally unsuitable for samples containing DNA mixtures, array-based SNP genotyping data can also be used for genetic genealogy searches (Greytak et al. 2019).

The ability to examine many more markers in parallel has become possible because of the new sequencing technologies described previously. For example, a Danish research group using the Ion Torrent NGS platform examined a 169plex SNP typing assay with 11 two-person mixtures with ratios ranging from 1:1 to 1:1000 (Børsting et al. 2014). They were able to observe all minor contributor SNP types in a 1:100 mixture when the overall number of reads was sufficiently high to cross a detection threshold for the minor allele. Maintaining a signal balance across all of the tested markers becomes challenging when more markers are examined. In a proof-of-concept study of a probe capture method for 451 target SNPs, the authors indicate an expected ability to detect 85% to 100% of alleles unique to the minor contributor with two-person male-male mixtures from 10 ng of total DNA template (Bose et al. 2018). This study observed allele drop-out when the minor contributor was approximately 10% or less (Bose et al. 2018).

The multi-allelic possibilities of microhaplotype (MH) markers, which are defined by two or more closely linked SNPs within a single PCR product (Figure 6.5), extend the possibilities for DNA mixture interpretation (Kidd et al. 2014). MH markers tend to be less polymorphic than STRs, so a greater number may be needed for identification purposes. An attractive aspect of using microhaplotype markers with DNA mixtures is the lack of stutter artifacts during PCR amplification. Although the absence of stutter artifacts should reduce the complexity of the interpretation, PCR-related issues still occur and need to be addressed as part of interpretation. This includes measuring rates of allele drop-out and defining minimum signal thresholds.

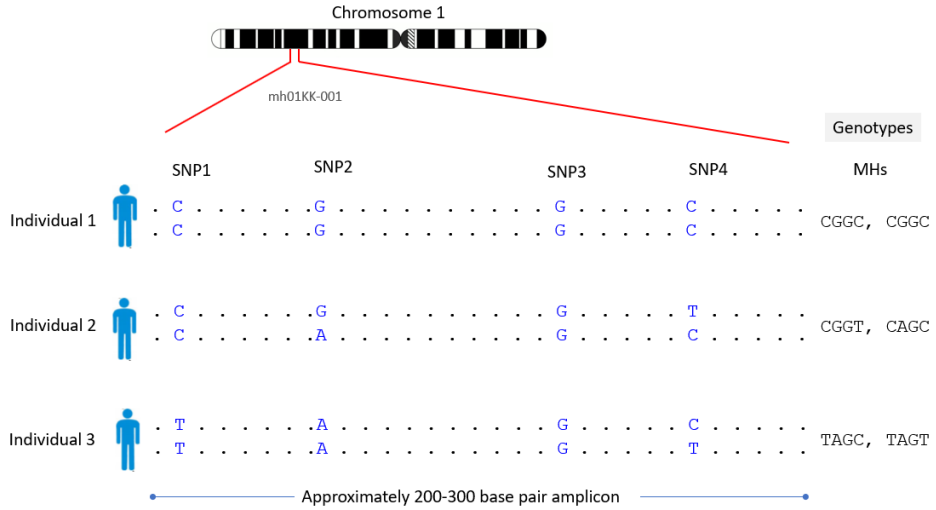

Figure 6.5. Schematic illustrating microhaplotypes in three individuals.

Different panels of microhaplotype markers have been developed by various research groups (e.g., van der Gaag et al. 2018, Chen et al. 2018, Voskoboinik et al. 2018, Bennett et al. 2019). Selection of standard MH markers and panels will be important as will more testing to explore the ability of these new markers to improve mixture interpretation in the future.

These research studies demonstrate the possibilities for new DNA markers to assist in mixture interpretation but will require much more extensive study before they can be

incorporated into laboratory workflows. Although these new markers may be free of PCR stutter artifacts, stochastic effects will still exist with PCR-based approaches. These stochastic effects, combined with overlapping alleles when there are multiple contributors, will continue to make DNA mixture interpretation challenging when small amounts of DNA are examined.

6.5. Summary and Key Takeaways

The ultimate decision to implement new technologies in forensic laboratories should be driven by a real-use case and by those responsible for producing and reporting the information. A vendor or members of the general public may encourage forensic DNA laboratories to adopt a new approach or technology without appreciating the investments required to make a change.

KEY TAKEAWAY #6.1: Fundamental measurement and interpretation issues surrounding DNA mixtures, as described in Chapter 2, should be understood before attempting to apply a new technology.

Consideration needs to be given to whether supporting factors and resources will be available upon implementation. This includes allele frequencies, analysis software, interpretation methods, training, and support for potential admissibility hearings.

KEY TAKEAWAY #6.2: Implementation requires a thorough understanding of the benefits and limitations of the new technology as well as the practical investment of time and effort put forth for its adoption by the laboratory.

An overall assessment is important and should include 1) how a new technology works, 2) what its limits are, and 3) how it might specifically address the problem to be solved. This assessment is a key component in evaluating whether implementation will be worthwhile.

Appendix 1: History of DNA Mixture Interpretation

Forensic DNA testing operates in an evolving environment with an increasingly complex set of technologies. Often important changes and advances have been made across the community because of experiences in high-profile court cases or awareness of issues raised through participation in interlaboratory studies or collaborative exercises. Several of these cases and studies are highlighted here. Histories of guidance documents and training courses related to DNA mixture interpretation are also described.

Forensic DNA analysis has undergone numerous changes in the three and a half decades since DNA methods were first applied to criminal investigations (Gill et al. 1985). An examination of the history of DNA mixture interpretation reveals an evolving set of technologies, DNA tests, and statistical approaches (e.g., Gill et al. 2015, Coble & Bright 2019). In recent years, many forensic laboratories have begun moving from "binary" approaches (i.e., is the genotype of interest present or not in the observed mixture?) to "probabilistic genotyping" methods (i.e., could the genotype of interest be present and, if so, how strongly does the data support this possibility?). This shift has occurred as techniques and approaches to mixture interpretation have evolved over time. As forensic DNA pioneer Peter Gill notes: "Interpretation of evidence continues to be the most difficult challenge that faces scientists, lawyers, and judges" (Gill 2019b).

This challenge comes in the face of change. Samples submitted to laboratories have changed from large visible stains to small invisible samples. Questions asked by the legal community have expanded from simply asking "to *whom* does the DNA belong?" to also asking "*how* did it get there?" The technology and marker sets have evolved from RFLP[30] to simple PCR assays to multiplex STRs with different commercial kits. Statistical approaches have changed in many jurisdictions e.g., from CPI to LRs. However, core principles underlying relevant and reliable DNA mixture interpretation remain the same (see Chapter 2 in this report).

A1.1. Early History of DNA Mixture Interpretation

A study of the early literature on DNA mixture interpretation is influenced by several authors, who are still active in the field. These authors include John Buckleton (Institute of Environmental Science and Research, New Zealand), Bruce Budowle (formerly at the FBI Laboratory and now at the University of North Texas Health Science Center), James Curran (University of Auckland, New Zealand), Ian Evett (formerly at the Forensic Science Service and now Principal Forensic Services, United Kingdom), Peter Gill (formerly at the Forensic Science Service, United Kingdom and now University of Oslo, Norway), and Bruce Weir (North Carolina State University and University of Washington).

A1.1.1. Early Method Development and Assessment of DNA Mixtures

Forensic DNA analysis began with restriction fragment length polymorphism (RFLP) techniques (Wyman & White 1980) and variable number of tandem repeat (VNTR) minisatellite probes (Jeffreys et al. 1985) that typically required hundreds of nanograms of

[30] Acronyms to be defined later in the document

DNA to obtain results. Sizable blood or semen stains were the most commonly examined evidence in initial forensic cases. Single-locus VNTR probes (Wong et al. 1987) quickly overtook the original multi-locus probes so that DNA mixtures could be more easily discerned (Kirby 1990, p. 140).

An early publication from Alec Jeffrey's laboratory at the University of Leicester in the UK claimed that autoradiograms of single-locus VNTR probes produced a signal "from 60 ng or less of human genomic DNA" and "depending on the genotypes of the individuals tested, … detect an admixture of 2% or less of one individual's DNA with another" (Wong et al. 1987). This same article notes: "locus-specific probes, unlike [multi-locus] DNA fingerprint probes, can be used to estimate the number of individuals represented in a mixed DNA sample" (Wong et al. 1987). In the late 1980s, in parallel with these developments in RFLP testing and its application to forensic analysis, a new technique helped improve DNA sensitivity. This involved generating millions of copies of targeted portions of each DNA molecule in a process known as the polymerase chain reaction (PCR).

PCR was originally developed in the mid-1980s (Saiki et al. 1985) and quickly became a valuable tool in molecular biology for examining small amounts of DNA. By the late 1980s and early 1990s, the first PCR methods were being implemented for forensic DNA testing purposes (Saiki et al. 1989, Blake et al. 1992). These initial methods were sensitive (i.e., detecting only a few cells' worth of DNA), but did not use highly polymorphic genetic markers (i.e., differentiating only a limited number of possible alleles and genotypes). Thus, these early tests were not extremely effective in distinguishing individual components of DNA mixtures. Many of these first PCR assays utilized single nucleotide polymorphisms (SNPs) that typically possess only two alleles (often generically designated "A" and "B") and thus, three genotypes ("AA," "AB," or "BB").

The first forensic PCR test involved the single-locus human leukocyte antigen (HLA) DQα with 6 possible alleles and 21 possible genotypes when examined with the AmpliType HLA DQα typing kit (Cetus Corporation, Emeryville, CA) using dot blot and reverse dot blot techniques (Walsh et al. 1991). A few years later, the AmpliType PM PCR Amplification and Typing Kit, which was developed by Roche Molecular Systems (Alameda, CA) and marketed by the PerkinElmer Corporation (Norwalk, CT), added five additional loci to the HLA DQα locus (Fildes & Reynolds 1995). These kits used either a "C" (control) dot or an "S" (sensitivity) dot "designed to be the lightest dot on the nylon strip and intended to act as a threshold for evaluating stochastic effects" (Budowle et al. 1995). According to the manufacturer, "the 'S' and the 'C' dots are designed not to be visible if the amount of template DNA is less than approximately 0.3 to 0.5 ng" (Fildes & Reynolds 1995).

In some of the earliest reported DNA mixture experiments, the FBI Laboratory performed validation experiments with the AmpliType DQα typing kit that involved two-person DNA mixtures, with combinations of non-overlapping heterozygous genotypes spanning ratios of 1:1 to 1,000:1 with DNA quantities in the 200 ng to 200 pg (0.2 ng) range (Comey & Budowle 1991). These authors note several limitations in the method used including (a) that mixture ratios appeared to matter more than the overall quantity of DNA in terms of dot

intensity and (b) that shared alleles between contributor genotypes could prevent mixture detection with a single-locus system exhibiting a limited number of possible alleles.

When the first multiplex PCR kit became available, a publication containing FBI validation studies of the AmpliType PM (PolyMarker) kit discussed the ability to detect mixed body-fluid samples created by combining saliva and semen:

> "the presence of two or more contributors to a sample generally is inferred by the presence of unbalanced dots and/or … extra dots in [two of the loci which were tri-allelic SNPs]" (Budowle et al. 1995).

However, these authors also note:

> "the exact percentage of samples that exhibit unbalanced allele dot intensities is difficult to determine, because the determination of unbalanced intensity is somewhat subjective." This study found that "the minor contributor of a 1:20 mixture of two samples was barely detectable, and the allele dot for the minor component was less intense than the S dot" (Budowle et al. 1995).

This study concludes:

> "Because of the potential for unbalanced allele dot intensities and the limitations for detecting some mixed samples containing equivalent amounts of DNA, caution should be exercised when interpreting evidentiary samples that potentially may be from more than one donor" (Budowle et al. 1995).

Thus, the FBI alerted specialists of the challenges posed by multi-donor samples.

A study involving seven laboratories, organized by the manufacturer of the AmpliType PM PCR Amplification and Typing Kit, was published about the same time as the FBI study detailed above. The publication described the kit's ability with mixture detection a little differently than the FBI researchers. Authors of the study wrote:

> "The balance of dots within a locus of the PM DNA probe strip proved to be a valuable asset of the system for the analysis of mixtures. This feature is an important benefit of the PM system since a high percentage of forensic casework involves the analysis of sexual assault samples" (Fildes & Reynolds 1995).

The differences in perspectives highlighted here illustrate that sometimes a disconnect can exist between researchers and commercial suppliers in the types of studies performed and the language used in sharing their results. For example, compare

> "because of the potential for unbalanced allele dot intensities, … caution should be exercised" (Budowle et al. 1995)

versus

> "the balance of dots … proved to be a valuable asset of the system for the analysis of mixtures" (Fildes & Reynolds 1995).

This observation exemplifies the reasoning of the President's Council of Advisors on Science and Technology (PCAST), who wrote in their 2016 report:

> "While it is completely appropriate for method developers to evaluate their own methods, establishing scientific validity also requires scientific evaluation by other scientific groups that did not develop the method" (page 80 of PCAST 2016).

In the early to mid-1990s when the AmpliType PM kit was used, most DNA mixtures seen in forensic laboratories derived from "incomplete separation of the sperm and female epithelial cell fractions from postcoital swab extractions" (Fildes & Reynolds 1995; see Box A1.1). After reviewing the field trial results from seven forensic laboratories, the authors of this study commented:

> "The potential for sample mixtures in forensic casework analysis has always required careful and thoughtful interpretation. Individual laboratories will need to develop their own policies for the interpretation of mixtures based on their experience and case history information" (Fildes & Reynolds 1995).

Box A1.1. Differential Extraction

Many sexual samples, particularly those coming from vaginal swabs collected from a sexual assault victim, typically contain DNA from both the victim and the perpetrator. In the 1985 *Nature* article that launched forensic DNA analysis, authors Peter Gill and David Werrett from the UK Forensic Science Service and Alec Jeffreys from the University of Leicester introduced differential extraction as a method to separate the perpetrator's sperm cells from the victim's epithelial cells based on the chemical composition of the sperm head (Gill et al. 1985). When DNA mixtures cannot be resolved into single-source components through techniques such as differential extraction, then mixture interpretation is required.

The developers of these early PCR test kits encouraged users to avoid interpreting low levels of DNA (i.e., attempting to interpret results below their "C" or "S" dots) to avoid problems with unbalanced allele detection. In a 1992 article, they note:

> "Preferential amplification due to stochastic fluctuation can occur when amplifying very low amounts of target DNA molecules; the possibility of an unequal sampling of the two alleles of a heterozygote is increased when only a few DNA molecules are used to initiate PCR. This problem can be avoided by adjusting the cycle number such that approximately 20 or more copies of target DNA [i.e., >120 pg genomic DNA assuming 6 pg per diploid copy of the genome] are required to give a typing result for that PCR system" (Walsh et al. 1992).

For the first decade of DNA testing (circa 1985 to 1995), where many nanograms of DNA were required to obtain a result, most of the samples examined involved visible bloodstains or sexual assault evidence. This meant that only a limited number of mixtures were observed in casework during the 1990s. For example, a review of DNA casework in a Spanish laboratory from 1997 through 2000 reported observing less than 7% mixture profiles (Torres et al. 2003). If mixtures were observed, they were often treated as "uninterpretable" (e.g., Fildes & Reynolds 1995).

By the mid-1990s, the field began to move towards multi-allelic short tandem repeat (STR) markers where multiple STR loci could be co-amplified and labeled using multiplex PCR (Caskey et al. 1989, Edwards et al. 1991, Frégeau & Fourney 1993, Kimpton et al. 1993). STR markers benefit mixture interpretation from the existence of sometimes a dozen or more alleles per marker compared to two and sometimes three alleles present in SNP loci (Butler et al. 2007), such as used in the AmpliType PM kit. In the 1990s, the UK Home Office's Forensic Science Service (FSS) led the forensic community in advancing knowledge of STR

markers and their application to forensic science including DNA mixture interpretation (Gill et al. 1995, Gill et al. 1997, Clayton et al. 1998, Gill et al. 1998, Evett et al. 1998).

Efforts were also made to extend interpretation of STR typing results to DNA quantities originating from less than approximately 20 cells (≈120 pg) (Gill et al. 2000) – a limit that had previously been recommended to avoid stochastic effects (Walsh et al. 1992). Commercial STR kits, either from Promega Corporation (Madison, WI) or Applied Biosystems[31] (previously Foster City and now South San Francisco, CA), have been widely used since the late 1990s to enable forensic DNA testing. More recently, Qiagen (Hilden, Germany) has begun offering STR typing kits.

A1.1.2. Initial Interpretation Approaches Explored for DNA Mixtures

The presence of a mixture can be identified by the observation of more than two alleles at an STR locus. Also, the occurrence of more than two alleles will typically be seen at two or more loci in the DNA profile for almost all mixtures. Exceptions exist for any rule though. Occasionally tri-allelic patterns have been reported at one STR locus in a single-source DNA profile (e.g., Clayton et al. 2004). Artifacts, such as stutter products created due to strand slippage during PCR amplification of STR markers (see chapter 3 in Butler 2015a), can give rise to additional DNA peaks and increase the complexity and challenge of mixture interpretation. For this reason, guidelines have been developed and refined over the past several decades to assist in designating STR alleles versus artifacts and interpreting DNA profiles (Gill et al. 1997, SWGDAM 2000, SWGDAM 2010, SWGDAM 2017a).

In some of the first articles describing mixture interpretation with STR markers, Peter Gill and his FSS colleagues noted the need to understand heterozygote peak balance within each locus to conduct mixture analysis (Gill et al. 1995, Gill et al. 1997). They point out that "interpretation of mixtures also needs to take account of the possible confusion between a true mixture and the presence of stutter bands" (Gill et al. 1995), which was described in more detail as part of the International Society for Forensic Genetics (ISFG) DNA Commission recommendations about a decade later (Gill et al. 2006b). Based on their observations with a 6-locus STR multiplex in use at the time, these FSS researchers share:

> "If the mixture [has components in the ratio of] 1:5 then reliable identification of the components of a [two-person] mixture is normally possible" (Gill et al. 1995).

They continue:

> "When mixtures are observed, and the components cannot be separated, there will inevitably be occasions when it will be more appropriate to present all the possible alternatives using statistical methods described by Evett et al. [Evett et al. 1991]" (Gill et al. 1995).

DNA mixture interpretation considers possible genotype combinations that could create the observed data. Different statistical approaches have been used to describe mixture results (Box A1.2).

[31] Applied Biosystems has undergone multiple name changes over the years and in 2019 is known as Thermo Fisher Scientific (for ten names spanning 1981 to 2014, see Butler 2015a, p. 26).

In 1991, Ian Evett of the FSS and several colleagues introduced a likelihood ratio (LR) approach (Evett et al. 1991). In this initial mixture interpretation article, which uses examples from RFLP single-locus probes available at the time, the authors note:

> "This paper has been *restricted to fairly simple case situations*; as the number of bands increases the evaluation is liable to become quite complicated. Also, it is important for caseworkers to recognize that *the evidential strength falls rapidly with increasing number of bands*…" (Evett et al. 1991, emphasis added).

The authors also observe:

> "In some cases, where there are unequal band intensities, it may be possible to determine which bands are paired. Thus, two very strong bands might be clearly distinguished from two weak ones. However, this would have to be considered carefully because there can be differences in intensities between the two bands from one individual…" (Evett et al. 1991).

Thus, this initial article using an LR approach recognizes the challenge of accounting for an increasing number of alleles as the number of contributors goes up. Furthermore, the article emphasizes that reliable allele pairing into contributor genotypes may be difficult and needs "to be considered carefully" because of the natural variation in heterozygote allele balance, which increases with lower amounts of starting DNA template.

Box A1.2. Statistical Approaches Used for DNA Mixture Interpretation
(as defined by SWGDAM 2017a)

RMP (random match probability): the probability of randomly selecting from the population an unrelated individual who could be a potential contributor to an evidentiary profile

CPI (combined probability of inclusion): produced by multiplying the probabilities of inclusion from each locus; probability of inclusion is the percentage of the population that can be included as potential contributors to a DNA mixture at a given locus; also known as Random Man Not Excluded (RMNE)

LR (likelihood ratio): the ratio of two probabilities of the same event under different and mutually exclusive hypotheses; typically, the numerator contains the prosecution's hypothesis and the denominator the defense's hypothesis

A1.1.3. Early U.S. Mixture Approaches – The NRC I and NRC II Reports

While LR approaches for mixtures were under development in the UK in the early 1990s, the National Research Council (NRC) in the United States completed a report in 1992 that mentions the combined probability of inclusion (CPI) as an appropriate method for mixture interpretation:

> "If the samples are mixtures from more than one person, one should see additional bands for all or most polymorphic probes, but not for a single-copy monomorphic probe. Mixed samples can be very difficult to interpret, because the components can be present in different quantities and states of degradation. It is important to examine the results of multiple RFLPs, as a consistency check. *Typically, it will be impossible to distinguish the individual genotypes of each contributor*. If a suspect's pattern is found within the mixed pattern, the appropriate frequency to assign such a 'match' is the sum of the frequencies of all genotypes that are contained within (i.e., that are a subset of) the mixed pattern" (page 59 of NRC 1992, emphasis added).

Thus, this early report recognizes some of the difficulties in mixture interpretation including distinguishing contributor genotypes when components vary in quantity and quality.

The NRC 1992 report emphasizes the following five principles: (1) that polymorphic loci containing many possible alleles enable mixtures to be more easily detected, (2) mixtures are complicated by the ratio of contributors and their possible states of degradation, (3) checking the consistency of the mixture across multiple loci aids quality assurance, (4) distinguishing the individual genotypes of each contributor is not always possible, and (5) when individual contributor genotypes cannot be distinguished, the CPI statistic should be used, which involves summing the frequencies of all genotypes that are contained with the mixed pattern.

It is important to note that at the time the first NRC report was written, high-quantities of DNA were needed to obtain an RFLP result and therefore the possibility of allele drop-out was not considered an issue. As emphasized in a more recent publication (Bieber et al. 2016), the CPI statistic is only fit-for-purpose at a tested locus if all alleles of all contributors present are detected in the DNA mixture. In other words, the CPI statistic cannot be applied to DNA mixture profiles with potential allele drop-out because it would not fully account for all possible genotypes. Therefore, the CPI statistic is not suitable for use with DNA mixture profiles containing low levels of DNA.

A second NRC report published in 1996 (known as the NRC II, NRC 1996), was intended to replace the 1992 report. The NRC II report observes:

> "In many cases, one of the contributors – for example, the victim – is known, and the genetic profile of the unknown is readily inferred. In some cases, it might be possible to distinguish the genetic profiles of the contributors to a mixture from differences in intensities of bands in an RFLP pattern or dots in a dot-blot typing; in either case, the analysis is similar to the unmixed case. However, when the contributors to a mixture are not known or cannot otherwise be distinguished, a likelihood-ratio approach offers a clear advantage and is particularly suitable" (pages 129 and 130, NRC 1996).

The report references a simple RFLP case example in which there are four distinguishable alleles coming from two individuals – and the CPI calculation is performed as recommended from the 1992 NRC report, page 59.

The NRC II report continues:

> "That [CPI] calculation is hard to justify because it does not make use of some of the information available, namely, the genotype of the suspect. The correct procedure, we believe, was described by Evett et al. (1991)" (page 130, NRC 1996).

After working through this example, the NRC II report notes:

> "We *have considered only simple cases*. With VNTRs, it is possible, though very unlikely, that the four bands were contributed by more than two persons, who either were homozygous or shared rare alleles. *With multiple loci, it will usually be evident if the sample was contributed by more than two persons*. Calculations taking those possibilities into account could be made if there were reason to believe that more than two persons contributed to the sample. Mixed samples are often difficult to analyze in systems where several loci are analyzed at once…. The problem is complex, and some forensic experts follow the practice of making several reasonable assumptions

> and then using the calculation that is most conservative. For a fuller treatment of mixed samples, see [Weir et al. (1997)]" (NRC 1996, emphasis added).

This report discusses the benefits of an LR approach with considering the suspect's genotype in the context of the case and notes that multiple assumptions and calculations may be needed particularly when going beyond simple cases.

About a decade after the NRC II report was released, an article was written discussing the merits of CPI and LR approaches (Buckleton & Curran 2008). The authors noted that LR results must assume a number of contributors and are more difficult to present in court. On the other hand, CPI (RMNE) statistics waste information and cannot be interpreted directly in the context of a court case.

A1.2. First High-Profile Case with DNA Mixtures

In June 1994, U.S. football star O.J. Simpson was accused of murdering his ex-wife Nicole Brown and her friend Ronald Goldman. The trial was televised and became a worldwide event with DNA evidence playing a prominent role in the trial (Weir 1995). Decisions during the O.J. Simpson case may well have impacted the early trajectory of mixture interpretation in U.S. courts and forensic laboratories (Box A1.3).

Box A1.3. Impact of O.J. Simpson Trial on U.S. Approach to DNA Mixtures

Experiences from the O.J. Simpson "trial-of-the century" in 1995 brought "the nature and strength of DNA evidence to wide public notice" (Weir 1995) and aided adoption of quality control measures in forensic DNA laboratories (see Butler 2009, pp. 84-85). Due to concerns raised during the trial, procedures for biological evidence collection and storage in many forensic laboratories were improved going forward. Within a few years, the FBI's DNA Advisory Board created the initial Quality Assurance Standards (QAS), which have been widely used and revised several times since then.

The O.J. Simpson trial had another impact that is perhaps not as well appreciated as the quality assurance improvements that were put in place around the country. Prior to this case in 1995, "no U.S. court had ever heard statistical testimony concerning mixtures" (Weir 2000). Towards the end of the trial, when reviewing statistics for DNA mixtures involved, Judge Ito denied the admissibility of likelihood ratio (LR) calculations performed (Kaye 2010). Reliance on the NRC I 1992 report apparently influenced this decision (Weir 1995), which may very well have delayed wider adoption of the LR approach in the United States for many years (Weir 2000) even though the NRC II 1996 report would be supportive of LRs (NRC 1996, p. 130). The trial experience did have the benefit of renewing the interest of some members of the community to work further on improving interpretation of mixtures (e.g., Weir et al. 1997, Buckleton et al. 1998) and generating the first book on the topic (Evett & Weir 1998).

Shortly after he appeared as a prosecution witness in the Simpson trial, Professor Bruce Weir, then at North Carolina State University, wrote:

> "Reliance on the [1992] NRC report prevented an adequate treatment of mixtures and population structure in the Simpson trial" (Weir 1995).

He continues:

> "It is incumbent on both prosecution and defense to explain the meaning of a conditional probability of a DNA profile… Simple frequencies do not address the issue of mixtures. When there are several contributors to a bloodstain (a mixed stain), the probability calculations can become quite complex…" (Weir 1995).

To improve mixture interpretation and remove some of the misconceptions that arose during the O.J. Simpson trial, Professor Weir and his collaborators began research that enabled the field to move forward in significant ways with DNA mixture interpretation.

A1.3. Development of LR Methods

In March 1997, Professor Bruce Weir and colleagues from New Zealand published an article titled "Interpreting DNA Mixtures" in the *Journal of Forensic Sciences* that described LR calculations with two-person mixtures based on assuming independence of alleles within and between loci (Weir et al. 1997). An example was even worked from an RFLP mixture result in the O.J. Simpson case using the "2p" rule. This rule had been introduced in the NRC II report for single-banded VNTR loci used in RFLP but declared inappropriate for PCR-based systems (see NRC 1996, p. 5). However, the authors note:

> "The '2p' rule is not always conservative, and we suggest caution in its use" (Weir et al. 1997).

Commenting on the value of LR calculations compared to the CPI approach, Professor Weir and colleagues state:

> "Interpretations based simply on the frequencies with which random members of a population would not be excluded from a mixed-stain profile [i.e., CPI] do not make use of all the information, and may overstate the strength of the evidence against included people," and they emphasize "only by comparing the probabilities of the evidentiary profile under alternative explanations [i.e., using LRs] is it possible to arrive at a complete analysis of mixtures" (Weir et al. 1997).

Thus, from the very beginning of mixture interpretation efforts, LR methods were emphasized as being superior to CPI calculations.

An important aspect of LR methods involves the number of potential contributors. Weir wrote:

> "the [LR] results given so far depend on the number of contributors to the mixed sample" (Weir et al. 1997).

Referring to an article from Charles Brenner, Rolf Fimmers, and Max Baur (the latter two of whom are German mathematicians) (Brenner et al. 1996), Professor Weir and colleagues note:

> "Whenever there is doubt as to the number of contributors, there can be considerable variation in the likelihood ratio." (Weir et al. 1997).

Using the formulas outlined in this initial article (Weir et al. 1997) and a follow-up one allowing for population sub-structure (Curran et al. 1999), a software program named *DNAMIX* was developed (Storey & Weir 1998). It is important to keep in mind that *DNAMIX* was built at a time when fairly high-levels of DNA were being tested and was not designed to account for the possibility of allele drop-out.

A1.4. Mixture Deconvolution

The UK Forensic Science Service, using in-house developed STR assays, published several landmark articles on mixture interpretation in 1998. This included approaches to using peak heights/areas to enable mixture deconvolution with simple two-person mixtures (Clayton et al. 1998, Evett et al. 1998, Gill et al. 1998). These articles are foundational and a valuable addition to training programs in DNA mixture interpretation.

In their article in which allele peak areas were used to interpret simple STR mixtures, FSS researchers examined 39 different mixtures prepared from five different individuals with mixture ratios ranging from 1:10 to 10:1 (Gill et al. 1998). They use a generic "mixture code" that enabled classifying mixture groups with similar levels of allele sharing while at the same time avoiding the need to list specific genotypes that could impact the privacy of donors.

This article also notes that with lower quality data, there was a poor fit to the model and so the correct genotype did not rank as well, and they suggest "caution should be exercised with low peaks" (Gill et al. 1998). In addition, this is the first attempt to define a "complex mixture" as a profile containing "more than four alleles at any locus" – and the authors note that their method does not apply to these complex mixtures (Gill et al. 1998).

Three important points and principles highlighted in this foundational article include:

(1) The lower the peak heights, the higher the variability in relative peak heights due to stochastic variation in PCR amplification of the mixture components. The report noted:

> "if the peak areas [or heights] are low, then the relative peak areas [or heights] become less predictable for a given mixture."

In other words, the variability and uncertainty in relative peak heights increases as overall peak heights decrease.

(2) The reproducibility of mixture results and relative peak heights of mixture components should be verified through repeated testing if there is sufficient DNA available. The authors of the article state:

> "it is important to repeat the experiment – possibly at a higher concentration of DNA."

(3) Reducing the number of loci, simultaneously amplified, improves the relative peak balance. The authors write:

> "Singleplex analysis (where just a single locus is amplified) is another option, to improve the signal strength" and "also improve the relative peak balance, so that peak areas better reflect the actual DNA concentration" (Gill et al. 1998).

The first commercial STR kits were becoming available at the same time that the FSS was sharing their mixture interpretation results. Applied Biosystems followed Technical Working Group on DNA Analysis Methods (TWGDAM) guidance when validating their first STR kit "AmpFlSTR Blue," which was a triplex amplifying DNA markers D3S1358, vWA, and FGA

(developmental validation published in Wallin et al. 1998). From these studies with two-person mixtures, which explored ratios of 1:1 to 1:50 at 1 ng or 5 ng total DNA template, they concluded:

> "The limit of detection for mixtures in which a total of 1 ng was amplified was 1:10" while "the limit of detection of mixtures in which a total of 5 ng was amplified was 1:30" (Wallin et al. 1998).

Therefore, the limit of detection for the minor contributor was in the range of 100 pg to 160 pg. These authors summarized:

> "Taken together, these mixture studies indicate that it is possible to detect a mixture and sometimes resolve the genotypes of each contributor, depending on the genomic DNA ratios, number of contributors, and particular combination of alleles present" (Wallin et al. 1998).

Experience gained from these early studies, as well as the increasing sensitivity of DNA tests (Gill et al. 2000) that quickly followed, would lead to the first international recommendations on DNA mixture interpretation (Gill et al. 2006b). Software programs were also developed to assist with mixture deconvolution (e.g., Bill et al. 2005, Wang et al. 2006).

A1.5. Increased Sensitivity in DNA Test Methods

The PCR method can be tuned to amplify and recover low quantities of DNA through increasing the number of amplification cycles or amount of PCR product injected into a genetic analyzer (see Butler 2012, pp. 311-346). As early as 1997, researchers demonstrated that with such tuning STR typing results could be obtained from DNA found in a single cell (Findlay et al. 1997). This capability encouraged attempts to recover DNA profiles from invisible samples left by touching a surface. Some laboratories, rather than using a specific enhanced detection method such as an increased number of PCR cycles, pushed the limits of their existing protocols by expanding their sampling approaches to include smaller and smaller quantities of biological material.

While information from invisible samples (sometimes called "touch evidence" or "trace DNA") can be helpful in an investigation, increasing the sensitivity of the PCR method to obtain results from invisible samples can impact reliability and relevance. From a historical perspective, this increase in DNA test method sensitivity and willingness to attempt examination of smaller quantities of DNA have resulted in an increase in samples and sample types submitted to forensic laboratories. This has led to more mixtures being observed, and to development of modern interpretation techniques discussed in Section A1.6.

A1.5.1. Low Copy Number (LCN) Method

As leaders in developing and implementing forensic DNA methods during the 1990s and early 2000s, the UK's Forensic Science Service ventured into increased sensitivity (Findlay et al. 1997, Gill et al. 2000) and new approaches for interpretation of evidence (Gill et al. 2007). The FSS method was initially referred to as low copy number (LCN) DNA testing and later as low template DNA (LT-DNA). The original FSS LCN method involved an in-house 6-plex STR assay or a commercial STR kit 10-plex amplified with 34 cycles (Gill et al. 2000,

Whitaker et al. 2001) rather than the traditional 28 cycles widely used at the time (e.g., Wallin et al. 1998).

In the foundational article "An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA," the authors report:

> "By increasing the PCR amplification regime to 34 cycles, we have demonstrated that it is possible routinely to analyze <100 pg DNA [representing around 15 diploid cells]…Compared to amplification of 1 ng DNA at 28 cycles, it was shown that increased imbalance of heterozygotes occurred, along with an increase in the size (peak area) of stutters. *The analysis of mixtures by peak area [or height] measurement becomes increasingly difficult*…Laboratory-based contamination cannot be completely avoided, even when analysis is carried out under stringent conditions of cleanliness…" (Gill et al. 2000, emphasis added).

Attempts to recover information from low amounts of DNA present in evidentiary samples using LCN methods inevitably led to increased imbalance in heterozygotes, higher levels of stutter products, allele drop-out, and allele drop-in (contamination). These phenomena are artifacts of stochastic, or random sampling, effects that occur in the early cycles of PCR amplification when there are a limited number of target molecules to amplify (Butler & Hill 2010).

To alleviate stochastic effects, LCN protocols typically involve forming consensus profiles using replicate amplifications from aliquots of a DNA extract. Alleles that are observed in replicate amplifications are deemed "reliable" and form a consensus profile (e.g., Benschop et al. 2011). While attempting to replicate alleles from single-source samples is straightforward, the replicate tests are unlikely to maintain relative peak height ratios and mixture ratios needed for traditional DNA mixture deconvolution and interpretation. It was recognized in the seminal LCN article that "these guidelines [for replicate testing and building consensus profiles] will be superseded by expert systems utilizing the Bayesian principles described in this paper" (Gill et al. 2000). Such expert systems would not be available for almost another decade.

A1.5.2. Reliability Concerns with Increased Sensitivity

A judge's ruling in the Omagh (Northern Ireland) terrorist bombing trial in 2007 raised concerns about the reliability of the FSS LCN method used in the case. This ruling, in turn, led to a formal investigation of LCN and the creation of the UK Forensic Science Regulator, which monitored quality assurance as well as some intense debates in several scientific meetings (see Butler 2012, pp. 313-319). Concerns regarding the use of LCN in criminal casework had been raised previously (e.g., Budowle et al. 2001), but this case revived scrutiny.

In addition, there was increased recognition of the challenges that higher-sensitivity DNA results brought to DNA mixture interpretation. For example, a December 2007 article states:

> "With the improved sensitivity of modern DNA methods coupled with the increased use of forensic genetics in crime case investigations, the number of DNA mixtures

> composed of full or partial DNA profiles from two or more contributors has increased dramatically" (Morling et al. 2007).

The authors continue:

> "The biostatistical interpretation of mixed DNA profiles is a challenge – especially if DNA profiles are incomplete" (Morling et al. 2007).

A1.5.3. Relevance Concerns with Increased Sensitivity

Along with the aforementioned sensitivity efforts, it was recognized early on that low levels of DNA template on items or surfaces might not be associated with the crime, but rather left innocently before the crime occurred (Gill 2001). Secondary or tertiary transfer of DNA due to casual contact, such as hand shaking, has been shown to vary. This variance is based on what has been termed the "shedder" status of the individuals involved (Lowe et al. 2002). Even as far back as 1997 in a landmark study in the journal *Nature,* "DNA fingerprints from fingerprints" (van Oorschot & Jones 1997) discussed the possibility of DNA transfer (see Chapter 5 in this report).

A 2013 review article "DNA transfer: review and implications for casework" increased awareness of relevance concerns with "trace DNA," which the authors termed "DNA that cannot be attributed to an identifiable body fluid" (Meakin & Jamieson 2013). When DNA cannot be attributed to an identifiable body fluid, it can no longer address source level questions on the hierarchy of propositions (offense, activity, and source levels), which were outlined by the FSS in 1998 (Cook et al. 1998b). Thus, sub-source (Gill 2001) and even sub-sub-source levels (Taylor et al. 2014, Taylor et al. 2018) become part of DNA mixture interpretation considerations. As discussed in Chapter 5 of this report, there is still a lot to learn in this area and many gaps remain to be filled (e.g., Burrill et al. 2019, van Oorschot et al. 2019). Activity-level propositions have been suggested as the most appropriate approach to dealing with small quantities of DNA detected due to increased sensitivity of DNA tests. In some cases, there has been a shift in focus by the court from questions about the source of the DNA to the mechanism by which it was deposited (Taylor et al. 2018, Gill et al. 2020a).

A1.6. Probabilistic Genotyping Software (PGS)

As techniques for obtaining DNA results from low amounts of DNA template were implemented around the turn of the century (e.g., Gill et al. 2000) and laboratories began expanding the sample types they were willing to attempt to analyze, dealing with the possibility of allele drop-out and missing information from DNA mixture profiles became important. This led to thinking probabilistically about DNA data (e.g., Balding & Buckleton 2009, Kelly et al. 2014) and the development of probabilistic genotyping software (PGS) systems.

A1.6.1. Development of PGS

In the late 1990s, the UK Forensic Science Service proposed the use of computer programs to assist in DNA mixture interpretation (Evett et al. 1998) and developed the initial theory for probabilistic genotyping. This theory incorporated the probability of drop-out when examining low quantities of DNA (Gill et al. 2000). At this same time, Cybergenetics (Pittsburgh, PA) was developing computer software to aid DNA mixture interpretation – first with automated

methods to cope with stutter products (Perlin et al. 1995) and then with mathematics to assist in deconvolution of mixture components (Perlin & Szabady 2001). U.S. patents on using a computer to calculate a likelihood ratio from a DNA mixture, which were filed as early as 2001, have been awarded to Cybergenetics and its TrueAllele software (e.g., Perlin 2017).

The FSS simulated and modeled each of the steps in the DNA analysis and interpretation process (Gill et al. 2005) and created the LoComatioN software (Gill et al. 2007) to assist with allele drop-out, which regularly occurs when examining low amounts of DNA template (Balding & Buckleton 2009). In addition, non-contributor assessments to explore the performance of probabilistic models were advocated (e.g., Gill & Haned 2013). As explained in further detail in Chapter 2 of this report, PGS systems are either (1) *discrete* (also called semi-continuous) if only alleles are considered or (2) *continuous* (also called fully-continuous) if peak height information is utilized (see Kelly et al. 2014).

Aspects of this FSS work were implemented in the LiRa system by former members of the FSS (Puch-Solis & Clayton 2014). David Balding also developed likeLTD (Balding 2013) which forms the basis of Lab Retriever (Inman et al. 2015). With European Union funding and an open-source software initiative, LRmix (Haned et al. 2012, Prieto et al. 2014) and EuroForMix (Bleka et al. 2016a) were developed and tested.

STRmix was developed by Duncan Taylor in South Australia and John Buckleton and Jo-Anne Bright in New Zealand (Taylor et al. 2013). STRmix was implemented in forensic laboratories across Australia and New Zealand in late 2012[32] and international sales began in early 2014. Developmental validation, which followed the SWGDAM 2015 guidelines (SWGDAM 2015), was published two years later (Bright et al. 2016).

A1.6.2. Movement to PGS in the United States

An increased awareness of the benefits of PGS for interpreting complex mixtures came at the NIST-FBI DNA Technical Leaders' Summit held in Norman, Oklahoma in November 2013 (see Table 6.5 in Butler 2015a), where more than 95% of public U.S. forensic laboratories were represented. The following June, a weeklong PGS workshop in St. Louis, Missouri sponsored by the Midwestern Association of Forensic Scientists (MAFS) informed attendees regarding the various software programs and their capabilities. At MAFS, vendors were provided an opportunity to demonstrate their PGS systems and answer questions. Concurrently, SWGDAM was drafting Guidelines for the Validation of Probabilistic Genotyping Systems, published the following year (SWGDAM 2015).

The first PGS publications in the U.S. came from Mark Perlin of Cybergenetics, demonstrating his fully-continuous TrueAllele Casework software; several of these articles were written in collaboration with scientists from the New York State Police (NYSP) or the Virginia Department of Forensic Science (VDFS) (Perlin et al. 2009, Perlin & Sinelnikov 2009, Perlin et al. 2011, Perlin et al. 2013, Perlin et al. 2014). The performance of TrueAllele Casework with two-, three-, and four-person mixtures were also explored by VDFS (Greenspoon et al. 2015) and results with five-person mixtures were described in another study involving the Kern County (California) Regional Crime Laboratory (Perlin et al. 2015).

[32] https://johnbuckleton.wordpress.com/strmix/

A discrete PGS system known as the Forensic Statistical Tool (FST) was developed in-house by the New York City Office of Chief Medical Examiner (OCME) to account for the possibility of allele drop-out and drop-in when testing low amounts of single-source and mixed DNA samples (Mitchell et al. 2011, Mitchell et al. 2012). OCME began using FST in forensic casework in April 2011 (Mitchell et al. 2011).

While FST was being developed and implemented in New York City, another discrete PGS system named Lab Retriever was created in California. Instead of a proprietary, in-house program like FST, Lab Retriever is an open-source, freely available program to calculate likelihood ratios for complex DNA profiles (Inman et al. 2015). This program is based on David Balding's likeLTD discrete PGS system (Balding 2013, Lohmueller & Rudin 2013).

STRmix has been used in the U.S. since 2014. According to information on the website of one of the STRmix developers[33], early U.S. adopters of STRmix included the United States Army Criminal Investigation Laboratory (USACIL) in November 2014, Erie County (NY) in July 2015, San Diego Police Department in October 2015, and the FBI Laboratory in December 2015. The FBI and the STRmix developers co-published a summary of the FBI internal validation studies of STRmix in a peer-reviewed journal (Moretti et al. 2017). Data from many early adopters of STRmix were also compiled and published, representing results for 2825 mixtures from 31 laboratories (Bright et al. 2018).

STRmix developers and colleagues have published their perspectives on the utility and validity of their PGS system. They conclude: "The efforts to bring [probabilistic genotyping] to fruition, including the initial theoretical development for human identification applications based on STR typing, span almost two decades, and thus its use today should not be misconstrued as some sudden novel technology" (Buckleton et al. 2019).

A1.6.3. FTCOE 2015 Landscape Study of PGS Systems

Given the growing interest in PGS systems among U.S. forensic laboratories, the National Institute of Justice (NIJ) funded a study to examine them. In July 2015, the NIJ Forensic Technology Center of Excellence (FTCOE) published a 45-page "Landscape Study of DNA Mixture Interpretation Software" (FTCOE 2015). This report explored the stated capabilities and limitations of 13 DNA mixture interpretation software tools available at the time: two with binary interpretation models (ArmedXpert and GeneMarker HID), six using discrete models (FST, GenoProof Mixture, Lab Retriever, LikeLTD, LiRa, and LRmix Studio), and five incorporating continuous models (DNAmixtures, DNA View Mixture Solution, LiRaHT, STRmix, and TrueAllele).

For each of these 13 software tools, the FTCOE assessment examined availability (commercial, proprietary, or open-source); the developer; statistical approaches utilized (RMP, CPI, LR); input data required (.fsa or .hid files, csv, text file); maximum number of unknown contributors that could be evaluated; whether training resources (yes/no), technical support (none, basic, extensive), or testimony support (yes/no) were available; whether CODIS output was possible (yes/no); whether a database could be queried (yes/no); whether

[33] https://johnbuckleton.wordpress.com/strmix/

Markov chain Monte Carlo (MCMC) simulations were performed (yes/no); whether the software could account for possible relatedness (yes/no); and frequency of system updates.

Since the 2015 study, there have been a few updates and additions to the PGS marketplace. PGS systems known to exist as of July 2019 are listed in a recent review article (Butler & Willis 2020, see also Coble & Bright 2019). Published direct comparisons of PGS systems are fairly limited as discussed in Chapter 4 of this report.

A1.7. Sources of Guidance on DNA Mixture Interpretation and Validation

Accredited laboratories follow written protocols and are regularly audited to assess their conformance to these protocols and compliance with applicable standards. Multiple advisory groups have provided recommendations on quality assurance measures and helpful validation studies to assess the capabilities and limitations of DNA mixture interpretation approaches (Butler 2013).

Numerous documents exist that provide guidance on DNA analysis in general and in some cases, mixture interpretation. A growing number are becoming available from various organizations around the world (see Table A1.1). A 2019 review noted that 34 guidance documents related to forensic DNA analysis and interpretation were published in the previous three years (Butler & Willis 2020). While many of these documents are designed to be specific for certain regions, there is value in knowing what others are doing and learning from them, as science knows no boundaries. Understanding the authority under which various documents are created, who is involved in creating them, and who uses or enforces the requirements or recommendations can be helpful.

Table A1.1. Documents that govern and influence DNA operations in accredited forensic laboratories. The order of the information does not imply preference. Abbreviations: AAFS = American Academy of Forensic Sciences, ANSI = American National Standards Institute, ANAB = ANSI National Accreditation Board, ASB = AAFS Standards Board, ASCLD/LAB = American Society of Crime Laboratory Directors/Laboratory Accreditation Board, ASTM = American Society for Testing and Materials, DAB = DNA Advisory Board, ENFSI = European Network of Forensic Science Institutes, FBI = Federal Bureau of Investigation, IEC = International Electrotechnical Commission, ILAC = International Laboratory Accreditation Cooperation, ISFG = International Society for Forensic Genetics, ISO = International Organization for Standardization, NDIS = National DNA Index System, OSAC = Organization of Scientific Area Committees for Forensic Science, QAS = Quality Assurance Standards, SDO = standards developing organization, SWGDAM = Scientific Working Group on DNA Analysis Methods, UK = United Kingdom, WG = Working Group.

Document	Authority	Who Creates	Who Uses or Enforces
FBI QAS (1998/1999 updated in 2009, 2011, 2020)	Law passed by Congress in 1994; issued by FBI Director	Originally DAB (1995-2000), now SWGDAM	FBI and ANAB auditors to assess U.S. forensic laboratories
ILAC G19 (2014) and **ISO/IEC 17025** (2017)	Standards community	ISO committee	Accrediting bodies (ANAB and formerly ASCLD/LAB)
Guidelines & Best Practices	Forensic practitioner community	SWGDAM, ENFSI DNA WG, ISFG DNA Commission	Forensic laboratories and practitioners (not required)

Document	Authority	Who Creates	Who Uses or Enforces
UK Forensic Science Code of Practice	UK Forensic Science Regulator	UK Forensic Science Regulator working group	UK forensic laboratories and practitioners
ASB/ASTM Standards (and OSAC Registry)	SDOs with forensic practitioner community input	SDOs (ASB, ASTM) and OSAC	Accrediting bodies as they are adopted

Groups that have commented on or proposed recommendations for DNA mixture interpretation include the ISFG DNA Commission (Gill et al. 2006b, Gill et al. 2012, Coble et al. 2016, Gill et al. 2018, Gill et al. 2020a), the German Stain Commission (Schneider et al. 2006b, Schneider et al. 2009), the European Network of Forensic Science Institutes DNA Working Group (Morling et al. 2007, ENFSI 2017), the Technical UK DNA Working Group on Mixture Interpretation (Gill et al. 2008), the Biology Specialist Advisory Group (BSAG) of the Australian and New Zealand forensic science community (Stringer et al. 2009), an FBI mixture committee (Budowle et al. 2009), the UK Forensic Science Regulator (UKFSR 2018a, UKFSR 2018b), AAFS Standards Board (ANSI/ASB 2018, ANSI/ASB 2019, Press 2020, ANSI/ASB 2020), and SWGDAM (SWGDAM 2010, SWGDAM 2015, SWGDAM 2017a). These efforts are briefly described below.

A1.7.1. ISFG DNA Commission and European Efforts in Mixture Interpretation

The International Society for Forensic Genetics (ISFG) has a DNA Commission that periodically addresses important topics in the field and makes recommendations. DNA mixture interpretation has been a part of five ISFG DNA Commissions (Gill et al. 2006b, Gill et al. 2012, Coble et al. 2016, Gill et al. 2018, Gill et al. 2020a).

In July 2006, the DNA Commission of the International Society for Forensic Genetics (ISFG) published nine recommendations (Box A1.4) covering multiple mixture interpretation principles (Gill et al. 2006b). In one of these recommendations, the ISFG DNA Commission endorsed the mixture deconvolution steps published in 1998 by the Forensic Science Service (Clayton et al. 1998). Since several forensic statisticians were part of this Commission, these recommendations favor approaches involving likelihood ratios that had previously been published (Evett et al. 1991, Weir et al. 1997).

An editorial accompanied the 2006 ISFG DNA Commission recommendations (Schneider et al. 2006a). The authors describe the purposes behind these initial DNA mixture interpretation recommendations:

> "…DNA evidence alone could be decisive for obtaining a conviction of an accused suspect. Thus, the interpretation of the observed DNA profile of a given stain in the context of the case needs to include a reasonable biostatistical evaluation of the weight of the evidence. At the same time, *the molecular biological tools available to the forensic geneticist have become more and more sensitive to the point where the genomic DNA from a few dozen cells may be sufficient to obtain a full STR profile from an unknown offender. As a result, the number of DNA mixtures composed from full or partial profiles from two or more contributors* (who could be offenders, victims, or individuals not associated with the crime event) *has increased*

> *significantly*. The biostatistical interpretation of such mixed DNA profiles is a very challenging task that sometimes leads to controversial views about correct mathematical approaches for estimating the weight of the evidence. Indeed, *diverse practices have already arisen between laboratories*, hence there is an urgent need to formulate recommendations… These recommendations have been written to serve two purposes: to define a generally acceptable mathematical approach for typical mixture scenarios and to address open questions where practical and generally accepted solutions do not yet exist…This paper is a 'high level' treatise on the mathematical principles to analyse complex mixtures. We realise that it will not be possible for most laboratories to immediately implement the methods described. *Our intention is primarily to specify a consensus approach to act as the foundation stone*. *Hopefully we will encourage the development of expert systems to take care of the onerous calculations*." (Schneider et al. 2006a, emphasis added).

Following the 2006 ISFG DNA Commission publication, a Technical UK DNA Working Group was formed to provide a detailed response that considered their national needs and court experiences with DNA mixture interpretation (Gill et al. 2008). An FBI Laboratory working group (Budowle et al. 2009) and SWGDAM (SWGDAM 2010) also built upon the 2006 ISFG DNA Commission foundational principles.

The December 2007 issue of *Forensic Science International: Genetics* contained a letter to the editor entitled "Interpretation of DNA mixtures – European consensus on principles" that was co-authored by chairs of the European DNA Profiling Group (EDNAP), the DNA Working Group of the European Network of Forensic Science Institutes (ENFSI), the German Stain Commission, and the Technical UK DNA Working Group (Morling et al. 2007). These groups expressed their support for the 2006 ISFG recommendations on mixture interpretation (Gill et al. 2006b). This letter to the editor emphasized "laboratories must invest in continuous education of the staff in the interpretation of DNA mixtures" (Morling et al. 2007). Appendix 2 of our report discusses this topic further.

The ISFG 2006 recommendations and principles were also supported by an Australian and New Zealand Biology Specialist Advisory Group (BSAG) (Stringer et al. 2009). The BSAG provided some additional commentary:

> "The likelihood ratio is a common approach to mixture interpretation in Australia and New Zealand. RMNE [random man not excluded] is considered an acceptable alternative approach to DNA interpretation. If the crime stain DNA profile is low level and some minor alleles are the same size as stutters of major alleles, and/or if drop-out is possible, then extra consideration needs to be given to the method of statistical interpretation… It is recommended that the scientist is trained in the primary methodology routinely used in their laboratory and has an understanding of other statistical approaches for DNA interpretation. The scientific community has a responsibility to support improvement of standards of scientific reasoning in the Justice system" (Stringer et al. 2009).

Box A1.4. ISFG 2006 Recommendations on DNA Mixture Interpretation

Recommendation 1: The likelihood ratio is the preferred approach to mixture interpretation. The RMNE [Random Man Not Excluded; also known as the Combined Probability of Inclusion, CPI] approach is restricted to DNA profiles where the profiles are unambiguous. If the DNA crime stain profile is low level and some minor alleles are the same size as stutters of major alleles, and/or if drop-out is possible, then the RMNE method may not be conservative.

Recommendation 2: Even if the legal system does not implicitly appear to support the use of the likelihood ratio, it is recommended that the scientist is trained in the methodology and routinely uses it in case notes, advising the court in the preferred method before reporting the evidence in line with the court requirements. The scientific community has a responsibility to support improvement of standards of scientific reasoning in the court-room.

Recommendation 3: The methods to calculate likelihood ratios of mixtures (not considering peak area) described by Evett et al. (Evett et al. 1991) and Weir et al. (Weir et al. 1997) are recommended.

Recommendation 4: If peak height or area information is used to eliminate various genotypes from the unrestricted combinatorial method, this can be carried out by following a sequence of guidelines based on Clayton et al. (Clayton et al. 1998).

Recommendation 5: The probability of the evidence under H_p is the province of the prosecution and the probability of the evidence under H_d is the province of the defense. The prosecution and defense both seek to maximize their respective probabilities of the evidence profile. To do this both H_p and H_d require propositions. There is no reason why multiple pairs of propositions may not be evaluated.

Recommendation 6: If the crime profile is a major/minor mixture, where minor alleles are the same size (height or area) as stutters of major alleles, then stutters and minor alleles are indistinguishable. Under these circumstances alleles in stutter positions that do not support H_p should be included in the assessment.

Recommendation 7: If drop-out of an allele is required to explain the evidence under H_p: (S = ab; E = a), then the allele should be small enough (height/area) to justify this. Conversely, if a full crime stain profile is obtained where alleles are well above the background level, and the probability of drop-out approaches $Pr(D) \approx 0$, then H_p is not supported.

Recommendation 8: If the alleles of certain loci in the DNA profile are at a level that is dominated by background noise, then a biostatistical interpretation for these alleles should not be attempted.

Recommendation 9: In relation to low copy number, stochastic effects limit the usefulness of heterozygous balance and mixture proportion estimates. In addition, allelic drop-out and allelic drop-in (contamination) should be taken into consideration of any assessment.

The German Stain Commission, a group of scientists from Germany's Institutes of Legal Medicine, introduced a three-part classification scheme for DNA mixtures: Type A (no major contributor), Type B (major and minor contributors distinguishable), and Type C (low-level DNA with stochastic effects). Their recommendations were first provided in German (Schneider et al. 2006b) and then republished in English (Schneider et al. 2009) to increase accessibility.

Under this classification scheme, Type A mixtures require a biostatistical analysis that can be performed with an LR or RMNE (CPI). Type B mixtures can be deconvoluted into the major and minor components, usually if they are present with consistent peak-to-height ratios of approximately 4:1. The major component following deconvolution can be treated as a single-

source profile and a random match probability calculated. For Type C mixtures, where all alleles may not be seen due to allele dropout, a biostatistical interpretation is not appropriate, and a clear decision about whether to include or exclude a suspect may be difficult to reach. It is important to keep in mind that these German Stain Commission categories were originally developed when two-person mixtures were most commonly seen in forensic laboratories (see Butler 2015a, p. 133) – and were not intended to address the complex mixtures processed today with PGS systems.

Later iterations of the ISFG DNA Commission provided recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods (Gill et al. 2012), the validation of software programs performing biostatistical calculations for forensic genetics applications (Coble et al. 2016), guidelines on formulating propositions for investigative and court-going purposes (Gill et al. 2018), and advice on evaluating low-level DNA results considering activity level propositions (Gill et al. 2020a).

In 2017, the ENFSI DNA Working Group[34], which has members from more than 50 organizations across 35 European countries, published a best practice manual, which outlined experiments for performing internal validation of probabilistic genotyping software used in DNA mixture interpretation (ENFSI 2017). This guidance builds upon the ISFG DNA Commission recommendations (Coble et al. 2016).

In 2018, the UK Forensic Science Regulator offered guidance on DNA mixture interpretation (UKFSR 2018a) and software validation for DNA mixture interpretation (UKFSR 2018b). For example, the software validation document promotes use of a validation library with supporting information covering software specifications, risk assessments, technical reports or scientific publications, a validation plan including the user acceptance criteria, information on the statistical models used, a statistical specifications report including underlying data on which any conclusions are based, the validation report with data summaries and assessment against the acceptance criteria, and a record of validation approval (UKFSR 2018b). A recent annual report[35] from the Regulator states:

> "There will always be limits to the complexity of DNA mixtures that can safely be interpreted, but the guidance published in FSR-G-222 [(UKFSR 2018a)] and FSR-G-223 [(UKFSR 2018b)] should ensure that interpretation does not stray beyond what is scientifically robust" (March 15, 2019, p. 47).

A1.7.2. SWGDAM and U.S. Efforts in Mixture Interpretation

In the United States, the FBI Laboratory has sponsored the Technical Working Group on DNA Analysis Methods (TWGDAM) from 1988 to 1998 and the Scientific Working Group on DNA Analysis Methods (SWGDAM)[36] from 1998 to the present. An important purpose of TWGDAM and SWGDAM continues to be a semi-annual gathering of forensic DNA scientists to share protocols and ideas and to write guidelines where appropriate. From 1995 to 2000, the FBI also had a Federal Advisory Committee known as the DNA Advisory Board (DAB) that crafted the original Quality Assurance Standards (QAS), which were first issued

[34] See http://enfsi.eu/about-enfsi/structure/working-groups/dna/
[35] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786137/FSRAnnual_Report_2018_v1.0.pdf (p. 47)
[36] See https://www.swgdam.org/about-us

in October 1998. Since 2000, when the DAB's charter expired, revisions to the QAS have been performed by SWGDAM.

Historically, DNA mixture interpretation has been minimally addressed in the QAS, with more detailed information included in SWGDAM guidance documents (see Table A1.2). For example, the 2011 version of the QAS contains one requirement regarding mixture interpretation in Standard 9.6.4:

> "Laboratories analyzing forensic samples shall have and follow a documented procedure for mixture interpretation that addresses major and minor contributors, inclusions and exclusions, and policies for the reporting of results and statistics."

Contemporaneous SWGDAM guidance documents then provided more detailed suggestions (SWGDAM 2010, SWGDAM 2012).

In February 2000, the FBI's DNA Advisory Board endorsed the use of CPI and LR methods for providing statistical support of an inclusion following mixture interpretation (DAB 2000). In their first publication regarding implementation of STRs in forensic casework, the FBI Laboratory discussed the importance of a stochastic threshold when performing mixture interpretation and using the CPI statistic (Moretti et al. 2001a, Moretti et al. 2001b). An FBI Mixture Committee provided further guidance on using stochastic thresholds with CPI a few years later (Budowle et al. 2009).

An interlaboratory study conducted by NIST in 2005, designated MIX05, demonstrated variation across the community in approaches being taken at the time with two-person mixtures (Butler et al. 2018a). Recognizing a need to address variability observed in approaches being taken with mixture interpretation, SWGDAM started a Mixture Committee in January 2007. The committee discussed topics surrounding mixture interpretation and drafted what was eventually published three years later as a 28-page document "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories" (SWGDAM 2010). The SWGDAM 2010 guidelines built upon many of the 2006 ISFG DNA Commission recommendations (Gill et al. 2006b), particularly in relationship to interpretation of peaks in the stutter position (see Butler 2015a, pp. 148-149).

Updates were made to the 2010 guidelines by the SWGDAM Autosomal STR Committee, and a 90-page document was released in 2017 providing a variety of examples in handling binary methods of DNA mixture interpretation (SWGDAM 2017a). Further revisions of the SWGDAM interpretation guidelines are under development to assist with guidance on probabilistic genotyping approaches. It is helpful to keep in mind that guidelines and standards take time to develop and are not always available when technology or interpretation approaches are initially implemented. Other documents from SWGDAM related to DNA mixture interpretation include verbal equivalents for likelihood ratios (SWGDAM 2018) and validation guidelines (see next section).

In September 2018, the U.S. Department of Justice issued a Uniform Language for Testimony and Reports (ULTR)[37] for forensic autosomal DNA examinations using probabilistic genotyping systems. This ULTR supports the LR verbal scale defined earlier by

[37] https://www.justice.gov/olp/uniform-language-testimony-and-reports

SWGDAM with qualitative equivalent categories of *uninformative* (LR=1), *limited support* (LR = 2 to <100), *moderate support* (LR = 100 to <10,000), *strong support* (LR = 10,000 to < 1 million), and *very strong support* (LR > 1 million) (SWGDAM 2018).

The Organization of Scientific Area Committees for Forensic Science (OSAC)[38] was created in 2014 as a joint venture between NIST and the Department of Justice (Butler 2015c). OSAC's goal is to facilitate the development of technically sound, science-based standards through a formal standards developing organization (SDO) process and placement of approved standards and guidelines on an OSAC Registry. In May 2020, the first two DNA standards were placed on the OSAC Registry (Press 2020). Both standards relate to DNA mixture interpretation: "Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol" (ANSI/ASB 2018 Standard 020) and "Standard for Forensic DNA Interpretation and Comparison Protocols" (ANSI/ASB 2019 Standard 040). These two documents were originally drafted by OSAC in 2015 and 2016 and then further developed and published by the AAFS Standards Board in 2018 and 2019 before being reviewed by OSAC for placement on the registry in 2020.

These new standards, which are meant to complement the FBI QAS and build upon SWGDAM guidelines, require laboratories to demonstrate that their protocols produce consistent and reliable conclusions with DNA samples different from the ones used in the initial validation studies. These standards also require that laboratories do not attempt to interpret DNA mixtures beyond the scope that they have validated and verified. For example, if a lab has tested its protocol for up to three-person DNA mixtures, it should not interpret casework that contains DNA from four or more people (Press 2020).

Additional standards to assist in DNA mixture interpretation in the future are in the OSAC pipeline and being finalized through the AAFS Standards Board DNA Consensus Body[39] with the SDO process.

A1.7.3. U.S. Validation Guidance Regarding DNA Mixture Interpretation

Validation studies assist in understanding the degree of reliability of scientific methods. This section briefly reviews FBI QAS validation requirements and SWGDAM guidance related to DNA mixture interpretation. For the forensic DNA community, levels of validation have been divided into developmental validation, often performed under the auspices of the vendor, and internal validation, performed within each user laboratory or laboratory system. The purpose of these studies is to explore the capabilities and limitations of the methods being used in the laboratory.

Often publications in the forensic DNA literature state, when describing the developmental validation of, for example, a new DNA test kit or methodology, that "SWGDAM validation guidelines were followed." In making such statements, authors of these publications may be trying to convey that because suggested mixture studies were performed, the method should be accepted as robust, reliable, and reproducible. In order for laboratory decision makers to

[38] https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science
[39] https://www.asbstandardsboard.org/aafs-standards-board-consensus-body-descriptions/

assess such statements, it is important to understand these guidelines as they relate to DNA mixture interpretation and how they have changed over the years. The general nature of current validation requirements or guidelines is such that variability can exist in the ways these studies are conducted.

Over the past several decades, SWGDAM has regularly updated its validation guidelines as well as validation requirements in the FBI Quality Assurance Standards (QAS) (Table A1.2). Validation guidelines were initially issued for RFLP techniques in 1989 (TWGDAM 1989) and for PCR techniques beginning in 1991 (TWGDAM 1991). PCR-based validation guidelines have been refined and updated in 1995, 2004, 2012, and 2016. In addition, validation guidelines for probabilistic genotyping software (PGS) systems were issued by SWGDAM in 2015 (SWGDAM 2015). Validation requirements contained in the FBI QAS were published in 1998, 1999, 2009, 2011, and 2020. Content related to DNA mixture interpretation in each of these documents is summarized in Table A1.2 with the exception of the SWGDAM PGS validation guidelines, which are covered separately below.

Table A1.2. A chronological review of validation guidelines or requirements prepared by SWGDAM or its predecessors that relate to DNA mixture interpretation.

Year	Document	**Information related to DNA mixture interpretation** (bold font used to add emphasis)
1989	TWGDAM Quality Assurance	*(no mention of mixtures)*
1991	TWGDAM Quality Assurance	4.1.5.5 Mixed Specimen Studies - **Investigate the ability of the system to detect the components of mixed specimens and define the limitations of the system.** 4.4.1.6 Where more than one locus is amplified in one sample mixture, **the effects of such amplification on each system (alleles) must be addressed and documented**.
1995	TWGDAM Quality Assurance	*(mixture information is the same as TWGDAM 1991)*

Year	Document	**Information related to DNA mixture interpretation** (bold font used to add emphasis)
1998 and 1999	DAB QAS Forensic and Database	8.1.2 Novel forensic methodologies shall undergo developmental validation to ensure the accuracy, precision and reproducibility of the procedure. The developmental validation shall include the following: … 8.1.2.2 Species specificity, sensitivity, stability and **mixture studies are conducted**. *(no mention of mixtures under 8.1.3 internal validation requirements)* 9.1.3 The laboratory shall have **a procedure for differential extraction** of stains that potentially contain semen. 9.6 The laboratory shall **have and follow written general guidelines for the interpretation of data.** *(no mention of mixtures)*
2004	SWGDAM Validation Guidelines	2.8 Mixture studies: **The ability to obtain reliable results from mixed source samples should be determined.** 3.5 Mixture studies: When appropriate, forensic casework laboratories must **define and mimic the range of detectable mixture ratios**, including detection of major and minor components. Studies should be conducted **using samples that mimic those typically encountered in casework** (e.g., postcoital vaginal swabs).
2009	FBI QAS	8.2.1 Developmental validation studies shall include, where applicable, characterization of the genetic marker, species specificity, sensitivity studies, stability studies, reproducibility, case-type samples, population studies, **mixture studies**, precision and accuracy studies, and PCR-based studies. PCR-based studies include reaction conditions, assessment of differential and preferential amplification, effects of multiplexing, assessment of appropriate controls, and product detection studies. All validation studies **shall be documented**. 8.3.1 Internal validation studies conducted after the date of this revision shall include as applicable: known and non-probative evidence samples or mock evidence samples, reproducibility and precision, sensitivity and stochastic studies, **mixture studies**, and contamination assessment. Internal validation studies **shall be documented and summarized**. The technical leader shall approve the internal validation studies. 8.3.2 Internal validation shall define quality assurance parameters and interpretation guidelines, including as applicable, **guidelines for mixture interpretation**. 9.6.4 Laboratories analyzing forensic samples **shall have and follow a documented procedure for mixture interpretation** that addresses major and minor contributors, inclusions and exclusions, and policies for the reporting of results and statistics.

Year	Document	Information related to DNA mixture interpretation (bold font used to add emphasis)
2011	FBI QAS	*(mixture information is the same as QAS 2009)*
2012	SWGDAM Validation Guidelines	2.2.2.2 Quality assurance parameters and interpretation guidelines shall be derived from internal validation studies. For example, lower template DNA may cause extreme heterozygote imbalance; as such, **empirical heterozygote peak-height ratio data could be used to formulate mixture interpretation guidelines and determine the appropriate ratio by which two peaks are determined to be heterozygotes**. In addition to establishing an analytical threshold, results from sensitivity studies could be used to determine the extent and parameters of quality control tests that reagents require prior to their being used in actual casework. 3.8 Mixture studies: The ability to obtain reliable results from mixed-source samples should be determined. These **studies will assist the laboratory to establish guidelines for mixture interpretation**, which may include determination of the number of contributors to the mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions. 4.4 Mixture studies: **Mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated**. These studies will assist a casework laboratory to establish guidelines for mixture interpretation, which may include determination of the number of contributors to the mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions. **A simplified mixture study may also assist a databasing laboratory to recognize mixtures and/or contamination**. Table 1 ***Mixture studies will be required if the assay is intended to distinguish different contributors (male/female, major/minor, etc.)**.
2016	SWGDAM Validation Guidelines	*(mixture information is the same as SWGDAM 2012)*

Year	Document	Information related to DNA mixture interpretation (bold font used to add emphasis)
2020	FBI QAS	2. Definitions: Interpretation Software is a tool to assist the analyst in assessing the analyzed data by applying quality assurance rules, **performing mixture deconvolution**, and/or evaluating comparisons. Interpretation software may include probabilistic genotyping software or expert systems. 2. Definitions: Sensitivity studies (for the purposes of Standard 8.8) are used to assess the ability of the system to reliably determine the presence of a contributor's DNA over a broad variety of evidentiary typing results (to include **mixtures** and low-level DNA quantities). 2. Definitions: Specificity studies (for the purposes of Standard 8.8) are used to evaluate the ability of the system to provide reliable results over a broad variety of evidentiary typing results (to include **mixtures** and low-level DNA quantities). 8.2.1 Developmental validation studies shall include, where applicable, characterization of the genetic marker, species specificity, sensitivity studies, stability studies, case-type samples, population studies, **mixture studies**, precision and accuracy studies, and PCR-based studies. PCR-based studies include reaction conditions, assessment of differential and preferential amplification, effects of multiplexing, assessment of appropriate controls, and product detection studies. All validation studies shall be documented. 8.3.1 Internal validation studies shall include as applicable: known and nonprobative evidence samples or mock evidence samples, precision and accuracy studies, sensitivity and stochastic studies, **mixture studies**, and contamination assessment studies. 8.3.2 Internal validation shall define quality assurance parameters and interpretation guidelines, including, as applicable, **guidelines for mixture interpretation** and the application of appropriate statistical calculations. 8.3.2.1 **Mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework.** 9.6.6 **Have and follow procedures for mixture interpretation** that address the following: 9.6.6.1 The assessment of the number of contributors. 9.6.6.2 The separation of contributors (e.g., major versus minor). 9.6.6.3 The criteria for deducing potential contributors. 9.10.5 The approaches to performing statistical calculations. 9.10.5.1 For autosomal STR typing, the procedure shall address homozygous and heterozygous typing results, multiple locus profiles, **mixtures**, minimum allele frequencies, and where appropriate, biological relationships.

As can be seen in Table A1.2, the amount of information regarding mixture interpretation has increased over the years in newer versions of the SWGDAM validation guidelines and the FBI QAS requirements. A more detailed comparison of topics covered in the various versions for developmental and internal validation requirements is available elsewhere (see Tables 7.2 and 7.3 on pages 179-181 in Butler 2012).

Historically, limited information was provided regarding the suggested and/or required studies to inform mixture interpretation protocols. Rather, the early emphasis was to

> "investigate the ability of the system [DNA testing method] to detect components of mixed specimens and define the limitations of the system" (TWGDAM 1991, section 4.1.5.5)

or to determine

> "the ability to obtain reliable results from mixed source samples" (SWGDAM 2004, section 2.8)

and to

> "define and mimic the range of detectable mixture ratios" in validation experiments (SWGDAM 2004, section 3.5).

The 2012 SWGDAM validation guidelines first emphasized performing validation studies that reflect the complexity of samples being examined in casework:

> "Mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated" (SWGDAM 2012, Guideline 4.4).

The 2012 guidelines do not specifically address the need to define the limitations of the system; rather, they suggest studies to help establish laboratory guidelines for mixture interpretation (SWGDAM 2012, Guideline 3.8). This text was maintained in the 2016 version of the document (SWGDAM 2016).

The 2020 update to the FBI QAS now requires that

> "mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework" (QAS 2020, Standard 8.3.2.1).

The 2009 version included a more open requirement, stating:

> "Laboratories analyzing forensic samples shall have and follow a documented procedure for mixture interpretation" (QAS 2009, Standard 9.6.4).

An observation made in conducting this scientific foundation review is that, historically, FBI QAS validation requirements and SWGDAM validation guidelines have become *task-driven* rather than *performance-based*. In other words, the requirements and guidelines may be treated by some as a checklist of studies that need to be completed to satisfy requirements rather than a demonstrated performance of the accuracy or reliability of results obtained using the method. Recommended studies include, for example, known and nonprobative evidence samples, sensitivity and stochastic studies, precision and accuracy assessments, mixture studies, and contamination assessment. Under mixture studies, the guidelines state:

> "mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated" (SWGDAM 2016, Section 4.4).

Ideally, developmental validation studies are conducted by vendors to meet specific performance measures, and internal validation experiments demonstrate similar performance under individual laboratory conditions.

Performance-based approaches are preferable over checklists of validation studies conducted because they can provide information on the limitations of the method. As noted in the previous section, a new documentary standard was published recently: "Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol" (ANSI/ASB 2018). This document discusses performance in more detail than previous guidance documents, but since it is new, feedback is not yet available regarding routine implementation by forensic DNA laboratories. For example, the standard requires:

> "The laboratory shall verify and document that the mixture interpretation protocols developed from the validation studies generate reliable and consistent interpretation and conclusions for the types of mixed DNA samples typically encountered by the laboratory" (ANSI/ASB 2018, standard 4.4)

and explains further that

> "DNA mixture data from different sets of contributors than used in the initial validation studies shall be used to verify the protocol" (ANSI/ASB 2018, p. 6).

Forensic laboratories are accredited to international standard ISO/IEC 17025:2017, which describes the types of information that can be used for method validation: (1) calibration or evaluation of bias and precision using reference materials, (2) systematic assessment of the factors influencing the result, (3) testing method robustness through variation of controlled parameters, (4) comparison of results achieved with other validated methods, (5) interlaboratory comparisons, and (6) evaluation of measurement uncertainty of the results based on the theoretical principles of the method and practical experience of the performance of the sampling or test method (ISO/IEC 17025:2017, Standard 7.2.2.1 note 2).

The ANAB accreditation requirements, under which most U.S. forensic laboratories are assessed, state:

> "The laboratory shall have a procedure for method validation that: (a) includes the associated data analysis *and interpretation*; (b) establishes the data required to report a result, opinion, or *interpretation*; and (c) identifies limitations of the method, reported results, opinions, and *interpretations*" (ANAB 2019, Section 7.2.2.2.1, emphasis added).

Historically, forensic DNA laboratories have conducted mixture studies during their internal validation experiments with emphasis on *robustness* (does the test produce a result?) and *detectability* (can minor alleles in a two-person mixture with multiple mixture ratios be detected?) rather than *reliability* (was interpretation of the mixture data accurate and consistent if repeated?). Publicly accessible performance-based validation data covering the desired factor space to achieve confidence in interpreting complex mixtures involving more than two contributors have been limited (see Chapter 4 in this report).

A1.7.4. Requirements and Expectations for PGS Validation

The ISFG DNA Commission from 2012 concluded:

> "The introduction of software solutions to interpret DNA profiles must be accompanied by a validation process ensuring conformity with existing standard laboratory procedures. … Software tools used for casework implementation must be evaluated with known samples and each laboratory will have to establish reporting guidelines and testimony training to properly present the results to courts" (Gill et al. 2012).

Several organizations and individual researchers have provided guidance on PGS validation. A brief history and overview of this guidance are provided here.

A1.7.4.1. Published Input from Software Developers

In 2006, the TrueAllele PGS developer, Mark Perlin, described his thoughts on scientific validation of mixture interpretation methods in a *Proceedings of the International Symposium for Human Identification* submission with a focus on precision, accuracy, and reproducibility (Perlin 2006).

In 2014, the STRmix developers, John Buckleton, Jo-Anne Bright, Duncan Taylor, and two colleagues, Ian Evett and James Curran, provided their thoughts on some recommended tests when validating PGS systems (Bright et al. 2015). Four experiments were suggested: (1) comparison of the expected LR with no drop-out or drop-in, (2) the effect of drop-out, (3) the effect of drop-in, and (4) reproducibility. Some examples were run with single-source profiles and simple two-person mixtures using STRmix, LRmix, and Lab Retriever. They conclude:

> "*An understanding of the models within each of the program[s] and their limitations is required* in order to validate interpretation software" (Bright et al. 2015, emphasis added).

They continue:

> "Gaining an understanding of the behavior of the software under certain conditions is central to the developmental validation process prior to use in casework... [It is] an important part of the internal validation and training process prior to implementation of software. *This includes calibration based on ground-truth cases where the contributors are known* and case hardening to test how a program performs in the real world" (Bright et al. 2015, emphasis added).

Developers of the discrete PGS systems LRmix and Lab Retriever write that "model and software validation are inherently entangled" and provide an example of examining over 1,000 LR calculations for their LRmix validation (Haned et al. 2016). They describe four principle steps for software validation: (1) define the statistical specifications of the software (i.e., document the theory behind the model); (2) carry out analytical verification, which involves manually calculating LR values for simple cases and comparing results to the software output (while keeping in mind that as the model becomes more complex, analytical verification may not be possible); (3) compare results to data obtained from alternative

software, which may rely on a similar or a different probabilistic model; and (4) verify the code itself through visual inspection and recoding, which they note is most easily achievable through open-source software (Haned et al. 2016).

These authors also note:

> "The more complex the model, the greater the number of assumptions that are required. Increasing the number of variables incorporated into such a model also increases the chance of creating dependencies. Such models require a validation protocol that specifically addresses the additional interactions, and care must be taken to clearly define the variables. *We caution that complex models may at some point begin to produce unrealistic results, and hence become counter-productive*. More generally, the validation criteria should be explicit to the end users, and a determination made as to whether these criteria are fit for purpose" (Haned et al. 2016, emphasis added).

A1.7.4.2. SWGDAM 2015 PGS Validation Guidelines

The SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems were approved and posted on the SWGDAM website on June 15, 2015 (SWGDAM 2015). They begin:

> "These guidelines are not intended to be applied retroactively. It is anticipated that they will evolve with future developments in probabilistic genotyping systems."

Topics covered include documentation required, computer system control measures, developmental validation studies recommended to be performed by developers, internal validation studies to be performed by forensic laboratories, and performance checks with any software modifications (SWGDAM 2015). Suggested readings include three published references available at the time (Gill et al. 2012, Kelly et al. 2014, Steele & Balding 2014). The introduction states:

> "Prior to validating a probabilistic genotyping system, the laboratory should ensure that it possesses the appropriate foundational knowledge in the calculation and interpretation of likelihood ratios. Laboratories should also be aware of the features and limitations of various probabilistic genotyping programs and the impact that those items will have on the validation process."

The 2015 SWGDAM PGS validation guidelines state that the system shall be validated "prior to usage for forensic applications" (1.1), that "the laboratory shall document all validation studies in accordance with the FBI Quality Assurance Standards" (1.2), and the laboratory should "have access to documentation that explains how the software performs its operations and activities" in order "to identify aspects of the system that should be evaluated through validation studies" (1.3). In addition, the laboratory is reminded to "verify that the software is installed on computers suited to run the software, that the system has been properly installed, and that the configurations are correct" (2.1) and that the following system control measures are in place: "every software release should have a unique version number" (2.2.1), "appropriate security protection [should exist] to ensure only authorized users can access the software and data" (2.2.2), that "audit trails to track changes to system data and/or verification of system settings [are] in place each time a calculation is run" (2.2.3), and that

"user-level security [exists] to ensure that system users only perform authorized actions" (2.2.4).

The developmental validation section of these guidelines stresses the importance of demonstrating "any known or potential limitations of the system" and emphasizes that "the underlying scientific principle(s) of the probabilistic genotyping methods and characteristics of the software should be published in a peer-reviewed scientific journal" and that these principles may include "modeling of stutter, allelic drop-in and drop-out, Bayesian prior assumptions such as allelic probabilities, and statistical formulae used in the calculation and algorithms" (3.1).

According to the 2015 SWGDAM guidelines, studies that should be performed for developmental validation include sensitivity (3.2.1), specificity (3.2.2), precision (3.2.3), case-type samples (3.2.4), control samples (3.2.5), and accuracy (3.2.6). Studies should "assess the ability of the system to reliably determine the presence of a contributor's(s') DNA over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities)" with "various sample types (e.g., different numbers of contributors, mixture proportions, and template quantities)" (SWGDAM 2015). The 2015 SWGDAM guidelines emphasize the need to understand the sensitivity and specificity of performance over a variety of conditions.

Under section 4 on internal validation, these guidelines state:

> "Data should be selected to test the system's capabilities and to identify its limitations. In particular, complex mixtures and low-level contributors should be evaluated thoroughly during internal validation, as the data from such samples generally help to define the software's limitations…" (SWGDAM 2015).

Internal validation should address samples with known contributors (4.1.1), hypothesis testing with contributors and non-contributors (4.1.2), variable DNA typing conditions (4.1.3), allelic peak height including off-scale data (4.1.4), single-source samples (4.1.5), DNA mixtures with various contributor ratios (4.1.6.1), various total DNA template quantities (4.1.6.2), various number of contributors (4.1.6.3), over- and under-estimating the number of contributors (4.1.6.4), allele sharing among contributors (4.1.6.5), partial profiles (4.1.7), allele drop-in (4.1.8), forward and reverse stutter (4.1.9), intra-locus peak height variation (4.1.10), inter-locus peak height variation (4.1.11), use of a different data set to establish software parameters and perform validation studies (4.1.12), sensitivity, specificity and precision studies (4.1.13), and additional challenge testing, such as the inclusion of non-allelic peaks from bleed-through or voltage spikes (4.1.14) (SWGDAM 2015).

A1.7.4.3. ISFG 2016 DNA Commission on Software Validation

In November 2016, the ISFG DNA Commission provided 16 recommendations on validation of software programs used in forensic genetics to perform biostatistical calculations (Coble et al. 2016). These recommendations are summarized as follows:

(1) software should be supported by a user manual and scientific publications describing the data model(s) used to permit independent recalculation to verify reproducibility of all computations;
(2) validation should be according to specified requirements and intended use with publicly available or disclosed data sets;
(3) each software version should be distinguishable and independently validated;
(4) software developers should provide instructions to users on how to validate and configure their software;
(5) a user manual should accompany software to enable trained users to understand and explain results;
(6) laboratories are responsible to provide sufficient training resources and support for users;
(7) software source code should be placed in a secure repository and algorithms described in sufficient details to ensure continued availability of software in the future;
(8) software bugs and their fixes need to be disclosed and users notified about updates and any quality assurance issues;
(9) software using random permutation algorithms, such as MCMC, needs to have a feature to set this function to a stable mode for repeatability testing purposes;
(10) laboratories should develop a documented validation plan prior to initiating software validation and have supporting publications describing the models, propositions, and parameters used by the software;
(11) laboratories should test the software on representative data generated in-house with reagents, instruments, analysis software, and conditions used routinely for casework;
(12) laboratories should test true donors (H_1 true) and non-donors (H_2 true) as well as related and unrelated individuals across a range of situations that span or exceed the complexity of cases likely to be encountered in casework;
(13) laboratories should determine whether software results are consistent with previous interpretation procedures if the data and/or methods exist;
(14) laboratories should develop standard operating procedures based on their internal validation data and outline the types of cases and data to which the software can be applied;
(15) laboratories should develop and follow a policy or procedure for training software end users in the laboratory; and
(16) the forensic community is encouraged to establish a public repository of typing results, including results from different challenging scenarios like low-level mixtures and related contributors, in a universal, standardized file format and to have this repository governed by a neutral organization to permit equal access to all interested international parties.

A1.7.4.4. ENFSI DNA Working Group 2017 Best Practice Manual

In May 2017, the ENFSI DNA Working Group issued a "Best Practice Manual for the internal validation of probabilistic software to undertake DNA mixture interpretation" that was intended to build upon the ISFG 2016 recommendations (see previous section). This document focuses on internal validation performed within a forensic laboratory. Regarding previous developmental (termed "external") validation, this best practice manual notes:

> "It will be a decision for the laboratory to be satisfied that the external validation is 'fit-for-purpose' within the scope of its intended use" (ENFSI 2017).

Section 4.1 in this document states:

> "…a person(s) should be nominated to be responsible to act as the 'local expert' with the broadest knowledge about the software."

Section 4.2 recommends:

> "The software developer should create instructions on how to validate and configure software within the laboratory…and supply a user manual…for end users" (ENFSI 2017).

A documented validation plan should be developed to take into account the types of samples the laboratory plans to analyze (Section 6.1). Mock casework samples that span the kinds of samples routinely tested by that laboratory, where ground truth is known, should be used (Section 6.2) and, where possible, results produced by the software should be compared for consistency with previous interpretation procedures used by the laboratory (Section 6.3). The laboratory should "establish a series of criteria that define the limitations of testing," such as "if the profile of interest is predominantly below some defined level or a specified number of alleles have dropped-out (under the prosecution hypothesis)" (Section 6.4). This document emphasizes:

> "It is important that users have a clear understanding on the limitations. To facilitate this, users must be presented with examples considered unsuitable for testing" (ENFSI 2017).

This ENFSI guidance document also discusses the probability of drop-in (Section 6.5), proficiency testing (Section 7), training (Section 12.1), and presentation of evidence (Section 13) and contains an appendix on terminology for probabilistic mixture models (Section 16.1).

A1.7.4.5. UK Forensic Science Regulator 2018 Guidance

In July 2018, the UK Forensic Science Regulator issued a 53-page guidance document on software validation for DNA mixture interpretation (UKFSR 2018b). A few points are highlighted here.

Section 6.1 discusses validation considerations specific to likelihood ratio calculations given that there is no "true" value for an LR. Section 6.2 reviews desired performance parameters (e.g., the software should be capable of analyzing three-person mixtures at a minimum), principles that should be incorporated into a DNA mixture interpretation model (e.g., limitations of all approaches should be made apparent to the customer), and routine operating quality checks required and data input considerations (e.g., an assessment of the evidence profile in the context of case circumstances, where possible, should always be undertaken before the use of software).

Section 7 reviews the process of validation defined in the UK Forensic Science Regulator's *Codes of Practice and Conduct* available at the time (UKFSR 2017) and a 2014 guidance document on validation (UKFSR 2014). Three additional stages are included with DNA

mixture interpretation: (1) validation of the statistical model, (2) software development and testing, and (3) user acceptance testing.

Under Section 7.5 covering conceptual and operational validation of the statistical model, this guidance document states:

> "…ideally *the underlying data on which conclusions are based should also be made available*, for example, as supplementary material within the journal or access provided online to downloadable material *including all data and a full statistical description*. This enables other scientists in the field to inspect it independently and verify the results obtained in order to enable general acceptance of the model concept within the scientific community. Such transparency is essential for any software used within the [criminal justice system], for which there can be no 'secret science'" (UKFSR 2018b, p. 25, emphasis added).

The guidance continues:

> "…[software] testing should utilize a variety of ground-truth cases for which the composition is known, and are of varying degrees of quality and complexity that represent the full spectrum of data that may typically be encountered in casework" (UKFSR 2018b, p. 25).

Assessment of reproducibility is needed including the magnitude of the variation when a statistical model

> "does not return precisely the same number on replicate analyses of identical data" (UKFSR 2018b, p. 26).

Also encouraged are boundary testing to experimentally determine the impact of increasing the number of contributors and benchmarking exercises comparing results with other software models or manual calculations that may be feasible with less complex data assessments (UKFSR 2018b, p. 26).

In addition, Section 7.10 of the UK guidance encourages creation of a validation library to maintain documentation from validation studies conducted and associated supporting materials including published articles and technical reports. Sections 8.1.4 and 8.1.5 state:

> "...the existing evidence that has been produced by a third party, and on which reliance is placed, must be relevant, available and adequate" and "the details of the analysis undertaken are both transparent and accessible to third parties" (UKFSR 2018b, p. 35).

A1.7.4.6. ANSI/ASB 2020 PGS Validation Standard

In July 2020, the AAFS Standards Board published the first standard on PGS validation (ANSI/ASB 2020). The foreword states:

> "Validations of [PGS] systems provide the study results and conclusions necessary for customers or forensic science service providers to have confidence in the evidence provided."

This document continues:

> "…each laboratory will need to perform internal studies to demonstrate the reliability of the software and any potential limitations."

The bibliography cited in Annex B of the document includes 16 references.

Under this new standard, developmental validation (4.1.2) and internal validation (4.1.3) require accuracy, sensitivity, specificity, and precision studies with:

> "case-type profiles of known composition that represent (in terms of number of contributors, mixture ratios, and total DNA template quantities) the range of scenarios that would likely be encountered in casework. Studies shall not be limited to pristine DNA but shall also include compromised DNA samples (e.g., low template, degraded, and inhibited samples)" (ANSI/ASB 2020).

This standard also states:

> "The internal validation shall not exceed the scope of the conditions tested in the developmental validation" (4.1.3), "All validation and performance check studies conducted by the laboratory shall be documented and retained by the laboratory" (4.5), and "Prior to implementation, the laboratory shall verify the functionality of its defined software settings and parameters utilizing different data sets than what were originally used to establish those settings and parameters" (4.7) (ANSI/ASB 2020).

Annex A with supporting information states:

> "Repeated testing and data analysis are critical to the understanding of variability. While specific requirements for the minimum number of studies and sample sets used for validation studies are not detailed in this standard, the laboratory shall *perform sufficient studies* to address the variability inherent to the various aspects of DNA testing, data generation, analysis and interpretation of data and user input parameters" (4.1.3) (ANSI/ASB 2020, emphasis added).

It continues:

> "All internal validation and performance check studies shall be documented and retained by the laboratory. Any validation and performance check studies may take a significant amount of time and are likely to result in a considerable amount of documentation output material. It is incumbent upon any laboratory performing these studies to *retain these results for the examination and evaluation by third parties*. The *results should be documented in such a way that the performance checks and validations can be reproduced* and decisions made on the basis of these studies documented…" (ANSI/ASB 2020, emphasis added).

As emphasized in previous guidance documents, internal validation studies of PGS software need to be sufficient to assess variability across the types of DNA mixtures expected to be seen in a laboratory, and results from these studies should be available for third-party review.

A1.8. History of DNA Mixture Interpretation Training

The 2007 article "Interpretation of DNA mixtures – European consensus on principles" emphasizes that:

> "laboratories must invest in continuous education of the staff in the interpretation of DNA mixtures" (Morling et al. 2007).

This point had been made previously by the ISFG DNA Commission:

> "Our discussions have highlighted a significant need for continuing education and research in this area [DNA mixture interpretation]" (Gill et al. 2006b).

A brief history of training workshops on this topic is included below. Further thoughts on needs in this area may be found in Appendix 2.

A1.8.1. Initial U.S. Training Workshop on Mixtures

The first DNA mixture training course in the United States was held as part of a scientific conference in Annapolis, Maryland, sponsored by International Business Communications on July 31, 1998 (IBC 1998). This workshop, titled "Resolution and Interpretation of Mixtures," included presentations by Peter Gill of the UK Forensic Science Service ("Distinguishing between Alleles, Artifacts and Genetic Anomalies in Mixture Interpretation"); James Curran, then working with Bruce Weir in the statistics department of North Carolina State University ("Calculating the Evidentiary Strength of Mixed DNA Profiles"); and Charles Brenner, a consultant in forensic mathematics ("Some Considerations of Race, Number and Accuracy").

Peter Gill began his July 1998 workshop presentation with the admonition: "Don't do mixture interpretation unless you have to!" He explained that forensic cases often have multiple stains and that a selection should be made, where possible, of samples that do not contain mixtures. He also emphasized that it was important to ensure that any mixtures obtained were consistent with casework circumstances (IBC 1998).

At this workshop, James Curran taught

> "if numbers are to be provided, they must be calculated with the same attention to appropriate methods as is given to the generation of the profiles in the first place" and
> "the key issue is to decide upon possible explanations for the mixed stains."

He worked through some examples in calculating likelihood ratios and the underlying assumptions (IBC 1998). Both James Curran and Peter Gill acknowledged John Buckleton's contribution to their work. All of the individuals who participated in this first DNA mixture workshop over 20 years ago are still active in the field, and the primary issues discussed have not changed.

A1.8.2. Training on Principles

To assist forensic DNA analysts in understanding issues and principles underpinning DNA mixture interpretation, more than 50 training workshops and presentations were organized or given by researchers from the National Institute of Standards and Technology (NIST) and collaborators (see below) between 2005 and 2014 (see Butler 2015a, Table 6.5). Slides for many of these workshops (e.g., AAFS 2008, AAFS 2011, ISHI 2010, ISHI 2011, ISHI 2012) are available on the NIST STRBase website[40].

Researchers from Boston University (BU) received a training grant from the National Institute of Justice (NIJ) that funded DNA mixture interpretation training workshops in 2010, 2011, and 2012. In addition, BU created a training website[41] with 12 lessons that examine the

[40] See https://strbase.nist.gov/
[41] http://www.bu.edu/dnamixtures/

various steps of mixture interpretation. In addition, the BU website contains more than 2,700 .fsa files with single-source, two-person, three-person, and four-person mixtures at different mixture ratios and DNA amounts that can be downloaded and used in training programs. An even more extensive set of DNA mixture profiles, known as PROVEDIt (Alfonse et al. 2018), is available[42] from Professor Catherine Grgicak now at Rutgers University.

The ISFG also maintains educational workshop materials shared at biennial conferences for its members[43] on a variety of topics including DNA mixture interpretation.

A1.8.3. Training on Probabilistic Genotyping Software

With the development and implementation of PGS systems, software-specific training courses have been created. In 2012, the European Forensic Genetics Network of Excellence (EuroForGen-NoE) created an online training academy[44] with webinars discussing DNA mixture interpretation using an open-source PGS system LRmix. The EuroForGen-NoE group demonstrated that training and use of a common PGS system could lead to uniformity of results obtained with DNA mixtures (Prieto et al. 2014).

Vendors providing PGS programs conduct training courses to support their appropriate use. For example, the STRmix team has provided almost 100 training courses between 2014 and 2018 with durations ranging from one to five days[45].

More recently, a webinar series organized by the FBI Laboratory has introduced hundreds of DNA analysts to PGS theory, methods, and software (Table A1.3).

Table A1.3. Webinar series on DNA mixture interpretation and probabilistic genotyping organized by FBI Laboratory and NIJ's Forensic Technology Center of Excellence. Original webinars were held from May 1, 2019 to July 17, 2019 and are now available in archived format at https://forensiccoe.org/webinar/online-workshop-series-probabilistic-genotyping-of-evidentiary-dna-typing-results/. Abbreviations: DOJ = Department of Justice, ESR = Institute of Environmental Science and Research, FBI = Federal Bureau of Investigation, FSSA = Forensic Science South Australia, NFI = Netherlands Forensic Institute, NYC OCME = New York City Office of Chief Medical Examiner, UNTHSC = University of North Texas Health Science Center.

Lesson	Topics	Presenters
Module 1	Mixture interpretation and introduction to probabilistic genotyping software (PGS)	**Tamyra Moretti** (FBI Laboratory, USA) **Peter Gill** (University of Oslo, Norway) **Lynn Garcia** (Texas Forensic Science Commission, USA)
Module 2	Statistical aspects of PGS	**David Balding** (University of Melbourne, Australia) **Mike Coble** (UNTHSC, USA) **Steven Myers** (California DOJ, USA) **John Buckleton** (ESR, New Zealand)

[42] https://lftdi.camden.rutgers.edu/provedit/files/
[43] https://www.isfg.org/Members+Area/Education
[44] https://www.euroforgen.eu/training/online-training-academy/
[45] See https://johnbuckleton.files.wordpress.com/2018/08/training.pdf

Lesson	Topics	Presenters
Module 3	PGS software and output: instructive overviews	**John Buckleton** (ESR, New Zealand) **Mike Coble** (UNTHSC, USA) **Peter Gill** (University of Oslo, Norway) **Mark Perlin** (Cybergenetics, USA)
Module 4	Validation of PGS	**Tamyra Moretti** (FBI Laboratory, USA) **Sarah Noël** (Montreal, Canada) **Duncan Taylor** (FSSA, Australia)
Module 5	Representation of statistical weight to stakeholders and the court	**David Kaye** (Penn State Law School, USA) **Tamyra Moretti** (FBI Laboratory, USA) **Steven Myers** (California DOJ, USA)
Module 6	PGS in U.S. courts	**John Buckleton** (ESR, New Zealand) **Jerrilyn Conway** (FBI Laboratory, USA) **Dawn Herkenham** (Leidos, USA) **Mark Perlin** (Cybergenetics, USA)
Module 7	Uncertainty and limitations of PGS	**Amke Caliebe** (University of Kiel, Germany) **Zane Kerr** (ESR, New Zealand) **Klaas Slooten** (NFI, The Netherlands) **Bianka Szkuta** (Victoria Police, Australia)
Module 8	PGS summation and special topics	**Jo-Anne Bright** (ESR, New Zealand) **Ted Hunt** (USDOJ, USA) **Klaas Slooten** (NFI, The Netherlands)

A1.9 Summary and Key Takeaways

Since initially described in 1985, DNA methods have changed and become more sensitive. This change has necessitated new approaches to DNA mixture interpretation. Guidance documents and training efforts have played an important role in the history of DNA mixture interpretation.

KEY TAKEAWAY #A1.1: Over the past 35 years, there has been an evolution of new technologies (different markers, kits, instruments, and software) for DNA analysis and interpretation strategies for DNA mixtures (manual deconvolution, binary and probabilistic models) along with a steady stream of peer-reviewed publications.

Forensic DNA testing operates in an evolving environment with an increasingly complex set of technologies. Often important changes and advances have been made across the community because of experiences in high-profile court cases or awareness of issues raised through participation in interlaboratory studies or collaborative exercises and several of these cases and studies are highlighted.

KEY TAKEAWAY #A1.2: Recommendations on DNA mixture interpretation from the 2006 ISFG DNA Commission (see Box A1.4) **serve as core foundational principles.**

These recommendations emphasize the value of an LR approach with mixture deconvolution and review difficulties when interpreting minor components in the presence of (a) artifacts like STR stutter products and (b) stochastic variation inherent with low amounts of DNA.

KEY TAKEAWAY #A1.3: Limited information has been provided in guidance documents, such as the FBI Quality Assurance Standards or the SWGDAM guidelines, regarding suggested or required studies to inform mixture interpretation protocols.

Future needs include promoting performance-based approaches to validation studies (see Chapter 4) and continuing education and research in DNA mixture interpretation (see Appendix 2 and Chapter 5). It would be helpful to have training workshops and seminars on validation to assist the forensic DNA community and stakeholders in strengthening DNA mixture interpretation.

Appendix 2: Training and Continuing Education

Effective training and continuing education of forensic practitioners are crucial to keep up with the evolving forensic DNA technologies and applications (e.g., see Appendix 1). Given these ongoing changes, "laboratories must invest in continuous education of the staff in the interpretation of DNA mixtures" (Morling et al. 2007). Stakeholders in the criminal justice system (e.g., law enforcement personnel, lawyers, and judges) using DNA results also benefit from regular training and continuing education to understand capabilities and limitations. A 2013 European review of education needs in forensic genetics (Poulsen & Morling 2013) found that on-the-job training played a larger role in developing DNA knowledge and skills than university studies due to ongoing developments in the field – particularly for DNA mixture interpretation. A culture of critical thinking and clear communication regarding DNA mixture interpretation is crucial as probabilistic genotyping software programs are implemented and results from low-level, complex mixtures are shared in written reports and court testimony. Defining what analysts need to know *rather than what they* need to do *will increase confidence and enhance practice.*

Technology alone cannot bring the desired and required improvements. Implementation of technology involves validation of specific methods in the laboratory as well as training for forensic scientists and the consumers of their data. The current requirements for training and continuing education are examined in this appendix. Considerations regarding professional development, ongoing literature awareness and access, training in searching and reading the literature, training for DNA technical leaders, and knowledge assessment are also discussed.

A2.1. Training and Continuing Education Needed for Expertise

In a recent annual report, the UK Forensic Science Regulator, Dr. Gillian Tully stated:

> "It is a clear expectation of the courts that expert evidence is presented by people who are indeed experts in their field. This necessitates an up-to-date knowledge of developments in the relevant field, which in turn necessitates access to scientific literature and sufficient time to ensure that each expert has the current relevant knowledge that they need" (UK Forensic Science Regulator 2018c, p. 10).

New information is regularly becoming available with each laboratory experiment or published article, thus studying scientific literature is crucial. An "up-to-date knowledge of developments in the relevant field" is an admirable goal, yet not all forensic scientists have access to the journals where relevant articles are published. Also, practitioners might not have time in their typical workday schedule to regularly study the latest developments in their field.

For forensic scientists in the trenches working cases, keeping up with an ever-growing body of literature from published research and sets of guidelines and standards from various organizations (e.g., see Butler & Willis 2020) can seem like an impossible task. During deliberations with our DNA Mixture Resource Group as part of this scientific foundation review (see Chapter 1), we discussed training and continuing education. A brief history of

training workshops covering DNA mixture interpretation is available in Appendix 1 of this report (section A1.9).

A2.1.1. Status of Education and Training in Europe

We are unaware of any published reports on education needs in U.S. forensic DNA laboratories; however, a study on education and training needs in Europe was conducted in 2012 and published the following year (Poulsen & Morling 2013). We acknowledge that more recent information is unavailable on the current state of education and training.

In a March 2013 report, the European Forensic Genetics Network of Excellence (EuroForGen-NoE) described information collected on the status of education, training, and career development in forensic genetics. A questionnaire was provided to national contact persons representing 28 European countries. Based on feedback received, the authors of this report conclude:

> "The most urgently needed courses are: interpretation of results and weight of evidence in crime cases [i.e., DNA mixture interpretation], interpretation of results in complex relationship cases, biostatistics in general, disaster victim identification and ethics" (Poulsen & Morling 2013).

The report states:

> "The rapid pace of changes…has resulted in a situation where most scientists currently responsible for analyzing [complex DNA mixture] results have not been formally educated in this field, but rather been 'trained at the job'… The possibility to analyze complex mixtures from multiple contributors, and the increased sensitivity…has pushed the methods to the limits of detection and interpretation…[and] have led to complex, and sometimes controversial, discussions about the reproducibility of borderline results and the best approach for a biostatistical interpretation taking into account all types of stochastic events… Consequently, this has led to an ever-increasing demand for continuing education to keep up-to-date with these developments… [multiple] groups have voiced a clear demand for more education in this field… For the time being, no institution has the capacity to provide special seminars or workshops to meet this demand, due to the lack of funding and, equally important, the lack of trained staff ready to take up this challenge…" (Poulsen & Morling 2013).

This expressed desire for additional DNA mixture interpretation training was echoed by members of our DNA Mixture Resource Group during our 2018 and 2019 discussions.

A2.2. Current DNA Training Requirements and Guidance on Continuing Education

In the United States, the FBI Quality Assurance Standards (QAS) governs requirements for accredited laboratories performing forensic DNA testing or utilizing the Combined DNA Index System (CODIS). Specific training and continuing education requirements are included in the QAS. In addition, the FBI's Scientific Working Group on DNA Analysis Methods (SWGDAM) has provided training guidelines (SWGDAM 2013, SWGDAM 2020).

The Organization of Scientific Area Committees for Forensic Science (OSAC)[46] is also developing documentary standards on training related to DNA testing.

A2.2.1. FBI Quality Assurance Standards

For training, the FBI QAS require that forensic DNA technical leaders, analysts, technical reviewers, and technicians meet minimum levels of education and experience and complete a competency test to demonstrate technical abilities (QAS 2011: Standard 5.2, 5.4, 5.5, and 5.6; QAS 2020: Standard 5.2, 5.4, 5.5, and 5.6 and Standard 6). The newest QAS version (which took effect July 1, 2020) provides new details on training requirements.

QAS 2020 Standard 6.1 and its subcomponents state:

> "The laboratory shall have a training program documented in a training manual for qualifying analysts and technicians. The training program shall (1) address all DNA analytical, interpretation, and/or statistical procedures used in the laboratory, (2) include practical exercises encompassing the examination of a range of samples routinely encountered in casework, (3) teach and assess the technical skills and knowledge required to perform DNA analysis … [including for analysts] the skills and knowledge required to conduct a technical review, (4) include an assessment of oral communication skills and/or a mock court exercise, and (5) include requirements for competency testing" (QAS 2020).

The QAS defines competency testing as "a test or series of tests (practical, written, and/or oral) designed to establish that an individual has demonstrated achievement of technical skills and met minimum standards of knowledge necessary to perform forensic DNA analysis" (QAS 2020). Thus, a competency test serves to inform a laboratory's technical leader whether a trainee is prepared to conduct independent casework analysis (see QAS 2020, Standard 5.2.5.4).

According to QAS 2020 Standard 6.3:

> "All analyst/technician(s), regardless of previous experience, shall successfully complete competency testing covering the routine DNA methods, interpretation, and/or statistical procedures that the analyst/technician will perform prior to participating in independent casework. Competency testing for a new analyst shall include a practical component and written and/or oral components" (QAS 2020).

There are currently no standardized specifications for competency test performance or for designing competency tests beyond having "a practical component and written and/or oral components." Each laboratory and technical leader set their own requirements for their training program, which is reviewed by external scientists in periodic assessments to the QAS along with records of competency testing.

QAS 2020 Standard 6.5 and its subcomponent continue:

> "For an analyst, currently or previously qualified within the laboratory, … the laboratory shall teach and assess the technical skills and knowledge required to

[46] https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/biologydna-scientific-area-committee

> interpret data, reach conclusions, and generate reports using the additional technology, typing test kit, platform, or interpretation software. Before use…the analyst shall successfully complete competency testing…[including] a practical component" (QAS 2020).

To address the topic of retraining when necessary, QAS 2020 Standard 6.12 and its subcomponent state:

> "The laboratory shall have and follow a policy for addressing retraining of personnel when necessary. The technical leader shall be responsible for evaluating the need for and assessing the extent of retraining. The retraining plan shall be approved by the technical leader. The individual shall successfully complete competency testing prior to his/her return to participation in casework analyses. This competency testing shall include a practical component" (QAS 2020).

Again, successful completion of a competency test with a practical component serves an important role in qualifying to begin casework, including DNA mixture interpretation. Once qualified as an analyst, technical reviewer, or technician, ongoing participation in a semi-annual external proficiency testing program is expected (see QAS 2020: Standard 13).

A competency test is intended to assess understanding of concepts and methods used at the time of an individual's initial training in DNA mixture interpretation and other aspects of DNA analysis. However, degrees of difficulty are likely to vary for competency tests administered in the hundreds of forensic DNA laboratories that exist across the U.S. and worldwide. Knowledge and skill levels of analysts will vary, which may play an important role in variation observed when interlaboratory studies are conducted (e.g., Prieto et al. 2014, Butler et al. 2018a).

We are unaware of any survey data of competency test requirements across laboratories or over time within a laboratory. If competency testing in individual laboratories continues to serve a primary role to becoming a qualified analyst in that laboratory, then it would be beneficial to have some kind of standardized competency testing to demonstrate appropriate knowledge and skill level for DNA mixture interpretation.

For continuing education, the FBI QAS requires that forensic DNA analysts, technical leaders, and technical reviewers maintain their qualifications through participation in continuing education (QAS 2011: Standard 5.1.3; QAS 2020: Standard 16.1). These individuals are encouraged to "stay abreast of topics relevant to the field of forensic DNA analysis" through attending seminars, conferences, or specific training for "at least eight (8) hours per year" and document attendance through a certificate (QAS 2020: Standard 16.1.1). Furthermore, the QAS requires that laboratories provide access to "a collection of current books, reviewed journals, or other literature applicable to DNA analysis" and that there is a documented "ongoing reading of the scientific literature" (QAS 2020: Standard 16.1.2). Individual laboratories and technical leaders determine which topics are relevant.

These minimum requirements for continuing education are a valuable starting place and enable assessment during an accreditation audit by external scientists, e.g., through

inspection of a certificate of attendance or initialed scientific articles. Accreditation audits tend to focus on practical competency (e.g., success with competency and proficiency tests) rather than assessing understanding of theoretical underpinnings and limitations of methods used. That said, interviews of analysts in which auditors ask questions related to basic theory and protocols can play a role in assessing the analyst's understanding.

Without a system of performance assessment following the continuing education activity, it is not possible to externally or uniformly evaluate what has been learned from the meeting attended or an article read by a DNA analyst.

A2.2.2. SWGDAM Training Guidelines

The 2020 SWGDAM Training Guidelines (and the previous 2013 version) encourage laboratories to develop a documented training program with a training manual and documented completion of specified tasks and competency tests (SWGDAM 2013, SWGDAM 2020). These guidelines provide a framework of information and topics to be covered, including laboratory introduction, fundamental and applied scientific knowledge, sample and/or evidence control, laboratory analysis, interpretation, reports and notifications, legal issues, and final evaluation.

Current guidance states that each laboratory is encouraged to develop a list of references "tailored to its specific needs" and to review and update the training manual each year (SWGDAM 2020). One member of our Resource Group noted:

> "From the perspective of training and continuing education, at the present time, all analysts need to know details and principles behind procedures such as DNA extraction, differential extraction, quantitative PCR, PCR, capillary electrophoresis, and mixture interpretation. Therefore, there should be a large common knowledge base [across the entire community] with a much smaller list of information that would be tailored to specific needs of a laboratory (e.g., use of a specific robotic platform or an unusual type of DNA extraction)."

The SWGDAM Training Guidelines state: "Updated references should be added to the laboratory's list during this review period or when new methodologies or technologies are incorporated into the laboratory protocols" (SWGDAM 2020). The 2013 guidelines list 98 recommended references, with 4 of these references[47] being related to DNA mixture interpretation (Buckleton & Curran 2008, Budowle et al. 2009, Gill & Buckleton 2010a, SWGDAM 2010). The 2020 guidelines list 129 references and have added 23 new articles on DNA mixture interpretation (SWGDAM 2020).

Selection of appropriate articles that are tailored to a laboratory's specific needs can be dependent on a DNA technical leader's experience and exposure. Consensus decisions from an advisory group (e.g., Butler 2013) on what knowledge would be relevant and necessary for a DNA analyst to be effective could help create a common knowledge base for the field. Developing and maintaining a centralized, online, up-to-date resource on DNA mixture

[47] These four references were determined by examining articles listed on pp. 23-24 under the Mixture Interpretation / Population Genetics / Statistics section (SWGDAM 2013).

interpretation with a relevant reference list (and electronic copies of articles, where possible) would be helpful. Given a continually growing scientific literature, it is challenging for DNA analysts to gain and maintain expert knowledge and to "stay abreast of topics relevant to the field of forensic DNA analysis" (as required by the FBI QAS, see above).

The next section covers several ideas regarding development of expert knowledge considered during our deliberations for the DNA mixture interpretation scientific foundation review.

A2.3. Considerations in Development of Expert Knowledge

Topics involving training and continuing education were discussed during several of our DNA Mixture Resource Group meetings (see Chapter 1). Information in this section came from those discussions and from ASTM Standard E2917-19 "Standard Practice for Forensic Science Practitioner Training, Continuing Education, and Professional Development Programs," which was published in February 2019 (ASTM 2019).

ASTM E2917-19 (ASTM 2019) defines ***training*** as:

> "the formal, structured process through which a forensic science practitioner reaches a level of scientific competency after acquiring the knowledge, skills, and abilities (KSAs) required to conduct specific forensic analyses"

and ***continuing education*** as:

> "the mechanism through which a forensic science practitioner increases or updates knowledge, skills, or abilities (KSAs), reinforces knowledge, or learns of the latest research, developments, or technology related to his or her profession."

DNA analysts benefit from at least three levels of expert knowledge: (1) education in basic science covering biochemistry, biology, chemistry, genetics, molecular biology, population genetics, and statistics, (2) training in forensic science and specific methods and protocols used in their laboratory to develop competency needed to perform casework, and (3) continued education and professional development to keep up-to-date as the field evolves and new methods become available.

After conducting this scientific foundation review, we believe that improvements in training and continuing education are needed to strengthen DNA mixture interpretation. Changes will be difficult without some specific funding (e.g., from the National Institute of Justice (NIJ) or the National Science Foundation (NSF)) and sustained, coordinated effort on the part of advisory groups (e.g., SWGDAM, OSAC), laboratory leadership, individual technical leaders and analysts, and the community at large (including stakeholders who use DNA results). Virtual training courses on DNA mixture interpretation could be offered by the NIJ Forensic Technology Center of Excellence[48], the Center for Statistics and Applications in Forensic Evidence (CSAFE)[49], or academic groups.

Improvements needed include:

[48] https://forensiccoe.org/
[49] https://forensicstats.org/

- An agreed upon, defined body of knowledge for DNA mixture interpretation and a means to update and remove outdated information as methods evolve
- Access to appropriate relevant literature for technical leaders and analysts
- Dedicated time in the workday to read the literature so that technical leaders and analysts can keep up to date with developments
- Uniformly documented knowledge assessment
- A method to acknowledge competence in a specific area to allow true expertise in testimony (e.g., DNA transfer and activity assessments, see van Oorschot et al. 2019)
- Training for technical leaders in experimental design and data analysis to assist with validation studies and protocol development.

Additional thoughts on these needs arising from deliberations and discussions with our Resource Group are included below.

A2.3.1. A Defined Body of Knowledge

There should be a defined standard body of knowledge for a DNA analyst to have a shared understanding with others in the field. This defined body of knowledge should include use of a consistent vocabulary with agreed upon terminology.

This defined body of knowledge should be monitored and updated by a group which functions independently of forensic laboratories but gathers input from these laboratories. An important part of any scientific effort is to understand and build upon previous documented work in the field. Such a body of knowledge could include foundational and historical literature, validation literature, and current literature. Defining what analysts *need to know* rather than what they *need to do* will increase confidence and enhance practice. References cited in this report can serve as a useful starting point as can a textbook like *Fundamentals of Forensic DNA Typing* (Butler 2009).

Lists of relevant articles in specific areas of interest to forensic casework analysts could be created from quality literature reviews exploring the breadth and depth of DNA mixture interpretation topics. Such listings of recommended articles in particular areas will be subjective and require ongoing curation to remain relevant. And maintenance will be an ongoing challenge. For example, the NIST STRBase[50] has provided some literature references on mixture interpretation, but even these lists are not up-to-date and do not contain many of the references utilized in our foundation review.

A2.3.2. Literature Awareness, Access, and Acumen

DNA technical leaders and analysts would benefit from receiving regular updates on relevant and available articles. Literature awareness and exposure to general forensic science articles can be obtained through voluntary community efforts, such as Forensic Library Service Bureau emails (flsblibrary@wsp.wa.gov) organized by Jeff Teitelbaum of the Washington State Patrol Forensic Laboratory Services Bureau (Seattle, WA). However, an additional gap

[50] https://strbase.nist.gov/mixture.htm

can exist in accessing articles of interest. A new service directly focusing on forensic DNA mixture interpretation topics could be helpful, particularly if the associated articles were made available without violating publisher copyrights or embargoes. Ongoing funding and continued commitment to creating and maintaining a national library service is needed. This will be important for success in this endeavor as will regular, active participation from DNA casework analysts with support from their laboratory management.

Access to appropriate and relevant literature can be challenging given a growing body of knowledge coming from a variety of active researchers. The laboratory should arrange for access to relevant journals. Partnering with a university could be a way to address this need.

The American Academy of Forensic Sciences (AAFS) provides access for individual members to the *Journal of Forensic Sciences* while membership in the International Society for Forensic Genetics (ISFG) provides access to *Forensic Science International: Genetics*. Other peer-reviewed journals with relevant information on DNA mixture interpretation include *Forensic Science International, International Journal of Legal Medicine, Legal Medicine, Science & Justice, PLOS ONE* (open access)*, Investigative Genetics* (open access, no longer active)*, Frontiers in Genetics* (open access)*, Electrophoresis, Croatian Medical Journal* (open access)*, and Law, Probability & Risk.* For journals that are not open access, it would be helpful for funding agencies like NIJ to support researchers' use of grant funding to cover open access fees and make their published work accessible to all.

To ensure maximum value is obtained from the scientific literature, DNA technical leaders and analysts could benefit from training on effective searching and reading of the literature (e.g., Butler 2016). Academic researchers should be encouraged to assist in this effort with the support of funding agencies, such as NIJ and NSF.

To help begin the process of identifying the most valuable publications in the field, a literature list was prepared and a workshop presented at the AAFS 2021 meeting titled "MVPs of Forensic DNA: Examining the Most Valuable Publications in the Field." A literature list with 497 articles in 26 categories along with explanatory slides is available on the NIST STRBase website[51]. A precursor of this literature list has also been adopted by the OSAC Biology Scientific Area Committee as informative literature for forensic biology and DNA[52].

A2.3.3. Knowledge Assessment

Practical work, a written competency exam, and an oral competency exam are important in assessing knowledge for various aspects of the role of a DNA analyst. These roles include laboratory work, report writing, and court testimony. Assessment methods with an appropriate level of difficulty are needed with a defined score required for passing and a policy agreed upon by laboratory management regarding remediation when an individual fails an assessment.

[51] See https://strbase.nist.gov/pub_pres/AAFS2021-W19-Handouts.pdf
[52] See https://www.nist.gov/osac/biology-scientific-area-committee

Proficiency tests, along with regular intralaboratory and interlaboratory tests in which analysts evaluate the same DNA mixture sample and/or profile, can identify differences in analyst interpretation and understanding of concepts. Additionally, such tests inform what types of specific training would be helpful within a laboratory or across the community in general. Self-organized regional interlaboratory studies and discussion groups could be useful to identify training gaps and needs while remaining relatively inexpensive.

A2.3.4. Additional Thoughts on Training and Continuing Education

Training is an ongoing process rather than a singular event when someone begins employment. Theory-based information and training should involve moving from simple to complex concepts.

Both individual and group training (e.g., independent study and team exercises) are necessary because people learn differently. Review of validation studies and the basis for laboratory-specific protocol development should be part of a training program. Training should include case assessment, critical thinking in interpretation, and report writing (Cook et al. 1998a), as well as understanding the hierarchy of propositions to appreciate what questions are being addressed in casework (Cook et al. 1998b).

The community should have access to online training modules covering topics in DNA mixture interpretation that could be taught via regularly scheduled webinars organized on a national level. In this manner, a large number of people could be trained on fundamental topics, and key articles and information could be covered. This type of online training platform was used to reach several hundred DNA analysts during May, June, and July 2019 in a series of eight webinars on probabilistic genotyping[53] coordinated by the FBI Laboratory and the NIJ's Forensic Technology Center of Excellence (see Table A1.3 in Appendix 1). Effective training must be coupled with time for study and subsequent demonstration of knowledge assessment to evaluate a learner's level of understanding. A certificate of attendance by itself ***is not sufficient*** for demonstrating that training or continuing education materials have been understood.

Dedicated time in the workday is needed for professional development, which is defined by ASTM E2917-19 (ASTM 2019) as:

> "the mechanism through which a forensic science practitioner improves personal skills, successfully handles increasing responsibility, makes contributions to the profession, and reinforces ethical behaviors."

Professional development includes continuing education and knowledge of the scientific literature. ASTM E2917-19 6.3.3.1 requires mechanisms "for the documented review of scientific literature" and 6.4.3 states that "continuing education and professional development can be delivered in-person, online, self-directed or computer-based" (ASTM 2019). If forensic casework analysts are expected to keep up to date with new developments in DNA mixture interpretation, some portion of their paid time should be devoted to examining

[53] Probabilistic Genotyping of Evidentiary DNA Typing Results – An Online Workshop Series: https://forensiccoe.org/webinar/online-workshop-series-probabilistic-genotyping-of-evidentiary-dna-typing-results/ (accessed May 27, 2020).

relevant books and articles published in the scientific literature. We discussed the benefits of a suggested 5% of paid time or two hours each week. ASTM E2917-19 6.1.1 requires "an annual average of at least 16 hours…over a three-year period" and emphasizes the need for management and their parent agency to "provide support and opportunities for this continuing professional development" (ASTM 2019).

A2.3.5. Specialized Training for DNA Technical Leaders

Being a DNA technical leader is a hard job. Given the responsibilities that they have under the FBI QAS requirements, technical leaders would benefit from additional training to design appropriate validation studies. This includes assessing, for example, probabilistic genotyping software and next-generation sequencing technologies. Training on design of validation experiments and statistical analysis could focus on types of controls, materials to test, and impacts of varying numbers of samples for testing.

Technical leaders need to be ahead of their DNA analysts in their knowledge to effectively assess and train analysts within their laboratories. As methods become more sophisticated, additional training in statistics and data analysis would be helpful. Many technical leaders also have a supervisory role and would benefit from management training to strengthen their skill sets in these areas.

It is not realistic to expect a technical leader who received a master's degree 10 to 15 years ago to use/adopt probabilistic genotyping, next-generation sequencing, or any new technology with only a week or two of training. It requires an extended period of time to learn and digest new information and practice new leadership skills in performing the functions of a technical leader.

A2.3. Future Considerations

FUTURE CONSIDERATIONS #A2.1: It would be helpful for the community (or advisory groups) to define the minimum standards of knowledge necessary to perform DNA mixture interpretation and to provide further guidance on competency test design.

FUTURE CONSIDERATIONS #A2.2: It would be beneficial to standardize competency testing to demonstrate appropriate knowledge and skill level for DNA mixture interpretation.

FUTURE CONSIDERATIONS #A2.3: With an evolving and complex field like DNA mixture interpretation, further guidance on what should be studied and understood for foundational knowledge would be helpful not only for ongoing learning within forensic laboratories, but also in academic programs seeking to prepare students to participate in the field.

FUTURE CONSIDERATIONS #A2.4: Consensus decisions from an advisory group on what knowledge would be relevant and necessary for a DNA analyst to be effective could help create a common knowledge base for the field.

FUTURE CONSIDERATIONS #A2.5: Developing and maintaining a centralized, online, up-to-date resource on DNA mixture interpretation with a relevant reference list (and electronic copies of articles, where possible) would be helpful.

FUTURE CONSIDERATIONS #A2.6: A culture of critical thinking and clear communication regarding DNA mixture interpretation is crucial as probabilistic genotyping software programs are implemented and as the appropriate relevance of results from low-level, complex mixtures are shared in written reports and court testimony.

FUTURE CONSIDERATIONS #A2.7: Technical leaders should ensure that analysts are familiar with fundamental principles and the complications of DNA mixtures before probabilistic genotyping software tools are employed.

FUTURE CONSIDERATIONS #A2.8: Some portion of DNA analysts' paid time should be devoted to examining relevant books and articles published in the scientific literature. DNA technical leaders would benefit from training on how to design validation experiments and perform data analysis.

Bibliography

AAFS 2008 DNA Mixture Workshop: https://strbase.nist.gov/training/AAFS2008_MixtureWorkshop.htm

AAFS 2011 DNA Mixture Workshop: https://strbase.nist.gov/training/AAFS2011_MixtureWorkshop.htm

[AFSP 2009] Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice* 49:161-164.

Aitken C, Nordgaard A, Taroni F, Biedermann A (2018) Commentary: Likelihood Ratio as Weight of Forensic Evidence: A Closer Look. *Frontiers in Genetics* 9:224.

Aiken C, Nordgaard A (2018) Letter to the Editor-The roles of participants' differing background information in the evaluation of evidence. *Journal of Forensic Sciences* 63(2):648-649.

Alfonse LE, Tejada G, Swaminathan H, Lun DS, Grgicak CM (2017) Inferring the number of contributors to complex DNA mixtures using three methods: exploring the limits of low-template DNA interpretation. *Journal of Forensic Sciences* 62(2):308-316.

Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM (2018) A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Science International: Genetics* 32:62-70.

Alladio E, Omedei M, Cisana S, D'Amico G, Caneparo D, Vincenti M, Garofano P (2018) DNA mixtures interpretation - A proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples. *Forensic Science International: Genetics* 37:143-150.

Almirall J, Arkes H, Lentini J, Mowrer F, Pawliszyn J (2017) *Forensic Science Assessments: A Quality and Gap Analysis – Fire Investigation.* American Association for the Advancement of Science. Available at https://www.aaas.org/resources/fire-investigation (accessed June 24, 2020).

Alonso A, Müller P, Roewer L, Willuweit S, Budowle B, Parson W (2017) European survey on forensic applications of massively parallel sequencing. *Forensic Science International: Genetics* 29:e23-e25.

Alonso A, Barrio PA, Müller P, Köcher S, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B (2018) Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis* 39:2655-2668.

ANAB (2019) ISO/IEC 17025:2017 – Forensic Science Testing and Calibration Laboratories – Accreditation Requirements (AR 3125 2019/04/29). Available at https://anab.qualtraxcloud.com/ShowDocument.aspx?ID=12371 (accessed June 5, 2020).

ANSI/ASB Standard 020 (2018) Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol. Available at https://asb.aafs.org/wp-content/uploads/2018/09/020_Std_e1.pdf (accessed August 12, 2020).

ANSI/ASB Standard 040 (2019) Standard for Forensic DNA Interpretation and Comparison Protocols. Available at https://www.asbstandardsboard.org/wp-content/uploads/2019/10/Std_040_e1.pdf (accessed August 12, 2020).

ANSI/ASB Standards 018 (2020) Standard for Validation of Probabilistic Genotyping Systems. Available at http://www.asbstandardsboard.org/wp-content/uploads/2020/07/018_Std_e1.pdf (accessed August 12, 2020).

ANZPAA (2017) An introductory guide to evaluative reporting. http://www.anzpaa.org.au/forensic-science/our-work/products/publications.

Aranda R (2015) A Large-scale Study of DNA Mixture Interpretation: Inter- and Intra-Laboratory Variability. *Proceedings of the 2015 International Symposium on Forensic Science Error Management* (NIST Special Publication 1206) Editor: JM Butler. Abstract available at https://doi.org/10.6028/NIST.SP.1206 (see p. 60) and slides available at https://www.nist.gov/system/files/documents/2016/11/22/large_scale_study_of_dna_mixture_interpretation.aranda-iv.crim1_.pdf. (accessed 13 November 2020).

ASTM (2019) E2917-19 Standard Practice for Forensic Science Practitioner Training, Continuing Education, and Professional Development Programs. Available at https://www.astm.org/Standards/E2917.htm.

Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* 64(2-3):125-140.

Balding DJ, Buckleton J (2009) Interpreting low template DNA profiles. *Forensic Science International: Genetics* 4(1):1-10.

Balding DJ (2013) Evaluation of mixed-source, low-template DNA profiles in forensic science. *Proceedings of the National Academy of Sciences of the United States of America* 110(30):12241-12246.

Ballantyne J, Hanson EK, Perlin MW (2013) DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser capture cell populations: combining quantitative data for greater identification information. *Science & Justice* 53:103-114.

Barber M, Evett I, Pope S, Sullivan K, Whitaker J (2015) HOS/13/037 Forensic Science Regulator Codes of Practice and Conduct – Proficiency testing guidance for DNA mixtures. Final report. January 2015. Principal Forensic Services Ltd.

Barrio PA, Crespillo M, Luque JA, Aler M, Baeza-Richer C, Baldassarri L, Carnevali E, Coufalova P, Flores I, García O, García MA, González R, Hernández A, Inglés V, Luque GM, Mosquera-Miguel A, Pedrosa S, Pontes ML, Porto MJ, Posada Y, Ramella MI, Ribeiro T, Riego E, Sala A, Saragoni VG, Serrano A, Vannelli S (2018) GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: results and evaluation. *Forensic Science International: Genetics* 35: 156-163.

Basset P, Castella V (2018) Lessons from a study of DNA contaminations from police services and forensic laboratories in Switzerland. *Forensic Science International: Genetics* 33:147-154.

Basset, P, Castella, V (2019) Positive impact of DNA contamination minimization procedures taken within the laboratory. *Forensic Science International: Genetics* 38:232-235.

Bauer DW, Butt N, Hornyak JM, Perlin MW (2020) Validating TrueAllele® interpretation of DNA mixtures containing up to ten unknown contributors. *Journal of Forensic Sciences* 65(2):380-398.

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions Royal Society of London* 53: 370-418.

Bennett L, Oldoni F, Long K, Cisana S, Maddela K, Wootton S, Chang J, Hasegawa R, Lagace R, Kidd KK, Podini D (2019) Mixture deconvolution by massively parallel sequencing of microhaplotypes. *International Journal of Legal Medicine* 133(3):719-729.

Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA, and Sijen, T. (2011) Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International: Genetics* 5(4): 316-328.

7422 Benschop CC, Haned H, de Blaeij TJ, Meulenbroek AJ, Sijen T (2012) Assessment of mock cases involving
7423 complex low template DNA mixtures: A descriptive study. *Forensic Science International: Genetics* 6(6):697-
7424 707.
7425
7426 Benschop C, Haned H, Sijen T (2013) Consensus and pool profiles to assist in the analysis and interpretation
7427 of complex low template DNA mixtures. *International Journal of Legal Medicine* 127(1):11-23.
7428
7429 Benschop CC, Sijen T (2014) LoCIM-tool: An expert's assistant for inferring the major contributor's alleles in
7430 mixed consensus DNA profiles. *Forensic Science International: Genetics* 11:154-165.
7431
7432 Benschop CC, Haned H, Yoo SY, Sijen T (2015a) Evaluation of samples comprising minute amounts of DNA.
7433 *Science & Justice* 55(5): 316-322.
7434
7435 Benschop CC, Haned H, Jeurissen L, Gill PD, Sijen T (2015b) The effect of varying the number of
7436 contributors on likelihood ratios for complex DNA mixtures. *Forensic Science International: Genetics* 19:92-
7437 99.
7438
7439 Benschop CC, Connolly E, Ansell R, Kokshoorn B (2017a) Results of an inter and intra laboratory exercise on
7440 the assessment of complex autosomal DNA profiles. *Science & Justice* 57(1): 21-27.
7441
7442 Benschop CCG, van de Merwe L, de Jong J, Vanvooren V, Kempenaers M, van der Beek CP, Barni F, Reyes
7443 EL, Moulin L, Pene L, Haned H, Sijen T (2017b) Validation of SmartRank: A likelihood ratio software for
7444 searching national DNA databases with complex DNA profiles. *Forensic Science International: Genetics*
7445 29:145-153.
7446
7447 Benschop CCG, Nijveld A, Duijs FE, Sijen T (2019a) An assessment of the performance of the probabilistic
7448 genotyping software EuroForMix: Trends in likelihood ratios and analysis of Type I & II errors.
7449 *Forensic Science International: Genetics* 42:31-38.
7450
7451 Benschop CCG, Hoogenboom J, Hovers P, Slagter M, Kruise D, Parag R, Steensma K, Slooten K, Nagel JHA,
7452 Dieltjes P, van Marion V, van Paassen H, de Jong J, Creeten C, Sijen T, Kneppers ALJ (2019b)
7453 DNAxs/DNAStatistX: Development and validation of a software suite for the data management and
7454 probabilistic interpretation of DNA profiles. *Forensic Science International: Genetics* 42:81-89.
7455
7456 Benschop CCG, Hoogenboom J, Bargeman F, Hovers P, Slagter M, van der Linden J, Parag R, Kruise D,
7457 Drobnic K, Klucevsek G, Parson W, Berger B, Laurent FX, Faivre M, Ulus A, Schneider P, Bogus M,
7458 Kneppers ALJ, Sijen T (2020) Multi-laboratory validation of DNAxs including the statistical library
7459 DNAStatistX. *Forensic Science International: Genetics* 49:102390.
7460
7461 Berger CEH, Slooten K (2016) The LR does not exist. *Science & Justice* 56: 388-391.
7462
7463 Bhoelai B, de Jong B, Sijen T (2013) Can mRNA markers distinguish traces generated by different types of
7464 contact? *Forensic Science International: Genetics Supplement Series* 4(1):e3–e4.
7465
7466 Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD (2016) Evaluation of forensic DNA mixture
7467 evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of
7468 inclusion. *BMC Genetics* 17(1):125.
7469
7470 Biedermann A, Taroni F (2012) Bayesian networks for evaluating forensic DNA profiling evidence: A review
7471 and guide to literature. *Forensic Science International: Genetics* 6(2):147-157.
7472
7473 Biedermann A, Bozza S, Taroni F, Aitken C (2016a) Reframing the debate: a question of probability, not of
7474 likelihood ratio. *Science & Justice* 56:392-396.
7475

Biedermann A, Champod C, Jackson G, Gill P, Taylor D, Butler J, Morling N, Hicks T, Vuille J, Taroni F (2016b) Evaluation of forensic DNA traces when propositions of interest relate to activities: Analysis and discussion of recurrent concerns. *Frontiers in Genetics* 7:215. https://doi.org/10.3389/fgene.2016.00215.

Bill M, Gill P, Curran J, Clayton T, Pinchin R, Healy M, Buckleton J (2005) PENDULUM – a guideline-based approach to the interpretation of STR mixtures. *Forensic Science International* 148:181-189.

Bille T, Bright JA, Buckleton J (2013) Application of random match probability calculations to mixed STR profiles. *Journal of Forensic Sciences* 58(2):474-485.

Bille TW, Weitz SM, Coble MD, Buckleton J, Bright JA (2014) Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis* 35:3125-3133.

Bille T, Weitz S, Buckleton JS, Bright JA (2019) Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors. *Forensic Science International: Genetics* 40:150-159.

Blake E, Mihalovich J, Higuchi R, Walsh PS, Erlich H (1992) Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQ alpha oligonucleotide typing on biological evidence samples: casework experience. *Journal of Forensic Sciences* 37(3):700-726.

Blastland M, Freeman ALJ, van der Linden S, Marteau TM, Spiegelalter D (2020) Five rules for evidence communication. *Nature* 587:362-364.

Bleka Ø, Storvik G, Gill P (2016a) EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics* 21:35-44.

Bleka Ø, Benschop CC, Storvik G, Gill P (2016b) A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. *Forensic Science International: Genetics* 25:85-96.

Bleka Ø, Prieto L, Gill P (2019) CaseSolver: An investigative open source expert system based on EuroForMix. *Forensic Science International: Genetics* 41:83-92.

Bolivar PA, Tracey M, McCord B (2016) Assessing the risk of secondary transfer via fingerprint brush contamination using enhanced sensitivity DNA analysis methods. *Journal of Forensic Sciences* 61(1):204–211.

Bonsu DOM, Higgins D, Austin JJ (2020) Forensic touch DNA recovery from metal surfaces - A review. *Science & Justice* 60(3):206-215.

Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N (2014) Evaluation of the Ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Science International: Genetics* 12:144-154.

Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics* 18:78-89.

Bose N, Carlberg K, Sensabaugh G, Erlich H, Calloway C (2018) Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples. *Forensic Science International: Genetics* 34:186-196.

Bouzga MM, Dorum G, Gundersen K, Kohler P, Hoff-Olsen P, Fonnelop AE (2020) Is it possible to predict the origin of epithelial cells? - A comparison of secondary transfer of skin epithelial cells versus vaginal mucous membrane cells by direct contact. *Science & Justice* 60(30):234-242.

Bowman ZE, Mosse KSA, Sungaila AM, van Oorschot RAH, Hartman D (2018) Detection of offender DNA following skin-to-skin contact with a victim. *Forensic Science International: Genetics* 37:252-259.

Bozzo WR, Colussi AG, Ortíz MI, Laborde L, Pilili JP, Carini G, Lojo MM (2015) Analysis of DNA from fingernail samples in criminal cases. *Forensic Science International: Genetics Supplement Series* 5:e601-e602.

Brayley-Morris H, Sorrell A, Revoir AP, Meakin GE, Syndercombe Court D, Morgan RM (2015) Persistence of DNA from laundered semen stains: Implications for child sex trafficking cases. *Forensic Science International: Genetics* 19:165-171.

Breathnach M, Williams L, McKenna L, Moore E (2016) Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment. *Forensic Science International: Genetics* 20:53-60.

Bregu J, Conklin D, Coronado E, Terrill M, Cotton RW, Grgicak CM (2013) Analytical thresholds and sensitivity: establishing RFU thresholds for forensic DNA analysis. *Journal of Forensic Sciences* 58(1):120-129.

Brenner CH, Fimmers R, Baur MP (1996) Likelihood ratios for mixed stains when the number of donors cannot be agreed. *International Journal of Legal Medicine* 109(4):218-219.

Bright JA, Curran JM (2014) Investigation into stutter ratio variability between different laboratories. *Forensic Science International: Genetics* 13:79-81.

Bright JA, Taylor D, Curran J, Buckleton J (2014) Searching mixed DNA profiles directly against profile databases. *Forensic Science International: Genetics* 9:102-110.

Bright JA, Evett IW, Taylor D, Curran JM, Buckleton J (2015) A series of recommended tests when validating probabilistic DNA profile interpretation software. *Forensic Science International: Genetics* 14:125-131.

Bright JA, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, Buckleton J (2016) Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics* 23:226-239.

Bright JA, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, McWhorter A, Ciecko A, Peck B, Baumgartner C, Buettner C, McWilliams S, McKenna C, Gallacher C, Mallinder B, Wright D, Johnson D, Catella D, Lien E, O'Connor C, Duncan G, Bundy J, Echard J, Lowe J, Stewart J, Corrado K, Gentile S, Kaplan M, Hassler M, McDonald N, Hulme P, Oefelein RH, Montpetit S, Strong M, Noel S, Malsom S, Myers S, Welti S, Moretti T, McMahon T, Grill T, Kalafut T, Greer-Ritzheimer MM, Beamer V, Taylor DA, Buckleton JS (2018) Internal validation of STRmix™– a multi laboratory response to PCAST. *Forensic Science International: Genetics* 34:11-24.

Bright JA, Cheng K, Kerr Z, McGovern C, Kelly H, Moretti TR, Smith MA, Bieber FR, Budowle B, Coble MD, Alghafri R, Allen PS, Barber A, Beamer V, Buettner C, Russell M, Gehrig C, Hicks T, Charak J, Cheong-Wing K, Ciecko A, Davis CT, Donley M, Pedersen N, Gartside B, Granger D, Greer-Ritzheimer M, Reisinger E, Kennedy J, Grammer E, Kaplan M, Hansen D, Larsen HJ, Laureano A, Li C, Lien E, Lindberg E, Kelly C, Mallinder B, Malsom S, Yacovone-Margetts A, McWhorter A, Prajapati SM, Powell T, Shutler G, Stevenson K, Stonehouse AR, Smith L, Murakami J, Halsing E, Wright D, Clark L, Taylor DA, Buckleton J (2019a) STRmix™ collaborative exercise on DNA mixture interpretation. *Forensic Science International: Genetics* 40:1-8.

Bright JA, Taylor D, Kerr Z, Buckleton J, Kruijver M (2019b) The efficacy of DNA mixture to mixture matching. *Forensic Science International: Genetics* 41:64-71.

Bright J-A, Coble M (2020) *Forensic DNA Profiling: A Practical Guide to Assigning Likelihood Ratios*. CRC Press: Boca Raton.

Brookes C, Bright JA, Harbison S, Buckleton J (2012) Characterising stutter in forensic STR multiplexes. *Forensic Science International: Genetics* 6(1):58-63.

Buckingham AK, Harvey ML, van Oorschot RAH (2016) The origin of unknown source DNA from touched objects. *Forensic Science International: Genetics* 25:26–33.

Buckleton JS, Evett IW, Weir BS (1998) Setting bounds for the likelihood ratio when multiple hypotheses are postulated. *Science & Justice* 38(1):23-26.

Buckleton J, Triggs CM, Walsh SJ (Editors) (2005) *Forensic DNA Evidence Interpretation*. CRC Press: Boca Raton.

Buckleton JS, Curran JM, Gill P (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *Forensic Science International: Genetics* 1(1):20-28.

Buckleton J, Curran J (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Science International: Genetics* 2(4):343-348.

Buckleton J (2009) Validation issues around DNA typing of low level DNA. *Forensic Science International: Genetics* 3(4):255-260.

Buckleton J, Bright JA, Taylor D, Evett I, Hicks T, Jackson G, Curran JM (2014) Helping formulate propositions in forensic DNA analysis. *Science & Justice* 54(4):258-261.

Buckleton JS, Bright J-A, Taylor D (Editors) (2016) *Forensic DNA Evidence Interpretation, Second Edition*. CRC Press: Boca Raton.

Buckleton JS, Bright J-A, Cheng K, Budowle B, Coble MD (2018) NIST Interlaboratory studies involving DNA mixtures (MIX13): a modern analysis. *Forensic Science International: Genetics* 37:172-179.

Buckleton JS, Bright J-A, Gittelson S, Moretti TR, Onorato AJ, Bieber FR, Budowle B, Taylor DA (2019) The probabilistic genotyping software STRmix: utility and evidence for its validity. *Journal of Forensic Sciences* 64(2):393-405.

Budowle B, Lindsey JA, DeCou JA, Koons BW, Giusti AM, Comey CT (1995) Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQ alpha using a multiplex amplification and typing procedure. *Journal of Forensic Sciences* 40(1):45-54.

Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification* (pp. 73–88). Madison, Wisconsin: Promega Corporation. Available at https://www.promega.com/-/media/files/resources/conference-proceedings/ishi-02/oral-presentations/17.pdf?la=en (accessed June 20, 2020).

Budowle B, Hobson DL, Smerick JB, Smith JAL (2001) Low copy number – consideration and caution. *Proceedings of the Twelfth International Symposium on Human Identification*. Madison, WI: Promega Corporation. Available at https://promega.media/-/media/files/resources/conference-proceedings/ishi-12/oral-presentations/budowle.pdf?la=en (accessed August 12, 2020).

Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences* 54(4):810-821.

Burrill J, Daniel B, Frascione N (2019) A review of trace "touch DNA" deposits: Variability factors and an exploration of cellular composition. *Forensic Science International: Genetics* 39:8-18.

Burrill J, Daniel B, Frascione N (2020) Illuminating touch deposits through cellular characterization of hand rinses and body fluids with nucleic acid fluorescence. *Forensic Science International: Genetics* 46:102269.

Butler JM, Buel E, Crivellente F, McCord BR (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25:1397-1412.

Butler JM (2007) Short tandem repeat typing technologies used in human identity testing. *Biotechniques* 43(4):Sii-Sv.

Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine and Pathology* 3(3):200-205.

Butler JM (2009) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego.

Butler JM, Hill CR (2010) Scientific issues with analysis of low amounts of DNA. *Profiles in DNA,* 13(1). Available at https://www.promega.com/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/ (accessed June 20, 2020).

Butler JM (2012) *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Academic Press: San Diego.

Butler JM (2013) Forensic DNA advisory groups: DAB, SWGDAM, ENFSI, and BSAG. In *Encyclopedia of Forensic Sciences, Second Edition*, pp. 339-343. Elsevier: San Diego.

Butler JM (2015a) *Advanced Topics in Forensic DNA Typing: Interpretation*. Elsevier Academic Press: San Diego.

Butler JM (2015b) The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B* 370:20140252.

Butler JM (2015c) U.S. initiatives to strengthen forensic science & international standards in forensic DNA. *Forensic Science International: Genetics* 18:4-20.

Butler JM (2016) Why search and read the forensic science literature? Presentation given during American Academy of Forensic Sciences workshop "Information Does Exist Beyond the First Page of Your Google Search!" Available at https://strbase.nist.gov/training/2_AAFS2016-W1-WhyRead.pdf (slides for other presentations in the workshop are available at https://strbase.nist.gov/training/AAFS2016_LiteratureWorkshop.htm).

Butler JM, Kline MC, Coble MD (2018a) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. *Forensic Science International: Genetics* 37: 81-94.

Butler JM, Iyer H, Press R, Taylor M, Vallone PM, Willis S (2018b) DNA mixture interpretation principles: insights from the NIST scientific foundation review. *Proceedings of the 29th International Symposium on Human Identification.* Available at https://promega.media/-/media/files/products-and-services/genetic-identity/ishi-29-oral-abstracts/butler.pdf (accessed June 24, 2020).

Butler JM, Iyer H, Press R, Taylor M, Vallone PM, Willis S (2019) NIST scientific foundation review on DNA mixture interpretation. Poster presented at the International Society for Forensic Genetics. Available at https://strbase.nist.gov/pub_pres/ButlerISFG2019poster.pdf (accessed June 24, 2020).

Butler JM, Willis S (2020) INTERPOL review of forensic biology and forensic DNA typing 2016-2019. *Forensic Sci. Int.: Synergy* 2:352-367. Presentation available at https://strbase.nist.gov/pub_pres/Butler-INTERPOL-DNAreview-Oct2019.pdf (accessed June 24, 2020).

Cale CM, Earll ME, Latham KE, Bush GL (2016) Could secondary DNA transfer falsely place someone at the scene of a crime? *Journal of Forensic Sciences* 61(1):196-203.

Casey DG, Clayson N, Jones S, Lewis J, Boyce M, Fraser I, Kennedy F, Alexender K (2016) Correspondence: A response to Meakin and Jamieson DNA transfer: Review and implications for casework. *Forensic Science International: Genetics* 21:117-118.

Caskey CT, Edwards A, Hammond HA (1989) DNA: the history and future use in forensic analysis. *Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis*. U.S. Government Printing Office: The Laboratory Division of the Federal Bureau of Investigation (June 19-23, 1989), pp. 3-9.

Castella V, Mangin P (2008) DNA profiling success and relevance of 1739 contact stains from caseworks. *Forensic Science International: Genetics Supplement Series* 1:405-407.

Cerri N, Verzeletti A, Cortellini V, Cincotta A, De Ferrari F (2009) Prevalence of mixed DNA profiles in fingernail swabs from autoptic cases. *Forensic Science International: Genetics Supplement Series* 2(1):163-164.

Chambers GK, Cordiner SJ, Buckleton JS, Robertson B, Vignaux GA (1997) Forensic DNA profiling: the importance of giving accurate answers to the right questions. *Criminal Law Forum* 8(3): 445-459.

Champod C (2013) DNA transfer: Informed judgment or mere guesswork? *Frontiers in Genetics* 4:300. https://doi.org/10.3389/fgene.2013.00300.

Chan Mun Wei J, Zhao Z, Li SC, Ng YK (2018) NGS-based likelihood ratio for identifying contributors in two- and three-person DNA mixtures. *Computational Biology and Chemistry* 74:428-433.

Chen P, Yin C, Li Z, Pu Y, Yu Y, Zhao P, Chen D, Liang W, Zhang L, Chen F (2018) Evaluation of the microhaplotypes panel for DNA mixture analyses. *Forensic Science International: Genetics* 35:149-155.

Clayton TM, Whitaker JP, Sparkes R, Gill P (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International* 91(1):55-70.

Clayton TM, Guest JL, Urquhart AJ, Gill PD (2004) A genetic basis for anomalous band patterns encountered during DNA STR profiling. *Journal of Forensic Sciences* 49(6):1207-1214.

Coble MD, Bright JA, Buckleton JS, Curran JM (2015) Uncertainty in the number of contributors in the proposed new CODIS set. *Forensic Science International: Genetics* 19:207-211.

Coble MD, Buckleton J, Butler JM, Egeland T, Fimmers R, Gill P, Gusmao L, Guttman B, Krawczak M, Morling N, Parson W, Pinto N, Schneider PM, Sherry ST, Willuweit S, Prinz M (2016) DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. *Forensic Science International: Genetics* 25:191-197.

Coble MD, Bright J-A (2019) Probabilistic genotyping software: an overview. *Forensic Science International: Genetics* 38:219-224.

Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit. *Journal of Forensic Sciences* 49(6):1265-1277.

Comey CT, Budowle B. (1991) Validation studies on the analysis of the HLA DQ alpha locus using the polymerase chain reaction. *Journal of Forensic Sciences* 36(6):1633-1648.

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998a) A model for case assessment and interpretation. *Science & Justice* 38(3):151-156.

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998b) A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice* 38(4):231-239.

Cooper S, McGovern C, Bright JA, Taylor D, Buckleton J (2015) Investigating a common approach to DNA profile interpretation using probabilistic software. *Forensic Science International: Genetics* 16:121-131.

Cotton EA, Allsop RF, Guest JL, Frazier RR, Koumi P, Callow IP, Seager A, Sparkes RL (2000) Validation of the AMPFlSTR SGM plus system for use in forensic casework. *Forensic Science International* 112(2-3):151-161.

Cowell RG (2009) Validation of an STR peak area model. *Forensic Science International: Genetics* 3(3):193-199.

Cowell RG, Graversen T, Lauritzen SL, Mortera J (2015) Analysis of forensic DNA mixtures with artefacts. *Journal of the Royal Statistical Society Applied Statistics Series C* 64(1):1-48. Available at https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12071 (accessed July 20, 2020).

Crespillo M, Barrio PA, Luque JA, Alves C, Aler M, Alessandrini F, Andrade L, Barretto RM, Bofarull A, Costa S, García MA, García O, Gaviria A, Gladys A, Gorostiza A, Hernández A, Herrera M, Hombreiro L, Ibarra AA, Jiménez MJ, Luque GM, Madero P, Martínez-Jarreta B, Masciovecchio MV, Modesti NM, Moreno F, Pagano S, Pedrosa S, Plaza G, Prat E, Puente J, Rendo F, Ribeiro T, Sala A, Santamaría E, Saragoni VG, Whittle MR (2014) GHEP-ISFG collaborative exercise on mixture profiles of autosomal STRs (GHEP-MIX01, GHEP-MIX02 and GHEP-MIX03): results and evaluation. *Forensic Science International: Genetics* 10:64-72.

Crow JF (1999) Hardy, Weinberg and language impediments. *Genetics* 152:821-825.

Curran JM, Triggs CM, Buckleton J, Weir BS (1999) Interpreting DNA mixtures in structured populations. *Journal of Forensic Sciences* 44(5):987-995.

Curran JM (2008) A MCMC method for resolving two person mixtures. *Science & Justice* 48(4):168-177.

Curran JM (2016) Admitting to uncertainty in the LR. *Science & Justice* 56:380-382.

Daly DJ, Murphy C, McDermott SD (2012) The transfer of touch DNA from hands to glass, fabric and wood. *Forensic Science International: Genetics* 6:41-46.

Dang HW, Mao J, Wang H, Huang JP, Bai XG (2012) Research status and prospects of DNA test on difficult specimens. *Fa Yi Xue Za Zhi* 28:52-54.

Daniel R, van Oorschot RAH (2011) An investigation of the presence of DNA on unused laboratory gloves. *Forensic Science International: Genetics Supplement Series* 3:e45-e46.

Davies C, Thomson J, Kennedy F (2015) Assessing primary, secondary and tertiary DNA transfer using the Promega ESI-17 Fast PCR chemistry. *Forensic Science International: Genetics Supplement Series* 5:e55-e57.

Dean L, Kwon YJ, Philpott MK, Stanciu CE, Seashols-Williams SJ, Cruz TD, Sturgill J, Ehrhardt CJ (2015) Separation of uncompromised whole blood mixtures for single source STR profiling using fluorescently-labeled human leukocyte antigen (HLA) probes and fluorescence activated cell sorting (FACS). *Forensic Science International: Genetics* 17:8-16.

Debernardi A, Suzanne E, Formant A, Pène L, Dufour AB, Lobry JR (2011) One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFℓSTR© Identifiler© STR kit. *Forensic Science International: Genetics* 5(1):43-49.

de Koeijer JA, Sierps MJ, Vergeer P, Berger CEH (2020) Combining evidence in complex cases – a practical approach to interdisciplinary casework. *Science & Justice* 60(1):20-29.

Dembinski GM, Sobieralski C, Picard CJ (2018) Estimation of the number of contributors of theoretical mixture profiles based on allele counting: Does increasing the number of loci increase success rate of estimates? *Forensic Science International: Genetics* 33:24-32.

Digréus P, Andersson AC, Nordgaard A, Ansell R (2011) Contamination monitoring in the forensic DNA laboratory and a simple model for unbiased EPG classification. *Forensic Science International: Genetics Supplement Series* 3:e299-e300.

Djuric M, Varljen T, Stanojevic A, Stojkovic O (2008) DNA typing from handled items. *Forensic Science International: Genetics Supplement Series* 1:411-412.

dMIQE Group (2020) The digital MIQE guidelines update: Minimum information for publication of quantitative digital PCR experiments for 2020. *Clinical Chemistry* 66(8):1012-1029.

[DAB 2000] DNA Advisory Board (2000) Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). *Forensic Sci. Comm.* 2(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm (accessed June 20, 2020).

DNA Identification Act of 1994 (1994) Public Law No. 103-322. Available at https://www.gpo.gov/fdsys/pkg/STATUTE-108/pdf/STATUTE-108-Pg1796.pdf (accessed June 25, 2020).

Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ (2001) NIST Mixed Stain Studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *Journal of Forensic Sciences* 46:1199-1210.

Duewer DL, Kline MC, Redman JW, Butler JM (2004) NIST Mixed Stain Study #3: signal intensity balance in commercial short tandem repeat multiplexes. *Analytical Chemistry* 76:6928-6934.

Duke KR, Myers SP (2020) Systematic evaluation of STRmix performance on degraded DNA profile data. *Forensic Science International: Genetics* 44:102174.

Durdle A, van Oorschot RAH, Mitchell RJ (2009) The transfer of human DNA by *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae). *Forensic Science International: Genetics Supplement Series* 2(1):180–182.

Durdle A, Mitchell RJ, van Oorschot RAH (2011) The change in human DNA content over time in the artefacts of the blowfly *Lucilia cuprina* (Meigen) (Diptera: Calliphoridae). *Forensic Science International: Genetics Supplement Series* 3(1):e289–e290.

Dziak R, Peneder A, Buetter A, Hageman C (2018) Trace DNA sampling success from evidence items commonly encountered in forensic casework. *Journal of Forensic Sciences* 63(3):835–841.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* 49:746-756.

Eldridge H (2019) Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy* 1:24-34.

[ENFSI 2015] Willis S, McKenna L, McDermott S, O'Donell G, Barrett A, Rasmusson B, Nordgaard A, Berger CEH, Sjerps MJ, Lucena-Molina JJ, Zadora G, Aitken C, Lovelock T, Lunt L, Champod C, Biedermann A, Hicks TN, Taroni F (2015) *ENFSI Guideline for Evaluative Reporting in Forensic Science*. Available at http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.

[ENFSI 2017] ENFSI DNA Working Group (2017) Best practice manual for the internal validation of probabilistic software to undertake DNA mixture interpretation. Available at http://enfsi.eu/wp-content/uploads/2017/09/Best-Practice-Manual-for-the-internal-validation-of-probabilistic-software-to-undertake-DNA-mixture-interpretation-v1.docx.pdf (accessed August 12, 2020).

Ensenberger MG, Thompson J, Hill B, Homick K, Kearney V, Mayntz-Press KA, Mazur P, McGuckian A, Myers J, Raley K, Raley SG, Rothove R, Wilson J, Wieczorek D, Fulmer PM, Storts DR, Krenke BE (2010) Developmental validation of the PowerPlex 16 HS System: an improved 16-locus fluorescent STR multiplex. *Forensic Science International: Genetics* 4(4):257-264.

Ensenberger MG, Hill CR, McLaren RS, Sprecher CJ, Storts DR (2014) Developmental validation of the PowerPlex(®) 21 System. *Forensic Science International: Genetics* 9:169-178.

Ensenberger MG, Lenz KA, Matthies LK, Hadinoto GM, Schienman JE, Przech AJ, Morganti MW, Renstrom DT, Baker VM, Gawrys KM, Hoogendoorn M, Steffen CR, Martín P, Alonso A, Olson HR, Sprecher CJ, Storts DR (2016) Developmental validation of the PowerPlex(®) Fusion 6C System. *Forensic Science International: Genetics* 21:134-144.

Evett IW, Buffery C, Willott G, Stoney D (1991) A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *Journal of Forensic Science Society* 31(1):41-47.

Evett IW (1995) Avoiding the transposed conditional. *Science & Justice* 35(2):127-131.

Evett IW, Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sinauer Associates: Sunderland, Massachusetts.

Evett IW, Gill PD, Lambert JA (1998) Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences* 43(1):62-69.

Evett IW, Jackson G, Lambert JA (2000a) More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. *Science & Justice* 40(1):3-10.

Evett IW, Jackson G, Lambert JA, McCrossan S (2000b) The impact of the principles of evidence interpretation on the structure and content of statements. *Science & Justice* 40(4):233-239.

Evett IW, Gill PD, Jackson G, Whitaker J, Champod C (2002) Interpreting small quantities of DNA: The hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences* 47(3):520-530.

Farash K, Hanson EK, Ballantyne J (2015) Enhanced genetic analysis of single human bioparticles recovered by simplified micromanipulation from forensic 'touch DNA' evidence. *Journal of Visualized Experiments* 97:e52612. Video available at https://www.jove.com/video/52612/enhanced-genetic-analysis-single-human-bioparticles-recovered.

Farash K, Hanson EK, Ballantyne J (2018) Single source DNA profile recovery from single cells isolated from skin and fabric from touch DNA mixtures in mock physical assaults. *Science & Justice* 58(3):191-199.

Farmen RK, Jaghø R, Cortez P, Frøyland ES (2008) Assessment of individual shedder status and implication for secondary DNA transfer. *Forensic Science International: Genetics Supplement Series* 1:415-417.

7917 Ferreira STG, Paula KA, Moraes AV (2013) DNA typing of trace DNA recovered from different areas of
7918 sandals found at a homicide crime scene investigation: A comparative study. *Forensic Science International:*
7919 *Genetics Supplement Series* 4(1):e372-e373.
7920
7921 Fildes N, Reynolds R (1995) Consistency and reproducibility of AmpliType PM results between seven
7922 laboratories: field trial results. *Journal of Forensic Sciences* 40(2):279-286.
7923
7924 Findlay I, Taylor A, Quirke P, Frazier R, Urquhart A (1997) DNA fingerprinting from single cells. *Nature*
7925 389:555-556.
7926
7927 Flanagan N, McAlister C (2011) The transfer and persistence of DNA under the fingernails following digital
7928 penetration of the vagina. *Forensic Science International: Genetics* 5(5):479-483.
7929
7930 Fonneløp AE, Egeland T, Gill P (2015a) Secondary and subsequent DNA transfer during criminal
7931 investigation. *Forensic Science International: Genetics* 17:155-162.
7932
7933 Fonneløp AE, Johannessen H, Gill P (2015b) Persistence and secondary transfer of DNA from previous users
7934 of equipment. *Forensic Science International: Genetics Supplement Series* 5:e191-e192.
7935
7936 Fonneløp AE, Johannessen H, Egeland T, Gill P (2016) Contamination during criminal investigation:
7937 Detecting police contamination and secondary DNA transfer from evidence bags. *Forensic Science*
7938 *International: Genetics* 23:121-129.
7939
7940 Fonneløp AE, Ramse M, Egeland T, Gill P (2017) The implications of shedder status and background DNA on
7941 direct and secondary transfer in an attack scenario. *Forensic Science International: Genetics* 29:48-60.
7942
7943 Fonneløp AE, Johannessen, H, Heen, G, Molland, K, Gill, P. (2019) A retrospective study on the transfer,
7944 persistence and recovery of sperm and epithelial cells in samples collected in sexual assault casework.
7945 *Forensic Science International: Genetics* 43:102153.
7946
7947 Fontana F, Rapone C, Bregola G, Aversa R, de Meo A, Signorini G, Sergio M, Ferrarini A, Lanzellotto R,
7948 Medoro G, Giorgini G, Manaresi N, Berti A (2017) Isolation and genetic analysis of pure cells from forensic
7949 biological mixtures: The precision of a digital approach. *Forensic Science International: Genetics* 29:225-241.
7950
7951 Fordyce SL, Mogensen HS, Børsting C, Lagacé RE, Chang CW, Rajagopalan N, Morling N (2015) Second-
7952 generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™.
7953 *Forensic Science International: Genetics* 14:132-140.
7954
7955 Forr C, Schei B, Stene LE, Ormstad K, Hagemann CT (2018) Factors associated with trace evidence analyses
7956 and DNA findings among police reported cases of rape. *Forensic Science International* 283:136-143.
7957
7958 [FTCOE 2015] Forensic Technology Center of Excellence (2015) *Landscape Study of DNA Mixture*
7959 *Interpretation Software*. Available at https://forensiccoe.org/report-dna-mixture-interpretation-software/
7960 (accessed June 4, 2020).
7961
7962 Frégeau CJ, Fourney RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and
7963 accurate approach to human identification. *BioTechniques* 15(1):100-119.
7964
7965 Geng T, Novak R, Mathies RA (2014) Single-cell forensic short tandem repeat typing within microfluidic
7966 droplets. *Analytical Chemistry* 86:703-712.
7967
7968 Geng T, Mathies RA (2015) Minimizing inhibition of PCR-STR typing using digital agarose droplet
7969 microfluidics. *Forensic Science International: Genetics* 14:203-209.
7970
7971 Gettings KB, Borsuk LA, Steffen CR, Kiesler KM, Vallone PM (2018) Sequence-based U.S. population data
7972 for 27 autosomal STR loci. *Forensic Science International: Genetics* 37:106-115.

Gettings KB, Ballard D, Bodner M, Borsuk LA, King JL, Parson W, Phillips C (2019) Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting. *Forensic Science International: Genetics* 43:102165.

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints.' *Nature* 318:577-579.

Gill P, Kimpton CP, Urquhart A, Oldroyd N, Millican ES, Watson SK, Downes TJ (1995) Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future. *Electrophoresis* 16:1543-1552.

Gill P, Sparkes R, Kimpton C (1997) Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International* 89(3):185-197.

Gill P, Sparkes R, Pinchin R, Clayton T, Whitaker J, Buckleton J (1998) Interpreting simple STR mixtures using allele peak areas. *Forensic Science International* 91(1):41-53.

Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Science International* 112(1):17-40.

Gill P (2001) Application of low copy number DNA profiling. *Croatian Medical Journal* 42(3):229-232.

Gill P (2002) Role of short tandem repeat DNA in forensic casework in the UK—Past, present, and future perspectives. *Biotechniques* 32(2):366-372.

Gill P, Curran J, Elliot K (2005) A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Research* 33(2):632-643.

Gill P, Fereday L, Morling N, Schneider PM (2006a) The evolution of DNA databases—recommendations for new European STR loci. *Forensic Science International* 156(2-3):242-244.

Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006b) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160:90-101.

Gill P, Kirkham A, Curran J (2007) *LoComatioN*: a software tool for the analysis of low copy number DNA profiles. *Forensic Science International* 166:128-138.

Gill P, Brown RM, Fairley M, Lee L, Smyth M, Simpson N, Irwin B, Dunlop J, Greenhalgh M, Way K, Westacott EJ, Ferguson SJ, Ford LV, Clayton T, Guiness J (2008) National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *Forensic Science International: Genetics* 2(1):76-82.

Gill P, Buckleton J (2010a) Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J Forensic Sci* 2009;54(4):810-21. *Journal of Forensic Sciences* 55(1):265-268. [authors response: pp. 269-272 from Budowle B, Chakraborty R, van Daal A]

Gill P, Buckleton J (2010b) A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Science International: Genetics* 4(4):221-227.

Gill P, Gusmao L, Haned H, Mayr WR, Morling N, Parson W, Prieto L, Prinz M, Schneider H, Schneider PM, Weir BS (2012) DNA Commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Science International: Genetics* 6(6):679-688.

8029 Gill P, Haned H (2013) A new methodological framework to interpret complex DNA profiles using likelihood
8030 ratios. *Forensic Science International: Genetics* 7(2):251-263.
8031
8032 Gill P (2014) *Misleading DNA Evidence: Reasons for Miscarriages of Justice.* Elsevier Academic Press: San
8033 Diego.
8034
8035 Gill P, Haned H, Bleka Ø, Hansson O, Dørum G, Egeland T (2015) Genotyping and interpretation of STR-DNA:
8036 Low-template, mixtures and database matches—Twenty years of research and development. *Forensic Science*
8037 *International: Genetics* 18:100-117.
8038
8039 Gill P (2016) Analysis and implications of the miscarriages of justice of Amanda Knox and Raffaele Sollecito.
8040 *Forensic Science International: Genetics* 23:9-18.
8041
8042 Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, van Oorschot RAH, Parson W,
8043 Prinz M, Schneider PM, Sijen T, Taylor D (2018) DNA Commission of the International Society for Forensic
8044 Genetics: Assessing the value of forensic biological evidence—Guidelines highlighting the importance of
8045 propositions. Part I: Evaluations of DNA profiling comparisons given (sub-) source propositions. *Forensic*
8046 *Science International: Genetics* 36:189-202.
8047
8048 Gill P (2019a) DNA evidence and miscarriages of justice. *Forensic Science International* 294:e1-e3.
8049
8050 Gill, P. (2019b) Interpretation continues to be the main weakness in criminal justice systems: developing roles
8051 of the expert witness and court. *WIREs Forensic Sci.* 1:e1321. Available at https://doi.org/10.1002/wfs2.1321.
8052
8053 Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, van Oorschot RAH, Parson W,
8054 Prinz M, Schneider PM, Sijen T, Taylor D (2020a) DNA Commission of the International Society for Forensic
8055 Genetics: Assessing the value of forensic biological evidence - guidelines highlighting the importance of
8056 propositions. Part II: Evaluation of biological traces considering activity level propositions. *Forensic Science*
8057 *International: Genetics* 44:102186.
8058
8059 Gill P, Bleka Ø, Hansson O, Benschop C, Haned H (2020b) *Forensic Practitioner's Guide to the*
8060 *Interpretation of Complex DNA Profiles.* Elsevier Academic Press: San Diego.
8061
8062 Gittelson S, Kalafut T, Myers S, Taylor D, Hicks T, Taroni F, Evett IW, Bright JA, Buckleton J (2016) A
8063 practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in
8064 an adversarial environment. *Journal of Forensic Sciences* 61(1):186-195.
8065
8066 Gittelson S, Berger CEH, Jackson G, Evett IW, Champod C, Robertson B, Curran JM, Taylor D, Weir BS,
8067 Coble MD, Buckleton JS (2018) A response to "Likelihood ratio as weight of evidence: A closer look" by
8068 Lund and Iyer. *Forensic Science International* 288:e15-e19.
8069
8070 Goodwin W, Linacre A, Hadi S (2010) *An Introduction to Forensic Genetics, Second Edition*. Wiley:
8071 Hoboken, NJ.
8072
8073 Goray M, Eken E, Mitchell RJ, van Oorschot RAH (2010a) Secondary DNA transfer of biological substances
8074 under varying test conditions. *Forensic Science International: Genetics* 4:62-67.
8075
8076 Goray M, Mitchell RJ, van Oorschot RAH (2010b) Investigation of secondary DNA transfer of skin cells
8077 under controlled test conditions. *Legal Medicine* 12:117-120.
8078
8079 Goray M, Mitchell JR, van Oorschot RAH (2012a) Evaluation of multiple transfer of DNA using mock case
8080 scenarios. *Legal Medicine* 14:40-46.
8081
8082 Goray M, van Oorschot RAH, Mitchell JR (2012b) DNA transfer within forensic exhibit packaging: Potential
8083 for DNA loss and relocation. *Forensic Science International: Genetics* 6(2):158-166.
8084

Goray M, van Oorschot RAH (2015) The complexities of DNA transfer during a social setting. *Legal Medicine* 17:82-91.

Goray M, Fowler S, Szkuta B, van Oorschot RAH (2016) Shedder status—An analysis of self- and non-self-DNA in multiple handprints deposited by the same individuals over time. *Forensic Science International: Genetics* 23:190-196.

Goray M, Pirie E, van Oorschot RAH (2019) DNA transfer: DNA acquired by gloves during casework examinations. *Forensic Science International: Genetics* 38:167-174.

Gosch A, Courts C (2019) On DNA transfer: The lack and difficulty of systematic research and how to do it better. *Forensic Science International: Genetics* 40:24-36.

Graham EA, Rutty GN (2008) Investigation into "normal" background DNA on adult necks: Implications for DNA profiling of manual strangulation victims. *Journal of Forensic Sciences* 53(5):1074-1082.

Gramlich J (2019) Only 2% of federal criminal defendants go to trial, and most who do are found guilty. Pew Research Center. Available at https://www.pewresearch.org/fact-tank/2019/06/11/only-2-of-federal-criminal-defendants-go-to-trial-and-most-who-do-are-found-guilty/ (accessed on August 12, 2020).

Grasso JA, Woodard JW (1967) DNA synthesis and mitosis in erythropoietic cells. *Journal of Cell Biology* 33(3):645-655.

Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics.* Wiley: New York.

Green RL, Lagacé RE, Oldroyd NJ, Hennessy LK, Mulero JJ (2013) Developmental validation of the AmpFℓSTR® NGM SElect™ PCR Amplification Kit: A next-generation STR multiplex with the SE33 locus. *Forensic Science International: Genetics* 7(1):41-51.

Greenspoon SA, Schiermeier-Wood L, Jenkins BC (2015) Establishing the limits of TrueAllele® Casework: a validation study. *Journal of Forensic Sciences* 60(5):1263-1276.

Greytak EM, Moore C, Armentrout SL (2019) Genetic genealogy for cold case and active investigations. *Forensic Science International* 299:103-113.

Gršković B, Markulin D, Zidkova A, Cmjac J, Anđelinović Š, Marijanović I, Tomašević L, Popović M, Primorac D, Mršić G (2014) Impact of donor age, gender and handling time on the DNA concentration left on different surfaces. *International Journal of Biomedicine* 4(3):169–174. Available at http://www.ijbm.org/articles/Article4_3_FR.pdf.

Haas C, Hanson E, Banemann R, Bento AM, Berti A, Carracedo A, Courts C, Cock G, Drobnic K, Fleming R, Franchi C, Gomes I, Hadzic G, Harbison SA, Hjort B, Hollard C, Hoff-Olsen P, Keyser C, Kondili A, Maronas O, McCallum N, Miniati P, Morling N, Niederstatter H, Noel F, Parson W, Porto MJ, Roeder AD, Sauer E, Schneider PM, Shanthan G, Sijen T, Syndercombe Court D, Turanska M, van den Berge M, Vennemann M, Vidaki A, Zatkalikova L, Ballantyne J (2015) RNA/DNA co-analysis from human skin and contact traces—Results of a sixth collaborative EDNAP exercise. *Forensic Science International: Genetics* 16:139–147.

Haned H, Slooten K, Gill P (2012) Exploratory data analysis for the interpretation of low template DNA mixtures. *Forensic Science International: Genetics* 6(6):762-774.

Haned H, Benschop CC, Gill PD, Sijen T (2015) Complex DNA mixture analysis in a forensic context: evaluating the probative value using a likelihood ratio model. *Forensic Science International: Genetics* 16:17-25.

Haned H, Gill P, Lohmueller K, Inman K, Rudin N (2016) Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. *Science & Justice* 56(2):104-108.

Hannig J, Riman S, Iyer H, Vallone PM (2019) Are reported likelihood ratios well calibrated? *Forensic Science International: Genetics Supplement Series* 7:572-574.

Hanson E, Haas C, Jucker R, Ballantyne J (2011) Identification of skin in touch/contact forensic samples by messenger RNA profiling. *Forensic Science International: Genetics Supplement Series* 3:e305-306.

Hanson E, Haas C, Jucker R, Ballantyne J (2012) Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. *Forensic Science International: Genetics* 6(5):548-558.

Hanson EK, Ballantyne J (2013) "Getting blood from a stone" Ultrasensitive forensic DNA profiling of microscopic bio-particles recovered from "touch DNA." *Methods in Molecular Biology* 1039:3-17.

Hares DR (2012) Expanding the CODIS core loci in the United States. *Forensic Science International: Genetics* 6(1):e52-e54.

Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics* 17:33-34.

Hardy GH (1908) Mendelian proportions in a mixed population. *Science* 17:49-50.

Hazell-Smithen M, Callahan T, Miller Coyle H (2014) Touch DNA and the ability to detect the correct source. *International Journal of Advanced Science and Technology* 1(1):45-54. Available at http://www.identacode.org/Touch-DNA-and-the-Ability-to-Detect-the-Correct-Source.pdf (accessed August 11, 2020).

Hellerud B, Johannessen H, Haltbakk H, Hoff-Olsen P (2008) Zip lock poly bags in drug cases—A valuable source for obtaining identifiable DNA results? *Forensic Science International: Genetics Supplement Series* 1(1):433-434.

Helmus J, Bajanowski T, Poetsch M (2016) DNA transfer—A never ending story. A study on scenarios involving a second person as carrier. *International Journal of Legal Medicine* 130(1):121-125.

Helmus J, Zorell S, Bajanowski T, Poetsch M (2018) Persistence of DNA on clothes after exposure to water for different time periods—A study on bathtub, pond, and river. *International Journal of Legal Medicine* 132(1):99-106.

Helmus J, Pfeifer M, Feiner LK, Krause LJ, Bajanowski T, Poetsch M (2019) Unintentional effects of cleaning a crime scene—When the sponge becomes an accomplice in DNA transfer. *International Journal of Legal Medicine* 133(3):759-765.

Henry J, McGowan P, Brown C (2015) A survey of environmental DNA in South Australia Police facilities. *Forensic Science International: Genetics Supplement Series* 5:e465-e466.

Hert DG, Fredlake CP, Barron AE (2008) Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 29(23): 4618-4626.

Hicks T, Biedermann A, de Koeijer JA, Taroni F, Champod C, Evett IW (2015) The importance of distinguishing information from evidence/observations when formulating propositions. *Science & Justice* 55(6):520-525.

Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS (2002) TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. *Journal of Forensic Sciences* 47(1):66-96.

Hundl C, Neuman M, Rairden A, Rearden P, Stout P (2020) Implementation of a blind quality control program in a forensic laboratory. *Journal of Forensic Sciences* 65(3):815-822.

Inman K, Rudin N (2001) *Principles and Practice of Criminalistics: The Profession of Forensic Science*. CRC Press: Boca Raton.

Inman K, Rudin N, Cheng K, Robinson C, Kirschner A, Inman-Semerau L, Lohmueller KE (2015) Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles. *BMC Bioinformatics* 16:298.

International Business Communications (IBC) (1998). Resolution and Interpretation of Mixtures. July 31, 1998. Post-conference workshop taught by Peter Gill, James Curran, and Charles Brenner. Copy of workshop materials held by attendee John Butler.

ISHI 2010 workshop: https://strbase.nist.gov/training/ISHI2010_MixtureWorkshop.htm (accessed August 12, 2020).

ISHI 2011 workshop: https://strbase.nist.gov/training/MixtureWorkshop-ISHI2011-no-lit.pdf (accessed August 12, 2020).

ISHI 2012 workshop: https://strbase.nist.gov/training/ISHI2012-MixtureWorkshop-Outline.pdf (accessed August 12, 2020). **M**ixtures **U**sing *SOUND* **S**tatistics, **I**nterpretation & **C**onclusions (ISHI 2012 Workshop: Butler, Coble, Cotton, Grgicak, Word) [Introduction] [Overview] [Profile1] [Statistics] [Profile Enhancement] [Different Assumptions] [BU Website] [Complex Mixtures] [Probabilistic Genotyping] [Complexity Threshold] [Court] [Future]

International Organization for Standardization (2016). ISO 18385:2016. Minimizing the risk of human DNA contamination in products used to collect, store and analyze biological material for forensic purposes – Requirements. Available at https://www.iso.org/standard/62341.html (accessed August 12, 2020).

ISO/IEC 17025:2017 (2017) General requirements for the competence of testing and calibration laboratories. Available at https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html (accessed August 12, 2020). *Section 7.2.2.1(e) mentions interlaboratory comparisons as a technique used for method validation and Section 7.7.2(b) encourages laboratories to monitor their performance by participation in interlaboratory comparisons.*

Iuvaro A, Bini C, Dilloo S, Salmo S, Pelotti S (2018) Male DNA under female fingernails after scratching: Transfer and persistence evaluation by RT-PCR analysis and Y-STR typing. *International Journal of Legal Medicine* 132(6):1603-1609.

Jackson G, Jones S, Booth G, Champod C, Evett IW (2006) The nature of forensic science opinion – a possible framework to guide thinking and practice in investigation and in court proceedings. *Science & Justice* 46(1):33-44.

Jackson G (2013) The impact of commercialization on the evaluation of DNA evidence. *Frontiers in Genetics* 4:227. https://doi.org/10.3389/fgene.2013.00227.

Jackson G, Biedermann A (2019) "Source" or "activity": What is the level of issue in a criminal trial? *Significance* 16(2):36-39.

Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, Bruand J, Schlesinger F, Pond SJK, Varlaro J, Stephens KM, Holt CL (2017) Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Science International: Genetics* 28:52-70.

8251 Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67-
8252 73.
8253
8254 Jeffreys AJ (1987) Highly variable minisatellites and DNA fingerprints. *Biochemical Society Transactions*
8255 15(3):309-317.
8256
8257 Jones S, Scott K, Lewis J, Davidson G, Allard JE, Lowrie C, McBride BM, McKenna L, Teppett G, Rogers C,
8258 Clayson N, Baird A (2016) DNA transfer through nonintimate social contact. *Science & Justice* 56(2):90-95.
8259
8260 Just RS, Irwin JA (2018) Use of the LUS in sequence allele designations to facilitate probabilistic genotyping
8261 of NGS-based STR typing results. *Forensic Science International: Genetics* 34:197-205.
8262
8263 Kadane JB (2011) *Principles of Uncertainty*. CRC Press: Boca Raton.
8264
8265 Kafarowski E, Lyon AM, Sloan MM (1996) The retention and transfer of spermatozoa in clothing by machine
8266 washing. *Canadian Society of Forensic Science Journal* 29(1):7-11.
8267
8268 Kalafut T, Schuerman C, Sutton J, Faris T, Armogida L, Bright JA, Buckleton J, Taylor D (2018)
8269 Implementation and validation of an improved allele specific stutter filtering method for electropherogram
8270 interpretation. *Forensic Science International: Genetics* 35:50-56.
8271
8272 Kamphausen T, Schadendorf D, von Wurmb-Schwark N, Bajanowski T, Poetsch M (2012) Good shedder or
8273 bad shedder—The influence of skin diseases on forensic DNA analysis from epithelial abrasions. *International*
8274 *Journal of Legal Medicine* 126:179-183.
8275
8276 Kamphausen T, Fandel SB, Gutmann JS, Bajanowski T, Poetsch M (2015) Everything clean? Transfer of
8277 DNA traces between textiles in the washtub. *International Journal of Legal Medicine* 129:709-714.
8278
8279 Kanokwongnuwut P, Martin B, Kirkbride KP, Linacre A (2018) Shedding light on shedders. *Forensic Science*
8280 *International: Genetics* 36:20-25.
8281
8282 Katsanis SH, Wagner JK (2013) Characterization of the standard and recommended CODIS markers. *Journal*
8283 *of Forensic Sciences* 58(Suppl 1):S169-S172.
8284
8285 Kaye DH (2010) *The Double Helix and the Law of Evidence*. Harvard University Press: Cambridge, MA. *See*
8286 *pp. 200-209, where Bruce Weir was requested to produce statistical analyses of multiple mixtures from the*
8287 *O.J. Simpson case, but Judge Ito would not permit LRs to be used (Bruce Weir's testimony took place on June*
8288 *22-23, 1995).*
8289
8290 Kayser M (2007) Uni-parental markers in human identity testing including forensic DNA analysis.
8291 *Biotechniques* 43(6):S16-S21.
8292
8293 Kelly H, Bright JA, Buckleton JS, Curran JM (2014) A comparison of statistical models for the analysis of
8294 complex forensic DNA profiles. *Science & Justice* 54(1):66-70.
8295
8296 Kelly H, Bright JA, Kruijver M, Cooper S, Taylor D, Duke K, Strong M, Beamer V, Buettner C, Buckleton J
8297 (2018) A sensitivity analysis to determine the robustness of STRmix™ with respect to laboratory calibration.
8298 *Forensic Science International: Genetics* 35:113-122.
8299
8300 Kelly H, Bright JA, Coble MD, Buckleton J (2020) A description of the likelihood ratios in the probabilistic
8301 genotyping software STRmix™. *WIREs Forensic Sciences* 2(6):e1377.
8302
8303 Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, Haigh E, Kidd JK (2014) Current
8304 sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic*
8305 *Science International: Genetics* 12:215-224.
8306

Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods & Applications* 3:13-22.

Kirby LT (1990) *DNA Fingerprinting: An Introduction.* Stockton Press: New York.

Kita T, Yamaguchi H, Yokoyama M, Tanaka T, Tanaka N (2008) Morphological study of fragmented DNA on touched objects. *Forensic Science International: Genetics* 3:32-36.

Kline MC, Redman JW, Duewer DL, Reeder DJ (1999) Results from the 1999 NIST Mixed-Stain Study #2: DNA quantitation, differential extraction, and identification of the unknown contributors. *Proceedings of the International Symposium on Human Identification* (Promega Corporation). Available at https://promega.media/-/media/files/resources/conference-proceedings/ishi-10/oral-presentations/42kline.pdf?la=en. (accessed 13 November 2020).

Kline MC, Duewer DL, Redman JW, Butler JM (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Analytical Chemistry* 75:2463-2469.

Kloosterman A, Sjerps M, Quak A (2014) Error rates in forensic DNA analysis: Definition, numbers, impact and communication. *Forensic Science International: Genetics* 12:77-85.

Kokshoorn B, Blankers BJ, de Zoete J, Berger CEH (2017) Activity level DNA evidence evaluation: On propositions addressing the actor or the activity. *Forensic Science International* 278:115-124.

Kokshoorn B, Aarts LHJ, Ansell R, Connolly E, Drotz W, Kloosterman AD, McKenna LG, Szkuta B, van Oorschot RAH (2018) Sharing data on DNA transfer, persistence, prevalence and recovery: Arguments for harmonization and standardization. *Forensic Science International: Genetics* 37:260-269.

Kolowski J, Barden C, Brown C, Leete J, Czyryca C (2016) Proficiency testing trends following the 2009 National Academy of Sciences report, "Strengthening Forensic Science in the United States: A Path Forward". *Forensic Science Policy and Management* 7: 98-105.

Kovács G, Pádár Z (2015) Misinterpretation of sample contamination in a Hungarian casework. *Forensic Science International: Genetics Supplement Series* 5:e425-e427.

Kraemer M, Prochnow A, Bussmann M, Scherer M, Peist R, Steffen C (2017) Developmental validation of QIAGEN Investigator® 24plex QS Kit and Investigator® 24plex GO! Kit: Two 6-dye multiplex assays for the extended CODIS core loci. *Forensic Science International: Genetics* 29:9-20.

Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ (2002) Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Sciences* 47(4):773-785.

Kwong K (2017) The algorithm says you did it: The use of black box algorithms to analyze complex DNA evidence. *Harvard Journal of Law & Technology* 31(1):275-301.

Lacerenza D, Aneli S, Omedei M, Gino S, Pasino S, Berchialla P, Robino C (2016) A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers. *Forensic Science International: Genetics* 22:44-53.

Lander ES, et al. (International Human Genome Sequencing Consortium) (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921. doi: 10.1038/35057062. Erratum in: Nature 2001 Aug 2;412(6846):565. Erratum in: Nature 2001 Jun 7;411(6838):720.

Lazaruk K, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, Menchen S, Scheibler D, Wenz HM, Holt C, Wallin J (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19(1):86-93.

Leclair B, Sgueglia JB, Wojtowicz PC, Juston AC, Frégeau CJ, Fourney RM (2003) STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *Journal of Forensic Sciences* 48(5):1001-1013.

Leclair B, Frégeau CJ, Bowen KL, Fourney RM (2004) Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. *Journal of Forensic Sciences* 49(5):968-980.

LehmannVJ, Mitchell RJ, Ballantyne KN, van Oorschot RAH (2013) Following the transfer of DNA: How far can it go? *Forensic Science International: Genetics Supplement Series* 4:e53-e54.

Lehmann VJ, Mitchell RJ, Ballantyne KN, van Oorschot RAH (2015) Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA? *Forensic Science International: Genetics* 19:68-75.

Li CC, Weeks DE, Chakravarti A. (1993) Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity* 43:45-52.

Lin M-H, Bright J-A, Pugh SN, Buckleton JS (2020) The interpretation of mixed DNA profiles from a mother, father, and child trio. *Forensic Science International: Genetics* 44:102175.

Lindley DV (1977) A problem in forensic science. *Biometrika* 64:207-213.

Lindley DV (2014) *Understanding Uncertainty*, Revised Edition. John Wiley and Sons: Hoboken.

Lohmueller KE, Rudin N (2013) Calculating the weight of evidence in low-template forensic DNA casework. *Journal of Forensic Sciences* 58 Suppl 1:S243-S249.

Lowe A, Murray C, Whitaker J, Tully G, Gill P (2002) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International* 129:25–34.

Ludeman MJ, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, Wang DY (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine* 132(6):1555-1573.

Lund SP, Iyer H (2017) Likelihood ratio as weight of forensic evidence: a closer look. *Journal of Research of National Institute of Standards and Technology* 122:27. Available at https://nvlpubs.nist.gov/nistpubs/jres/122/jres.122.027.pdf (accessed March 8, 2021).

Lynch PC, Cotton RW (2018) Determination of the possible number of genotypes which can contribute to DNA mixtures: non-computer assisted deconvolution should not be attempted for greater than two person mixtures. *Forensic Science International: Genetics* 37:235-240.

Magee AM, Breathnach M, Doak S, Thornton F, Noone C, McKenna LG (2018) Wearer and non-wearer DNA on the collars and cuffs of upper garments of worn clothing. *Forensic Science International: Genetics* 34:152-161.

Manabe S, Morimoto C, Hamano Y, Fujimoto S, Tamaki K (2017) Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model. *PLoS ONE* 12(11):e0188183.

Manoli P, Antoniou A, Bashiardes E, Xenophontos S, Photiades M, Stribley V, Mylona M, Demetriou C, Cariolou MA (2016) Sex-specific age association with primary DNA transfer. *International Journal of Legal Medicine* 130(1):103-112.

Mapes AA, Kloosterman AD, van Marion V, de Poot CJ (2016) Knowledge on DNA success rates to optimize the DNA analysis process: From crime scene to laboratory. *Journal of Forensic Sciences* 61(4):1055-1061.

Margiotta G, Tasselli G, Tommolini F, Lancia M, Massetti S, Carnevali E (2015) Risk of DNA transfer by gloves in forensic casework. *Forensic Science International: Genetics Supplement Series* 5:e527-e529.

Mari L, Carbone P (2012) Measurement fundamentals: a pragmatic view. *IEEE Transactions on Instrumentation and Measurement* 61(8):2107-2115.

Marquis R, Biedermann A, Cadola L, Champod C, Gueissaz L, Massonnet G, Mazzella WD, Taroni F, Hicks T (2016) Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings. *Science & Justice* 56:364-370.

Martire KA, Kemp RI (2018) Considerations when designing human performance tests in the forensic sciences. *Australian Journal of Forensic Sciences* 50(2):166-182.

Matte M, Williams L, Frappier R, Newman J (2012) Prevalence and persistence of foreign DNA beneath fingernails. *Forensic Science International: Genetics* 6(2):236-243.

McDonald A, Jones E, Lewis J, O'Rourke P (2015) Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. *Forensic Science International: Genetics* 15:84-89.

McGovern C, Cheng K, Kelly H, Ciecko A, Taylor D, Buckleton JS, Bright JA (2020) Performance of a method for weighting a range in the number of contributors in probabilistic genotyping. *Forensic Science International: Genetics* 48:102352.

McKenna L (2013) Understanding DNA results within the case context: Importance of the alternative proposition. *Frontiers in Genetics* 4:242.

Meakin G, Jamieson A (2013) DNA transfer: Review and implications for casework. *Forensic Science International: Genetics* 7:434-443.

Meakin G., Butcher EV, van Oorschot RAH, Morgan RM (2015) The deposition and persistence of indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics Supplement Series* 5:e498-e500.

Meakin GE, Butcher EV, van Oorschot RAH, Morgan RM (2017) Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly- and indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics* 29:38-47.

Meuwly D, Ramos D, Haraksim R (2017) A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Science International* 276:142-153.

Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Prinz M, Caragine T (2011) Likelihood ratio statistics for DNA mixtures allowing for drop-out and drop-in. *Forensic Science International: Genetics Supplement Series* 3:e240-e241.

Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Budimlija Z, Prinz M, Caragine T (2012) Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *Forensic Science International: Genetics* 6(6):749-761.

Montpetit S, O'Donnell P (2015) An optimized procedure for obtaining DNA from fired and unfired ammunition. *Forensic Science International: Genetics* 17:70-74.

Moore D, Clayton T, Thomson J (2020) A comprehensive study of allele drop-in over an extended period of time. *Forensic Science International: Genetics* 48:102332.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B (2001a) Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Sciences* 46(3):647-660.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B (2001b) Validation of STR typing by capillary electrophoresis. *Journal of Forensic Sciences* 46(3):661-676.

Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, Bright J-A, Onorato AJ (2016) Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. *Forensic Science International: Genetics* 25:175-181.

Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright JA, Taylor DA, Onorato AJ (2017) Internal validation of STRmix for the interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 29:126-144.

Morling N, Bastisch I, Gill P, Schneider PM (2007) Interpretation of DNA mixtures--European consensus on principles. *Forensic Science International: Genetics* 1(3-4):291-292.

Morrison GS, Enzinger E (2016) What should a forensic practitioner's likelihood ratio be? *Science & Justice* 56:374-379.

[NAS 2009] National Academy of Sciences (2009) *On Being a Scientist: A Guide to Responsible Conduct in Research* (Third Edition). The National Academies Press: Washington, DC.

National Research Council (1992) *DNA Technology in Forensic Science.* The National Academy Press: Washington DC. [NRC I]

National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. The National Academy Press: Washington DC. [NRC II]

[NRC 2009] National Research Council, Committee on Identifying the Needs of the Forensic Science Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press: Washington DC.

Neuhuber F, Dunkelmann B, Höckner G, Kiesslich J, Klausriegler E, Radacher M (2009) Female criminals: It's not always the offender! *Forensic Science International: Genetics Supplement Series* 2(1):145-146.

Neuhuber F, Kreindl G, Kastinger T, Dunkelmann B, Zahrer W, Cemper-Kiesslich J, Grießner I (2017) Police officer's DNA on crime scene samples—Indirect transfer as a source of contamination and its database-assisted detection in Austria. *Forensic Science International: Genetics Supplement Series* 6:e608-e609.

[NISTIR 8225] Butler, J.M., Iyer, H., Press, R., Taylor, M.K., Vallone, P.M. and Willis, S. (2020) NIST scientific foundation reviews. NISTIR 8225. Available at https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8225.pdf (accessed February 4, 2021).

Noël S, Lagacé K, Rogic A, Granger D, Bourgoin S, Jolicoeur C, Séguin D (2016) DNA transfer during laundering may yield complete genetic profiles. *Forensic Science International: Genetics* 23:240-247.

Noël S, Noël J, Granger D, Lefebvre JF, Séguin D (2019) STRmix™ put to the test: 300 000 non-contributor profiles compared to four-contributor DNA mixtures and the impact of replicates. *Forensic Science International: Genetics* 41:24-31.

Novroski NMM, Woerner AE, Budowle B (2018) Potential highly polymorphic short tandem repeat markers for enhanced forensic identity testing. *Forensic Science International: Genetics* 37:162-171.

Oldoni F, Castella V, Hall D (2015) Exploring the relative DNA contribution of first and second object's users on mock touch DNA mixtures. *Forensic Science International: Genetics Supplement Series* 5:e300-e301.

Oldoni F, Castella V, Hall D (2016) Shedding light on the relative DNA contribution of two persons handling the same object. *Forensic Science International: Genetics* 24:148-157.

Oleiwi AA, Morris MR, Schmerer WM, Sutton R (2015) The relative DNA-shedding propensity of the palm and finger surfaces. *Science & Justice* 55(5):329-334.

Oostdik K, Lenz K, Nye J, Schelling K, Yet D, Bruski S, Strong J, Buchanan C, Sutton J, Linner J, Frazier N, Young H, Matthies L, Sage A, Hahn J, Wells R, Williams N, Price M, Koehler J, Staples M, Swango KL, Hill C, Oyerly K, Duke W, Katzilierakis L, Ensenberger MG, Bourdeau JM, Sprecher CJ, Krenke B, Storts DR (2014) Developmental validation of the PowerPlex(®) Fusion System for analysis of casework and reference samples: A 24-locus multiplex for new database standards. *Forensic Science International: Genetics* 12:69-76.

Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. *Journal of Forensic Sciences* 50(6):1361-1366.

Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmao L, Hares DR, Irwin JA, King JL, de Knijff P, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C (2016) Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics* 22:54-63.

Perlin MW, Lancia G, Ng SK (1995) Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. *American Journal of Human Genetics* 57(5):1199-1210.

Perlin MW, Szabady B (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Journal of Forensic Sciences* 46(6):1372-1378.

Perlin MW (2006) Scientific validation of mixture interpretation methods. *Proceedings of the Seventeenth International Symposium on Human Identification*. Available at https://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/17th-ishi-oral-presentations/ (accessed June 23, 2020).

Perlin MW, Kadane JB, Cotton RW (2009) Match likelihood ratio for uncertain genotypes. *Law, Probability and Risk* 8(3):289-302.

Perlin MW, Sinelnikov A (2009) An information gap in DNA evidence interpretation. *PLoS ONE* 4(12):e8327.

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW (2011) Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences* 56(6):1430-1447.

Perlin MW, Belrose JL, Duceman BW (2013) New York State TrueAllele® casework validation study. *Journal of Forensic Sciences* 58(6):1458-1466.

Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S (2014) TrueAllele casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE* 9(3):e92837.

Perlin MW, Hornyak JM, Sugimoto G, Miller KW (2015) TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences* 60(4):857-868.

Perlin MW (2017) Method and system for DNA mixture analysis. United States Patent 9,708,642 B2 (July 18, 2017).

Pfeifer CM, Wiegand P (2017) Persistence of touch DNA on burglary-related tools. *International Journal of Legal Medicine* 131(4):941-953.

Phillips C, Gettings KB, King JL, Ballard D, Bodner M, Borsuk L, Parson W (2018) "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic Science International: Genetics* 34:162-169.

Phipps M, Petricevic S (2007) The tendency of individuals to transfer DNA to handled items. *Forensic Science International* 168(2–3):162-168.

Pickrahn I, Kreindl G, Muller E, Dunkelmann B, Zahrer W, Cemper-Kiesslich J, Neuhuber F (2017) Contamination incidents in the pre-analytical phase of forensic DNA analysis in Austria—Statistics of 17 years. *Forensic Science International: Genetics* 31:12-18.

Poetsch M, Bajanowski T, Kamphausen T (2013) Influence of an individual's age on the amount and interpretability of DNA left on touched items. *International Journal of Legal Medicine* 127:1093-1096.

Poetsch M, Pfeifer M, Konrad H, Bajanowski T, Helmus J (2018) Impact of several wearers on the persistence of DNA on clothes—A study with experimental scenarios. *International Journal of Legal Medicine* 132(1):117-123.

Possolo A (2015) *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results.* NIST Technical Note 1900. Available at https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.1900.pdf (accessed June 20, 2020).

Poulsen L, Morling N (2013) White Book on the current status of education and training in forensic genetics in Europe. Available at https://www.euroforgen.eu/fileadmin/websites/euroforgen/images/Training/White_book_final.pdf (accessed June 4, 2020).

[PCAST 2016] President's Council of Advisors on Science and Technology. Report to the President - Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. September 2016. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed June 24, 2020).

[PCAST 2017] President's Council of Advisors on Science and Technology. *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf (accessed July 13, 2020).

Press R (2020) Two New Forensic DNA Standards Added to the OSAC Registry: National forensic science organization approves standards for interpreting DNA mixtures. NIST press release. Available at https://www.nist.gov/news-events/news/2020/05/two-new-forensic-dna-standards-added-osac-registry (accessed June 4, 2020).

Preuße-Prange A, Renneberg R, Schwark T, Poetsch M, Simeoni E, von Wurmb-Schwark N (2009) The problem of DNA contamination in forensic case work: How to get rid of unwanted DNA? *Forensic Science International: Genetics Supplement Series* 2(1):185-186.

Prieto L, Haned H, Mosquera A, Crespillo M, Aleman M, Aler M, Alvarez F, Baeza-Richer C, Dominguez A, Doutremepuich C, Farfán MJ, Fenger-Grøn M, García-Ganivet JM, González-Moya E, Hombreiro L, Lareu MV, Martínez-Jarreta B, Merigioli S, Milans Del Bosch P, Morling N, Muñoz-Nieto M, Ortega-González E,

Pedrosa S, Pérez R, Solís C, Yurrebaso I, Gill P (2014) Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. *Forensic Science International: Genetics* 9:47-54.

Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, Balding D (2013) Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics* 7(5):555-563.

Puch-Solis R, Clayton T (2014) Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software. *Forensic Science International: Genetics* 11:220-228.

[QAS 1998] Federal Bureau of Investigation (2000). Quality assurance standards for forensic DNA testing laboratories. *Forensic Science Communications*, *2*(3). Available at https://www2.fbi.gov/hq/lab/fsc/backissu/july2000/codispre.htm or https://www2.fbi.gov/hq/lab/fsc/backissu/july2000/codis2a.htm (accessed June 29, 2020).

[QAS 1999] Federal Bureau of Investigation (2000). Quality assurance standards for convicted offender DNA databasing laboratories. *Forensic Science Communications*, *2*(3). Available at https://www2.fbi.gov/hq/lab/fsc/backissu/july2000/codis1a.htm (accessed June 29, 2020).

[QAS 2009] Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. (2009). Available at https://strbase.nist.gov/QAS/Final-FBI-Director-Forensic-Standards.pdf (accessed June 29, 2020).

[QAS 2011] Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. (2011). Available at https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view (accessed June 29, 2020).

[QAS 2020] Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. (2020). Available at https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view (accessed February 4, 2021).

Quinones I, Daniel B (2012) Cell free DNA as a component of forensic evidence recovered from touched surfaces. *Forensic Science International: Genetics* 6:26-30.

Ramos D, Gonzalez-Rodriguez J (2013) Reliable support: Measuring calibration of likelihood ratios. *Forensic Science International* 230:156-169.

Ramos D, Gonzalez-Rodriguez J, Zadora G, Aitken C (2013) Information-theoretical assessment of the performance of likelihood ratio computation methods. *Journal of Forensic Sciences* 58(6):1503-1518.

Ramos P, Handt O, Taylor D. (2020) Investigating the position and level of DNA transfer to undergarments during digital sexual assault. *Forensic Science International: Genetics* 47:102316.

Rand S, Schürenkamp M, Brinkmann B (2002) The GEDNAP (German DNA profiling group) blind trial concept. *International Journal of Legal Medicine* 116(4):199-206.

Rand S, Schürenkamp M, Hohoff C, Brinkmann B (2004) The GEDNAP blind trial concept part II. Trends and developments. *International Journal of Legal Medicine* 118(2):83-89.

Raymond JJ, van Oorschot RAH, Walsh SJ, Roux C (2008a) Trace DNA analysis: Do you know what your neighbour is doing? A multi-jurisdictional survey. *Forensic Science International: Genetics* 2(1):19-28.

Raymond JJ, Walsh SJ, van Oorschot RAH, Gunn PR, Evans L, Roux C (2008b) Assessing trace DNA evidence from a residential burglary: Abundance, transfer and persistence. *Forensic Science International: Genetics Supplement Series* 1:442-443.

Raymond JJ, van Oorschot RAH, Gunn PR, Walsh SJ, Roux C (2009a) Trace evidence characteristics of DNA: A preliminary investigation of the persistence of DNA at crime scenes. *Forensic Science International: Genetics* 4(1):26-33.

Raymond JJ, van Oorschot RAH, Gunn PR, Walsh SJ, Roux C (2009b) Trace DNA success rates relating to volume crime offences. *Forensic Science International: Genetics Supplement Series* 2(1):136-137.

Riman S, Iyer H, Borsuk LA, Vallone PM (2019a) Understanding the behavior of stutter through the sequencing of STR alleles. *Forensic Science International: Genetics Supplement Series* 7:115-116.

Riman S, Iyer H, Vallone PM (2019b) Exploring DNA interpretation software using the PROVEDIt dataset. *Forensic Science International: Genetics Supplement Series* 7:724-726.

Riman S, Iyer H, Borsuk LA, Vallone PM (2020) Understanding the characteristics of sequence-based single-source DNA profiles. *Forensic Science International: Genetics* 44:102192.

Riman S, Iyer H, Vallone PM (2021) Examining discrimination performance and likelihood ratio values for two different likelihood ratio systems using the PROVEDIt dataset. Available at https://www.biorxiv.org/content/10.1101/2021.05.26.445891v1 (accessed May 27, 2021).

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978):636-639.

Roberts P, Stockdale M (eds.) (2018) *Forensic Science Evidence and Expert Witness Testimony: Reliability Through Reform?* Edward Elgar Publishing: Northampton, MA.

Rodriguez JJRB, Bright J-A, Salvador JM, Laude RP, de Ungria MCA (2019) Probabilistic approaches to interpreting two-person DNA mixtures from post-coital specimens. *Forensic Science International:* 300:157-163.

Roux C, Talbot-Wright B, Robertson J, Crispino F, Ribaux O (2015) The end of the (forensic science) world as we know it? The example of trace evidence. *Philosophical Transactions of the Royal Society B* 370:20140260.

Rowan KE, Wellner GA, Grgicak CM (2016) Exploring the impacts of ordinary laboratory alterations during forensic DNA processing on peak height variation, thresholds, and probability of dropout. *Journal of Forensic Sciences* 61(1):177-185.

Ruan T, Barash M, Gunn P, Bruce D (2018) Investigation of DNA transfer onto clothing during regular daily activities. *International Journal of Legal Medicine* 132:1035-1042.

Rutty GN (2000) Human DNA contamination of mortuaries: Does it matter? *Journal of Pathology* 190(4):410–411.

Rutty GN, Hopwood A, Tucker V (2003) The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime. *International Journal of Legal Medicine* 117(3):170-174.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354.

Saiki RK, Walsh PS, Levenson CH, Erlich HA (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proceedings of the National Academy of Sciences of the United States of America* 86(16):6230-6234.

Samie L, Hicks T, Castella V, Taroni F (2016) Stabbing simulations and DNA transfer. *Forensic Science International: Genetics* 22:73-80.

Samie L, Champod C, Taylor D, Taroni F (2020) The use of Bayesian Networks and simulation methods to identify the variables impacting the value of evidence assessed under activity level propositions in stabbing cases. *Forensic Science International: Genetics* 48:102334.

Schneider PM, Gill P, Carracedo A (2006a) Editorial on the recommendations of the DNA commission of the ISFG on the interpretation of mixtures. *Forensic Science International* 160:89.

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2006b) Allgemeine Empfehlungen der Spurenkommission zur Bewertung von DNA-Mischspuren. *Rechtsmedizin* 16: 401-404. [German version was published in 2009 in English]

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *International Journal of Legal Medicine* 123(1):1-5. [originally published in German in 2006 – *Rechtsmedizin* 16: 401-404]

Schuerman C, Kalafut T, Buchanan C, Sutton J, Bright JA (2020) Using the nondonor distribution to improve communication and inform decision making for low LRs from minor contributors in mixed DNA profiles. *Journal of Forensic Sciences* 65(4):1072-1084.

Sense about Science (2017) *Making Sense of Forensic Genetics*. Available at http://senseaboutscience.org/activities/making-sense-of-forensic-genetics/ (accessed August 11, 2020).

Slooten K (2018) The information gain from peak height data in DNA mixtures. *Forensic Science International: Genetics* 36:119-123.

Smith PJ, Ballantyne J (2007) Simplified low-copy-number DNA analysis by post-PCR purification. *Journal of Forensic Sciences* 52(4):820-829.

Stanciu CE, Philpott MK, Kwon YJ, Bustamante EE, Ehrhardt CJ (2015) Optical characterization of epidermal cells and their relationship to DNA recovery from touch samples. *F1000Research* 4:1360.

Steele CD, Balding DJ (2014) Statistical evaluation of forensic DNA profile evidence. *Annual Review of Statistics and Its Application* 1:361-384.

Steele CD, Greenhalgh M, Balding DJ (2014) Verifying likelihoods for low template DNA profiles using multiple replicates. *Forensic Science International: Genetics* 13:82-89.

Steele CD, Greenhalgh M, Balding DJ (2016) Evaluation of low-template DNA profiles using peak heights. *Statistical Applications in Genetics and Molecular Biology* 15(5): 431-445.

Steensma K, Ansell R, Clarisse L, Connolly E, Kloosterman AD, McKenna LG, van Oorschot RAH, Szkuta B, Kokshoorn B (2017) An inter-laboratory comparison study on transfer, persistence and recovery of DNA from cable ties. *Forensic Science International: Genetics* 31:95-104.

Stokes NA, Stanciu CE, Brocato ER, Ehrhardt CJ, Greenspoon SA (2018) Simplification of complex DNA profiles using front end cell separation and probabilistic modeling. *Forensic Science International: Genetics* 36:205-212.

Storey JD, Weir BS (1998) *DNAMIX* software. Available at http://genomine.org/dnamix/index.html (accessed June 4, 2020).

Stringer P, Scheffer JW, Scott P, Lee J, Goetz R, Ientile V, Eckhoff C, Turbett G, Carroll D, Harbison S (2009) Interpretation of DNA mixtures—Australian and New Zealand consensus on principles. *Forensic Science International: Genetics* 3(2):144-145.

Swaminathan H, Garg A, Grgicak CM, Medard M, Lun DS (2016) CEESIt: A computational tool for the interpretation of STR mixtures. *Forensic Science International: Genetics* 22:149-160.

Swaminathan H, Qureshi MO, Grgicak CM, Duffy K, Lun DS (2018) Four model variants within a continuous forensic DNA mixture interpretation framework: effects on evidential inference and reporting. *PLOS ONE* 13(11): e0207599.

SWGDAM (2000) Short Tandem Repeat (STR) Interpretation Guidelines. *Forensic Science Communications* 2(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/strig.htm (accessed August 12, 2020).

SWGDAM (2004) Revised Validation Guidelines. *Forensic Science Communications* 6(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2004/standards/2004_03_standards02.htm (accessed August 12, 2020).

SWGDAM (2010) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.

SWGDAM (2012) SWGDAM Validation Guidelines for DNA Analysis Methods.

SWGDAM (2013) Training Guidelines. Available at https://www.swgdam.org/publications (accessed July 7, 2020).

SWGDAM (2015) SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. Available at https://www.swgdam.org/publications (accessed August 12, 2020).

SWGDAM (2016) SWGDAM Validation Guidelines for DNA Analysis Methods. Available at https://www.swgdam.org/publications (accessed July 20, 2020).

SWGDAM (2017a) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available at https://www.swgdam.org/publications (accessed July 20, 2020).

SWGDAM (2017b) Contamination Prevention and Detection Guidelines for Forensic DNA Laboratories. Available at https://www.swgdam.org/publications (accessed August 11, 2020).

SWGDAM (2018) Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios. Available at https://www.swgdam.org/publications (accessed July 20, 2020).

SWGDAM (2020) Training Guidelines. Available at https://www.swgdam.org/publications (accessed February 4, 2021).

Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2013) The potential transfer of trace DNA via high risk vectors during exhibit examination. *Forensic Science International: Genetics Supplement Series* 4(1):e55-e56.

Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2015a) DNA transfer by examination tools—A risk for forensic casework? *Forensic Science International: Genetics* 16:246-254.

Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2015b) Residual DNA on examination tools following use. *Forensic Science International: Genetics Supplement Series* 5:e495-e497.

Szkuta B, Ballantyne KN, van Oorschot RAH (2017a) Transfer and persistence of DNA on the hands and the influence of activities performed. *Forensic Science International: Genetics* 28:10-20.

Szkuta B, van Oorschot RAH, Ballantyne KN (2017b) DNA decontamination of fingerprint brushes. *Forensic Science International* 277: 41-50.

Szkuta B, Ballantyne KN, Kokshoorn B, van Oorschot RAH (2018) Transfer and persistence of non-self-DNA on hands over time: Using empirical data to evaluate DNA evidence given activity level propositions. *Forensic Science International: Genetics* 33:84-97.

Szkuta B, Ansell R, Boiso L, Connolly E, Kloosterman AD, Kokshoorn B, McKenna LG, Steensma K, van Oorschot RAH (2020) DNA transfer to worn upper garments during different activities and contacts: An inter-laboratory study. *Forensic Science International: Genetics* 46:102268.

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24(16):3189-3194.

Tan B, Zhao Z, Zhang Z, Li S, Li SC (2017) Search for more effective microsatellite markers for forensics with next-generation sequencing. *IEEE Trans Nanobioscience* 16(5):375-381.

Taroni F, Biedermann A, Vuille J, Morling N (2013) Whose DNA is this? How relevant a question? (a note for forensic scientists). *Forensic Science International: Genetics* 7:467-470.

Taylor D, Bright JA, Buckleton J (2013) The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 7(5):516-528.

Taylor D (2014) Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Science International: Genetics* 11:144-153.

Taylor D, Bright JA, Buckleton J (2014) The 'factor of two' issue in mixed DNA profiles. *Journal of Theoretical Biology* 363:300-306.

Taylor D, Buckleton J (2015) Do low template DNA profiles have useful quantitative data? *Forensic Science International: Genetics* 16:13-16.

Taylor D, Buckleton J, Evett I (2015) Testing likelihood ratios produced from complex DNA profiles. *Forensic Science International: Genetics* 16:165-171.

Taylor D, Bright JA, McGovern C, Hefford C, Kalafut T, Buckleton J (2016a) Validating multiplexes for use in conjunction with modern interpretation strategies. *Forensic Science International: Genetics* 20:6-19.

Taylor D, Buckleton J, Bright JA (2016b) Factors affecting peak height variability for short tandem repeat data. *Forensic Science International: Genetics* 21:126-133.

Taylor D, Abarno D, Hicks T, Champod C (2016c) Evaluating forensic biology results given source level propositions. *Forensic Science International: Genetics* 21:54-67.

Taylor D, Abarno D, Rowe E, Rask-Nielsen L (2016d) Observations of DNA transfer within an operational forensic biology laboratory. *Forensic Science International: Genetics* 23:33-49.

Taylor D, Bright JA, Kelly H, Lin MH, Buckleton J (2017a) A fully continuous system of DNA profile evidence evaluation that can utilise STR profile data produced under different conditions within a single analysis. *Forensic Science International: Genetics* 31:149-154.

Taylor D, Buckleton J, Bright JA (2017b) Does the use of probabilistic genotyping change the way we should view sub-threshold data? *Australian Journal of Forensic Sciences* 49(1):78-92.

Taylor D, Curran JM, Buckleton J (2017c) Importance sampling allows Hd true tests of highly discriminating DNA profiles. *Forensic Science International: Genetics* 27:74-81.

Taylor D, Biedermann A, Samie L, Pun KM, Hicks T, Champod C (2017d) Helping to distinguish primary from secondary transfer events for trace DNA. *Forensic Science International: Genetics* 28:155-177.

Taylor D, Kokshoorn B, Biedermann A (2018) Evaluation of forensic genetics findings given activity level propositions: A review. *Forensic Science International: Genetics* 36:34-49.

Taylor D, Samie S, Champod C (2019) Using Bayesian networks to track DNA movement through complex transfer scenarios. *Forensic Science International: Genetics* 42:69-80.

Taylor D, Balding D (2020) How can courts take into account the uncertainty in a likelihood ratio? *Forensic Science International: Genetics* 48:102361.

Thompson WC, Schumann EL (1987) Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior* 11(3):167-187.

Thompson WC (2009) Painting the target around the matching profile: The Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk* 8(3):257-276.

Thompson WC, Newman EJ (2015) Lay understanding of forensic statistics: evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law and Human Behavior* 39(4):332-349.

Thompson W, Black J, Jain A, Kadane J (2017) *Forensic Science Assessments: A Quality and Gap Analysis – Latent Fingerprint Examination.* American Association for the Advancement of Science. Available at https://www.aaas.org/resources/latent-fingerprint-examination (accessed June 24, 2020).

Thomson J (2018) AFSP Mixture Evaluation Collaborative Exercise. Presentation given at the English Speaking Working Group of the International Society for Forensic Genetics held September 2018 in St. Petersburg, Russia. Available at http://www.isfg.org/files/ESWG2018_Thomson_AFSP_Mixture.pdf (accessed June 27, 2020).

Tobias SHA, Jacques GS, Morgan RM, Meakin GE (2017) The effect of pressure on DNA deposition by touch. *Forensic Science International: Genetics Supplement Series* 6:e12-e14.

Torres Y, Flores I, Prieto V, López-Soto M, Farfán MJ, Carracedo A, Sanz P (2003) DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Science International* 134(2-3):180-186.

Toscanini U, Gusmao L, Alava Narvaez MC, Alvarez JC, Baldassarri L, Barbaro A, Berardi G, Betancor HE, Camargo M, Carreras-Carbonell J, Castro J, Costa SC, Coufalova P, Domínguez V, Fagundes de Carvalho E, Ferreira STG, Furfuro S, García O, Goios A, González R, de la Vega AG, Gorostiza A, Hernández A, Jiménez Moreno S, Lareu MV, León Almagro A, Marino M, Martínez G, Miozzo MC, Modesti NM, Onofri V, Pagano S, Pardo Arias B, Pedrosa S, Penacino GA, Pontes ML, Porto MJ, Puente-Prieto J, Pérez RR, Ribeiro T, Rodríguez Cardozo B, Rodríguez Lesmes YM, Sala A, Santiago B, Saragoni VG, Serrano A, Streitenberger ER, Torres Morales MA, Vannelli Rey SA, Velázquez Miranda M, Whittle MR, Fernández K, Salas A (2016) Analysis of uni and bi-parental markers in mixture samples: Lessons from the 22nd GHEP-ISFG Intercomparison Exercise. *Forensic Science International: Genetics* 25:63-72.

[TWGDAM 1989] Kearney J, et al. (1989). Guidelines for a quality assurance program for DNA restriction fragment length polymorphism analysis. *Crime Laboratory Digest* 16(2):40-59.

[TWGDAM 1991] Kearney J, et al. (1991). Guidelines for a quality assurance program for DNA analysis. *Crime Laboratory Digest* 18(2):44-75.

[TWGDAM 1995] Budowle B, et al. (1995). Guidelines for a quality assurance program for DNA analysis. *Crime Laboratory Digest* 22(2):20-43.

[UKFSR 2014] UK Forensic Science Regulator (2014) Validation, FSR-G-201. Available at https://www.gov.uk/government/publications/forensic-science-providers-validation (accessed September 9, 2020).

[UKFSR 2015] UK Forensic Science Regulator (2015) *Guidance: The Control and Avoidance of Contamination in Laboratory Activities Involving DNA Evidence Recovery and Analysis*. FSR-G-208. Available at https://www.gov.uk/government/publications/laboratory-dna-anti-contamination-guidance.

[UKFSR 2016a] UK Forensic Science Regulator (2016a) *Guidance: DNA Anti-Contamination—Forensic Medical Examination in Sexual Assault Referral Centres and Custodial Facilities*. FRS-G-207. Available at https://www.gov.uk/government/publications/sexual-assault-referral-centres-and-custodial-facilities-dna-anti-contamination.

[UKFSR 2016b] UK Forensic Science Regulator (2016b) *Guidance: The Control and Avoidance of Contamination in Crime Scene Examination Involving DNA Evidence Recovery*. FSR-G-206. Available at https://www.gov.uk/government/publications/crime-scene-dna-anti-contamination-guidance.

[UKFSR 2017] UK Forensic Science Regulator (2017) Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System. Available at https://www.gov.uk/government/collections/forensic-science-providers-codes-of-practice-and-conduct (accessed September 9, 2020).

[UKFSR 2018a] UK Forensic Science Regulator (2018a) DNA mixture interpretation, FSR-G-222. Available at https://www.gov.uk/government/publications/dna-mixture-interpretation-fsr-g-222 (accessed June 20, 2020).

[UKFSR 2018b] UK Forensic Science Regulator (2018b) Software validation for DNA mixture interpretation, FSR-G-223. Available at https://www.gov.uk/government/publications/software-validation-for-dna-mixture-interpretation-fsr-g-223 (accessed June 20, 2020).

[UKFSR 2018c] UK Forensic Science Regulator (2018c) Annual Report (November 2016 to November 2017) published on 19 January 2018 by Dr. Gillian Tully. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674761/FSR Annual_Report_2017_v1_01.pdf (accessed July 7, 2020).

[UKFSR 2020] UK Forensic Science Regulator (2020) Proficiency Testing Guidance for DNA Mixture Analysis and Interpretation. FSR-G-224. Available at https://www.gov.uk/government/publications/proficiency-testing-guidance-dna-mixture-analysis-and-interpretation (accessed October 20, 2020).

van den Berge M, Ozcanhan G, Zijlstra S, Lindenbergh A, Sijen T (2016) Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios. *Forensic Science International: Genetics* 21:81-89.

van der Gaag KJ, de Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros JFJ, de Knijff P (2016) Massively parallel sequencing of short tandem repeats-Population data and mixture analysis results for the PowerSeq™ system. *Forensic Science International: Genetics* 24:86-96.

van der Gaag KJ, de Leeuw RH, Laros JFJ, den Dunnen JT, de Knijff P (2018) Short hypervariable microhaplotypes: A novel set of very short high discriminating power loci without stutter artefacts. *Forensic Science International: Genetics* 35:169-175.

Vandewoestyne M, Van Hoofstat D, Franssen A, Van Nieuwerburgh F, Deforce D (2013) Presence and potential of cell free DNA in different types of forensic samples. *Forensic Science International: Genetics* 7(2):316-320.

van Oorschot RA, Jones MK (1997) DNA fingerprints from fingerprints. *Nature* 387:767.

van Oorschot RA, Treadwell S, Beaurepaire J, Holding NL, Mitchell RJ (2005) Beware of the possibility of fingerprinting techniques transferring DNA. *Journal of Forensic Sciences* 50(6):1417-1422.

van Oorschot RAH (2012) Assessing DNA profiling success rates: Need for more and better collection of relevant data. *Forensic Science Policy and Management* 3(3):37-41.

van Oorschot RA, Glavich G, Mitchell RJ (2014a) Persistence of DNA deposited by the original user on objects after subsequent use by a second person. *Forensic Science International: Genetics* 8(1):219-225.

van Oorschot RA, McArdle R, Goodwin WH, Ballantyne KN (2014b) DNA transfer: The role of temperature and drying time. *Legal Medicine* 16(3):161-163.

van Oorschot RAH, Found B, Ballantyne KN (2015) Considerations relating to the components of a laboratory DNA contamination minimisation monitoring (DCMM) program. *Forensic Science Policy and Management* 6(3-4):91-105.

van Oorschot RAH, Szkuta B, Meakin GE, Kokshoorn B, Goray M (2019) DNA transfer in forensic science: A review. *Forensic Science International: Genetics* 38:140–166.

Verdon TJ, Mitchell RJ, van Oorschot RAH (2013) The influence of substrate on DNA transfer and extraction efficiency. *Forensic Science International: Genetics* 7(1):167-175.

Verdon TJ, Mitchell RJ, Chen W, Xiao K, van Oorschot RA (2015) FACS separation of non-compromised forensically relevant biological mixtures. *Forensic Science International: Genetics* 14:194-200.

Vilsen SB, Tvedebrink T, Eriksen PS, Bøsting C, Hussing C, Mogensen HS, Morling N (2018a) Stutter analysis of complex STR MPS data. *Forensic Science International: Genetics* 35:107-112.

Vilsen SB, Tvedebrink T, Eriksen PS, Hussing C, Børsting C, Morling N (2018b) Modelling allelic drop-outs in STR sequencing data generated by MPS. *Forensic Science International: Genetics* 37:6-12.

Vincent FHR (2010) Report: Inquiry into the circumstances that led to the conviction of Mr. Farah Abdulkadir Jama. Available at https://www.parliament.vic.gov.au/papers/govpub/VPARL2006-10No301.pdf (accessed March 8, 2021).

Vosk T, Emery AF (2014) *Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists*. CRC Press: Boca Raton.

Voskoboinik L, Amiel M, Reshef A, Gafny R, Barash M (2017) Laundry in a washing machine as a mediator of secondary and tertiary transfer. *International Journal of Legal Medicine* 123(2):373-378.

Voskoboinik L, Motro U, Darvasi A (2018) Facilitating complex DNA mixture interpretation by sequencing highly polymorphic haplotypes. *Forensic Science International: Genetics* 35:136-140.

Wallin JM, Buoncristiani MR, Lazaruk KD, Fildes N, Holt CL, Walsh PS (1998) TWGDAM validation of the AmpFlSTR Blue PCR amplification kit for forensic casework analysis. *Journal of Forensic Sciences* 43(4):854-870.

Walsh PS, Fildes N, Louie AS, Higuchi R (1991) Report of the blind trial of the Cetus Amplitype HLA DQ alpha forensic deoxyribonucleic acid (DNA) amplification and typing kit. *Journal of Forensic Sciences* 36(5):1551-1556.

Walsh PS, Erlich HA, Higuchi R (1992) Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods & Applications* 1(4):241-250.

Walsh PS, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research* 24(14):2807-2812.

Wang DY, Chang CW, Lagacé RE, Calandro LM, Hennessy LK (2012) Developmental validation of the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit: an established multiplex assay with improved performance. *Journal of Forensic Sciences* 57(2):453-465.

Wang C, Stanciu CE, Ehrhardt CJ, Yadavalli VK (2017) Nanoscale characterization of forensically relevant epithelial cells and surface associated extracellular DNA. *Forensic Science International* 277:252-258.

Wang T, Xue N, Birdwell JD (2006) Least-square deconvolution: a framework for interpreting short tandem repeat mixtures. *Journal of Forensic Sciences* 51(6):1284-1297.

Weinberg W (1908) Uber den Nachwels der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* 64:369-382 [English translation: "On the demonstration of heredity in man," Papers on Human Genetics, Prentice-Hall (1963) translation by S.H. Boyer].

Weir BS (1995) DNA statistics in the Simpson matter. *Nature Genetics* 11:365-368.

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KA, Buckleton J (1997) Interpreting DNA mixtures. *Journal of Forensic Sciences* 42(2):213-222.

Weir BS (2000) Court experiences in the USA: People v. Simpson. *First International Conference on Forensic Human Identification*. London, 23-26 October 1999.

Weir BS (2001) DNA match and profile probabilities: comment on Budowle et al. (2000) and Fung and Hu (2000). *Forensic Science Communications* 3(1). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2001/weir.htm (accessed August 3, 2020).

Westen AA, Nagel JH, Benschop CC, Weiler NE, de Jong BJ, Sijen T (2009) Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing. *Journal of Forensic Sciences* 54(3):591-598.

Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Science International* 123(2-3):215-223.

Wickenheiser RA (2002) Trace DNA: A review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. *Journal of Forensic Sciences* 47(3):442-450.

Wickenheiser RA (2006) General guidelines for categorization and interpretation of mixed STR DNA profiles. *Canadian Society of Forensic Sciences Journal* 39:179-216.

Wickenheiser R, Farrell L (2020) Collaborative versus traditional method validation approach: discussion and business case. *Forensic Science International: Synergy* 2:230-237.

Williamson VR, Laris TM, Romano R, Marciano MA (2018) Enhanced DNA mixture deconvolution of sexual offense samples using the DEPArray™ system. *Forensic Science International: Genetics* 34:265-276.

Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annals of Human Genetics* 51(4):269-288.

Word CJ (2011) Mixture interpretation: why is it sometimes so hard? *Profiles in DNA (Promega)*. Available at https://www.promega.com/resources/profiles-in-dna/2011/mixture-interpretation-why-is-it-sometimes-so-hard/ (accessed on June 20, 2020).

Wyman AR, White R (1980) A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences of the United States of America* 77(11):6754-6758.

You Y, Balding D (2019) A comparison of software for the evaluation of complex DNA profiles. *Forensic Science International: Genetics* 40:114-119.

Young BA, Gettings KB, McCord B, Vallone PM (2019) Estimating number of contributors in massively parallel sequencing data of STR loci. *Forensic Science International: Genetics* 38:15-22.

Zadora G, Martyna A, Ramos D, Aitken C (2014) *Statistical Analysis in Forensic Science: Evidential Value of Multivariate Physicochemical Data.* Wiley: Hoboken, NJ.

Zeng X, King JL, Budowle B (2017) Investigation of the STR loci noise distributions of PowerSeq™ Auto System. *Croatian Medical Journal* 58(3):214-221.

Zoppis S, Muciaccia B, D'Alessio A, Ziparo E, Vecchiotti C, Filippini A (2014) DNA fingerprinting secondary transfer from different skin areas: Morphological and genetic studies. *Forensic Science International: Genetics* 11:137-143.