The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

| | |
|---|---|
| Document Title: | Black Box Evaluation of Bloodstain Pattern Analysis Conclusions |
| Author(s): | R. Austin Hicklin, Kevin R. Winer, Paul E. Kish, Connie L. Parks, William Chapman, Kensley Dunagan, Nicole Richetelli, Eric G. Epstein, Madeline A. Ausdemore, Thomas A. Busey |
| Document Number: | 304645 |
| Date Received: | April 2022 |
| Award Number: | 2018-DU-BX-0214 |

# Black Box Evaluation of Bloodstain Pattern Analysis Conclusions

## Final Technical Research Report

Funded by the National Institute of Justice
Award Number: 2018-DU-BX-0214

Authors:
R. Austin Hicklin (Noblis)
Kevin R. Winer (Kansas City Police Department Crime Laboratory)
Paul E. Kish (Forensic Consultant)
Connie L. Parks (Noblis)
William Chapman (Noblis)
Kensley Dunagan (Noblis)
Nicole Richetelli (Noblis)
Eric G. Epstein (Noblis)
Madeline A. Ausdemore (Noblis)
Thomas A. Busey (Indiana University)

Principal Investigator:
R. Austin Hicklin, Ph.D.
Director, Forensic Science Group
Noblis
2002 Edmund Halley Drive
Reston, Virginia 20191
hicklin@noblis.org
703-610-1995

Award recipient:
Noblis, Inc.

Project period:
January 1, 2019 to March 31, 2021

Award amount:
$1,086,976

Submitted March 29, 2021

*This project was supported by Award No. 2018-DU-BX-0214, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.*

## Abstract

Although the analysis of bloodstain evidence left at crime scenes relies on the expert opinions of bloodstain pattern analysts, the accuracy and reproducibility of their conclusions have never been rigorously evaluated at a large scale. We investigated conclusions made by 75 practicing bloodstain pattern analysts on 192 bloodstain patterns selected to be broadly representative of operational casework, resulting in 33,005 responses (from 27,038 classification prompts and 5,967 questions). We also collected 1,760 short text summary conclusions. On prompts for which the cause of the depicted bloodstain was known, 11.2% of responses were erroneous (calculated on responses from 11,634 classification prompts and 2,163 questions). The results show limited reproducibility of conclusions: 7.8% of responses contradicted other analysts. Both semantic differences and contradictory interpretations contributed to errors and disagreements, which could have serious implications if they occurred in casework.


*Note: The content of this technical report is largely comparable to a submitted manuscript detailing the findings of this study [1], with the exception of two additions: expanded data analysis detailing reporting tendencies and errors as a function of short text response types (see Sections 2.7 and 3.9) and detailed comparison of BPA terminology across various sources (see Appendix E).*

# Table of Contents

Executive Summary .......... vii
1 Introduction .......... 1
2 Materials and Methods .......... 2
2.1 Study Overview .......... 2
2.2 Participants .......... 2
2.3 Prompts and Responses .......... 4
2.3.1 Classification Prompts .......... 5
2.3.2 Questions .......... 6
2.3.3 Short Text Summary Conclusions .......... 7
2.4 Collection, Selection, and Preparation of Bloodstain Samples and Images .......... 8
2.4.1 Controlled Samples .......... 8
2.4.2 Casework Samples .......... 12
2.4.3 Sample Selection .......... 12
2.4.4 Image Preparation .......... 13
2.4.5 Sample Assignments .......... 13
2.4.6 Examples .......... 13
2.5 Assessing Known Cause for Classification Prompts and Questions .......... 16
2.6 Assessing Operational Consequence for Classification Prompts and Questions .......... 16
2.7 Evaluating Short Text Responses .......... 18
2.7.1 Extracting Classification Decisions from Short Text Responses .......... 18
2.7.2 Short Text Response Categorization .......... 20
2.8 Test Yield .......... 21
3 Results .......... 22
3.1 Accuracy .......... 22
3.2 Consensus .......... 25
3.2.1 Exclusion vs Inclusion .......... 27
3.3 Reproducibility .......... 28
3.3.1 Overall Agreement Rate and Overall Contradiction Rate .......... 30
3.3.2 Gwet's Agreement Coefficient .......... 31
3.4 Comparing Participants .......... 32
3.4.1 ROC Analysis – Detail .......... 34
3.5 Error by Participant .......... 35
3.6 Associating Participants' Performance and Background .......... 36
3.6.1 Variable Importance Analysis via Linear Regression & Random Forest .......... 37
3.6.2 Significance Testing via Kruskal-Wallis .......... 38
3.6.3 Reporting Criteria: Associating Attributes with Performance .......... 39
3.6.4 Evaluating Associations between Analyst Background and Performance .......... 40
3.7 Accuracy and Reproducibility by Pattern Type .......... 44
3.8 Semantic Issues .......... 45
3.9 Analysis of Short Text Responses .......... 46
3.9.1 Short Text Responses vs Known Cause .......... 46
3.9.2 Short Text Response Issue Categorization .......... 50
4 List of Products .......... 51
5 Dissemination of Research Findings .......... 51
6 Conclusions .......... 51
6.1 Limitations .......... 52
6.2 Recommendations .......... 53
6.2.1 Methodology and general terminology recommendations .......... 53
6.2.2 Implications for casework .......... 55
6.2.3 Lessons learned .......... 56
6.2.4 Specific terminology recommendations .......... 57
Acknowledgements .......... 65
References .......... 65
Appendix A Participant Instructions and Frequently Asked Questions (FAQs) .......... 68
Appendix B Participant Background Survey .......... 68

Appendix C Post-study Survey Results ........................................74
Appendix D Glossary/Acronyms ........................................75
Appendix D.1 ASB Terminology ........................................75
Appendix D.2 Terminology as Used in this Report ........................................77
Appendix E Comparison of BPA Terminology from Various Sources........................................79
Appendix E.1 Summary of BPA Terms from Various Sources........................................79
Appendix E.2 BPA Definitions from Various Sources........................................81

## List of Figures

Figure 1. Foam core pockets used as bullet targets ........ 9
Figure 2. Comparison of whole blood vs. reconstituted blood ........ 11
Figure 3. Comparison of whole blood vs. reconstituted blood ........ 11
Figure 4. Comparison of whole blood vs. reconstituted blood ........ 12
Figure 5. Examples of bloodstain patterns used in the study ........ 15
Figure 6. Sample #632 ........ 19
Figure 7. Sample #763 ........ 21
Figure 8. Mosaic plots of responses. ........ 23
Figure 9. Mosaic plots of responses for prompts assessed as "most consequential". ........ 23
Figure 10. Consensus on classifications and questions ........ 26
Figure 11. Sample #885 ........ 27
Figure 12. Reproducibility of responses ........ 28
Figure 13. Reproducibility of responses on prompts assessed as "most consequential" ........ 29
Figure 14. Reproducibility of responses conditioned on known cause ........ 30
Figure 15. Decisiveness plotted against six performance metrics ........ 33
Figure 16. ROC curves for analyst performance ........ 35
Figure 17. Distribution of performance measures for associations with country of practice ........ 42
Figure 18. Distribution of performance measures for associations with extent of advanced workshop training ........ 43
Figure 19. Distribution of responses as a function of extracted classifications ........ 47
Figure 20. Distribution of responses as a function of extracted classification type ........ 48
Figure 21. Participants' BPA cases by availability of evidence ........ 70

## List of Tables


Table 1. Counts of participants by number of samples completed ........ 2
Table 2. Examples of questions developed based on short text responses ........ 7
Table 3. Qualifying language rubric used for assigning decisions to extracted classifications ........ 8
Table 4. Types of bloodstain samples collected and used in study samples ........ 10
Table 5. Proportion of classifications and questions for which known cause was asserted ........ 16
Table 6. Counts of classification and question prompts, by known cause and operational consequence ........ 17
Table 7. Classification prompts ........ 18
Table 8. Rubric for coding of short text responses when classifications were not explicitly mentioned ........ 19
Table 9. Assigned types for each classification term used is this study. ........ 20
Table 10. Categories used to evaluate short text responses ........ 21
Table 11. Counts of classification and question prompts and responses ........ 22
Table 12. Counts of classification and question prompts, by known cause and operational consequence ........ 22
Table 13. Counts of responses by known cause and operational consequence ........ 24
Table 14. Accuracy rates. ........ 24
Table 15. Accuracy rates for classification prompts and questions assessed as most consequential ........ 25
Table 16. Definitions of rates reported in Table 14 and Table 15 ........ 25
Table 17. Reproducibility of responses ........ 30
Table 18. Reproducibility rates for all prompts ........ 31
Table 19. Reproducibility rates for most consequential prompts ........ 31
Table 20. Comparing OAR with Gwet's AC: all prompts ........ 31
Table 21. Comparing OAR with Gwet's AC: most consequential prompts. ........ 32
Table 22. Performance metrics used in comparing participants ........ 34
Table 23. Background attributes of interest ........ 36
Table 24. Attribute versus performance ........ 41
Table 25. Bonferroni-adjusted Dunn post-hoc analysis of Kruskal-Wallis test for performance measures exhibiting associations with country of practice ........ 42
Table 26. Bonferroni-adjusted Dunn post-hoc analysis of Kruskal-Wallis test for performance measures exhibiting associations with extent of advanced workshop training ........ 43
Table 27. Error rates and reproducibility contradiction rates for classification prompts by classification type ........ 45
Table 28. Erroneous definitive rate with respect to combinations of prompts and causes ........ 45
Table 29. Summary of short text response known cause assessments ........ 48
Table 30. Summary of short text response known cause assessments by summary type ........ 49
Table 31. Frequency of short text issue types ........ 50

## Executive Summary

Bloodstains are frequently encountered at crime scenes. The forensic discipline of bloodstain pattern analysis (BPA) involves the examination and interpretation of the attributes of bloodstains to determine causal mechanisms [2]–[5]. In some legal cases, BPA is critical evidence. For example, in the David Camm case [6]–[8] there were fundamentally contradictory opinions among BPA analysts regarding the classification of the bloodstain pattern that was the key evidence in the case. BPA analysts for the prosecution concluded that the bloodstain pattern on the defendant's clothing was backspatter from a gunshot, but BPA analysts for the defense concluded it was a transfer stain resulting from the defendant assisting his wounded children. BPA differs from many other forensic disciplines (e.g. DNA or latent fingerprint examination) in that the focus is not on source attribution (e.g. who was involved), but rather on addressing what happened at a crime scene [2], [3]. For example, BPA conclusions may provide information used in determining whether an incident was suicide or homicide, or whether a claim of self-defense is supported (or negated) by the evidence. Although BPA has been admissible in expert testimony for more than 150 years [9], the accuracy and reproducibility of conclusions by BPA analysts have never been rigorously assessed in a large-scale study. A 2009 report from the National Research Council of the National Academies strongly criticized BPA, stating "The uncertainties associated with bloodstain pattern analysis are enormous" and "In general, the opinions of bloodstain pattern analysts are more subjective than scientific" [10]. The National Research Council called for testing of error rates in forensic disciplines, which was echoed in a 2016 report by the President's Council for Science and Technology [11], [12].

We conducted this "black box" study [13] to evaluate the accuracy and reproducibility of conclusions made by practicing BPA analysts. Several BPA studies have previously been conducted [14]–[19], but not with the scale or breadth of the current study. Participation was limited to practicing BPA analysts. As the background survey results illustrate (Appendix B), our participants' formal education, training, and experience represent a diverse group of analysts from 14 countries (57% from the U.S.). Survey results indicated that the participants generally perform BPA only as one of their responsibilities: nearly half (47%) of the participants perform fewer than five BPA cases per year; 83% have testified in court.

Although the BPA discipline has adopted a terminology standard [20], there is no preexisting conclusion standard that could be adopted for use in this study. Therefore, we developed three complementary approaches to collect participants' assessments of the mechanism(s) that caused each sample: *classification prompts, questions,* and *short text summary conclusions*. Each participant who completed the study received 30 samples for which short text summary conclusions were required and 120 samples assigned for multiple classification prompts and/or questions.

Samples included bloodstain patterns collected under controlled conditions and from operational casework (123 and 69 samples, respectively). The controlled creation of bloodstains followed common practices used in the development of training materials for BPA analysts. The operational samples used in this study were from closed cases. In the post-study survey, 72% of the respondents said the difficulty of the samples in the study was similar to casework; 23% said it was harder or much harder than casework.

Classification prompts used terminology explicitly from the prevailing Academy Standards Board (ASB) standard [20], and participants responded using multiple choice options of *definitive*, *included*, or *excluded.* A response of *included* indicated there was insufficient support either for a *definitive* decision or for an *excluded* decision. The criteria for these decisions have not been clearly defined in the BPA discipline. Because of this lack of a standardized sufficiency threshold, *included* was considered as indeterminate in analyses, neither correct nor an error.

Questions were added over the course of the study as a means of evaluating the reproducibility of statements made by participants in the short text responses: participants were assigned samples for short text responses early in the study so that other participants would subsequently be assigned questions derived from those text responses. Our questions often addressed reconstruction issues encountered in BPA that go beyond strictly pattern classification, such as "Was this the result of two cast-off patterns?" or "Was the decedent standing up when [the] bloodletting event occurred?" Questions were assessed using multiple choice options of *yes*, *possible*, or *no*; *possible* was considered as indeterminate in analyses.

For short text responses, participants were instructed to write two concise statements (each less than about 75 words), stating their observations of the depicted bloodstain patterns(s), and discussing the perceived mechanism(s) responsible for these pattern(s) — i.e., make a pattern classification. Based on review of the Beta testing results, which showed notable variability in terminology use, participants were provided basic guidance for writing text responses; otherwise, participants were afforded complete latitude to express conclusions. Short text responses collected as part of this study served two primary purposes—to generate questions (which then fed into the primary study results) and to facilitate analyses specific to these open-ended responses.

Reproducibility of responses was assessed on all samples; accuracy was assessed only when the cause of the bloodstain was known. For each prompt (classification or question), a correct ("known cause") response was generally available for controlled collection samples and was not available, nor asserted, for casework samples. Although the overall mechanism for each controlled sample was known, for prompts considered debatable or semantic we left cause as "unknown." For casework samples we did not claim certain knowledge of the cause of each bloodstain. We asserted known cause for 47% of prompts (81% of prompts for controlled samples and 0% of prompts for casework samples). Note that asserting that the cause of a bloodstain is known does not necessarily imply that a given sample has sufficient information to make a definitive attribution of that cause. Since BPA has no standardized criteria for determining the types or quantities of characteristics needed to make a given decision, consensus among analysts is currently the only available means of assessing the appropriateness of an indeterminate response (*included* or *possible*).

In broad overview, our results showed that conclusions by BPA analysts were often erroneous and often contradictory of other analysts. Such errors could have serious implications if they occurred in casework, as would conflicting conclusions among BPA analysts if those resulted in conflicting court testimony. Many of the disagreements among BPA analysts — and some of the errors — may be attributable to semantic differences. There is often a lack of agreement on the meaning and usage of BPA terminology and classifications, suggesting a need for improved standards. The lack of criteria for classification decisions means that we cannot expect high rates of reproducibility among analysts.

Concerning accuracy more specifically, we found that out of all 11,634 classification prompts for which there was a known cause, 11.2% of responses contradicted the known cause and therefore were erroneous (weighted average of erroneous definitive and erroneous excluded responses). For the 2,163 questions for which there was a known cause, 11.0% of responses were erroneous (weighted average of erroneous yes and erroneous no responses). Responses were indeterminate for 30.1% of classifications and 43.8% of questions. If we consider only determinate responses (*definitive*, *excluded*, *yes*, *no*) on prompts with known cause, 83.0% of responses were correct with respect to known cause ("overall predictive value").

Two other calculations of accuracy are of note. First, in order to limit the effects of prompts that could be seen as minor or semantic, the BPA analysts on the study team evaluated each prompt to determine whether an error on that prompt would be highly consequential in an actual case. These were labeled "most consequential" and accounted for 22.5% of the prompts. If limited to the most consequential prompts, the error rate for classifications was 9.0% and the error rate for questions was 5.8%; the overall predictive value was 86.6%.

Second, of 1,760 short text statements, there were 1,052 that could be evaluated with respect to known cause, and of these, 4.8% entirely contradicted known cause, and 11.2% partially contradicted known cause (i.e., included both correct and incorrect statements). Team member BPA experts also evaluated the quality and thoroughness of the short text summary conclusions, assessing whether the analysts' observations and conclusions were adequately supported. The BPA experts determined that 11.3% had errors in reconstruction, observations, or unsupported conclusions.

We found that consensus was limited, and failures of consensus were widely distributed across prompts. Only 3% of prompts received unanimous responses, 33% of prompts had at least 75% consensus, and 81% of prompts had a majority consensus. Majority responses were rarely incorrect. For classification prompts, responses with a 95% supermajority never contradicted known cause; 75% supermajority responses contradicted known cause on four prompts (1.0% of 392 classification prompts with known cause); a majority contradicted known cause on an additional five prompts (9 total, 2.3%); and a plurality contradicted known cause on an additional eight prompts (17 total, 4.3%). On questions, a 75% supermajority never contradicted known cause, a majority (or plurality) contradicted known cause on one question (1.0% of 96 questions with known cause). When limited to the most consequential prompts, the majority was always correct.

Concerning reproducibility, we found that for every participant who responded *excluded* to a classification prompt, 65.3% of other participants also responded *excluded* to that prompt (agreement rate), 25.3% responded *included*, and 9.4% responded *definitive* (contradiction rate). Across all classification prompts and questions, the agreement rate (the proportion of other participants who had an identical response to a given prompt) was 54.6%; the contradiction rate (the proportion of other participants who had a diametrically opposed response) was 7.8%. If limited to the most consequential prompts, the agreement rate was 56.3% and contradiction rate was 6.2%. Contradictions were distributed broadly across prompts: 549 of the 815 classification prompts and 146 of the 223 questions resulted in contradictions.

When reproducibility of responses is conditioned on known cause, erroneous responses were often reproduced: 17.7% of erroneous *definitives* were reproduced, as were 34.1% of erroneous *excluded* responses, 24.2% of erroneous *yes* responses, and 22.5% of erroneous *no* responses. These results suggest that if two BPA analysts both analyze a pattern (such as occurs operationally during technical review) they cannot always be expected to agree, and if they do agree they may both be wrong.

An important result of our study was that many of the disagreements between analysts — and some of the errors — may be attributed to semantic differences rather than contradictory interpretations. Such semantic issues include inadequate delineation between certain pattern types and ambiguity of some definitions. For example, participants often incorrectly concluded that *splash* patterns were *drip patterns* (34% erroneous definitive rate), indicating a lack of general agreement on the delineation between *splash* and *drip patterns*. In the post-study survey participants indicated that "several examples stretched the semantic interpretation of definitions," in particular regarding *projected* and *impact* patterns; this was supported by analysis of the short text responses. Although some semantic disagreements would presumably be unlikely to have significant consequences in actual casework, their prevalence obscures the extent of serious disagreements. This lack of agreement on the meaning and usage of BPA terminology and classifications illustrates the need for improved standards.

The results presented here are intended to provide estimates for use in decision making, improving procedures and training, and future research. The results should not be taken to be precise measures of operational error rates. The discipline of bloodstain pattern analysis is not solely defined by pattern classification, but rather it includes multiple other aspects that were not evaluated within this study. This study differed from operational casework in that analysts were asked to provide responses based solely on photographs, were not provided case-relevant facts that may have aided in making conclusions, and were required to report conclusions in a manner atypical for the BPA discipline (i.e., in part, by multiple choice options). These results do not account for operational quality assurance measures, such as technical review or verification.

# 1 Introduction

Bloodstains are frequently encountered at crime scenes. The forensic discipline of bloodstain pattern analysis (BPA) involves the examination and interpretation of the attributes of bloodstains to determine causal mechanisms [2]–[5]. In some legal cases, BPA is critical evidence. For example, in the David Camm case [6]–[8] there were fundamentally contradictory opinions among BPA analysts regarding the classification of the bloodstain pattern that was the key evidence in the case: BPA analysts for the prosecution concluded that the bloodstain pattern on the defendant's clothing was backspatter from a gunshot, but BPA analysts for the defense concluded it was a transfer stain resulting from the defendant assisting his wounded children. BPA differs from many other forensic disciplines (e.g. DNA or latent fingerprint examination) in that it is not focused on source attribution (e.g. who was involved), but rather on addressing what happened at a crime scene [2], [3]. For example, BPA conclusions may provide information used in determining whether an incident was suicide or homicide, or whether a claim of self-defense is supported (or negated) by the evidence.

Although BPA has been admissible as expert testimony for more than 150 years [9], the accuracy and reproducibility of conclusions by BPA analysts have never been rigorously assessed in a large-scale study, this leaves a critical gap in the court testimony of the data presented in legal proceedings. A 2009 report from the National Research Council of the National Academies strongly criticized BPA, stating "The uncertainties associated with bloodstain pattern analysis are enormous" and "In general, the opinions of bloodstain pattern analysts are more subjective than scientific" [10]. The National Research Council called for testing of error rates in forensic disciplines, which was echoed in a 2016 report by the President's Council for Science and Technology [11], [12].

The goal of this black-box study was to assess the scientific basis of bloodstain pattern analysis BPA by rigorously measuring the accuracy and reproducibility of BPA conclusions, associating these results with the education and training of participants. This study directly addresses one of the National Institute of Justice's (NIJ's) key goals and objectives: "Perform studies that examine the degree of accuracy and reliability of methods used by forensic scientists to achieve a more complete understanding of the scientific basis of forensic evidence and the interpretation of that evidence."

# 2 Materials and Methods

## 2.1 Study Overview

Participants were asked to classify 150 distinct bloodstain patterns over a period of four months (2 Oct 2019 through 3 Feb 2020). Samples were selected from a pool of 192 samples, which were obtained from both controlled collection and operational casework. The study was conducted using custom web-based software that presented bloodstain images and recorded test responses. Medium-resolution images were shown in the browser, and full-resolution images could be downloaded. Prior to commencement of the study, Alpha and Beta pilot tests were conducted to assess software functionality, instructions, and analyst experience. The pilot study test materials were not reused in the formal study. The following sections provide details on the study design and data selection.

## 2.2 Participants

Participation was open to all practicing U.S. and international analysts who had conducted operational casework examinations within the 2 years preceding the study announcement. Seventy-five BPA analysts were included in analyses (not including six BPA analysts who completed three or fewer test samples each and were omitted from all analyses). Forty-five of the participants completed all (150) assigned test samples, 11 participants completed 58-149 samples, and 19 participants completed 5-45 samples (Table 1). For the subset of analyses in which we calculated individual rates for each participant to compare performance, we limited analyses to the 56 participants who completed at least 50 samples.

| *Samples completed* | *# participants* | |
|---|---|---|
| 1-3 | 6 | (omitted from all analyses) |
| 5-45 | 19 | (included in all analyses except for comparisons of participants) |
| 58-96 | 7 | (included in all analyses) |
| 125-149 | 4 | |
| 150 | 45 | |

*Table 1. Counts of participants by number of samples completed. 75 participants were used for analyses; 56 participants were used in comparisons of participants.*

Participation was solicited at the 2019 International Association for Identification (IAI) International Educational Conference, at the 2019 International Association of Bloodstain Pattern Analysts (IABPA) Conference, and via notices in the IAI *Identification News* and the IABPA website.

Use of BPA analysts in the study was approved by the Advarra Institutional Review Board (IRB), following NIJ requirements for human subjects research. Informed consent was obtained from all participating analysts. Anonymity was part of the IRB approval for this study: participants were assured that all results would be anonymous, and results would not be aggregated in a way that compromises anonymity. Anonymity was maintained through multiple levels of data anonymization, data segregation, and information flow control. Participant identifiers were anonymized prior to data analysis, precluding the analysis team's ability to cross associate participants' personally identifying information with questionnaire responses or test results. Cross-references between personal information and results were destroyed prior to the public presentation or publication of results.

Note that participants were volunteers and cannot be considered a randomly selected statistical sample. There is no practical way to know the size of the overall worldwide population of practicing BPA analysts, nor the demographic breakdown of all BPA analysts. However, note that out of 52 IAI certified bloodstain pattern analysts [21], 11 participated in the study.

Participants were required to complete a consent form and background questionnaire prior to the study (detailed in Appendix B). Participants who completed the study were asked to complete a brief post-study survey (detailed in Appendix C). The participants encompassed a range of experience, training, and affiliations. The participants were from 14 countries; 57% were from the U.S. Of the 75 participating analysts, 44% held advanced degrees (i.e., Master's or PhD); 35% had 11 or more years of BPA experience, and 39% had 5-10 years of experience; 15% were certified as BPA analysts by the IAI (an additional 15% were certified by other organizations). More than half of the participants (53%) reported receiving no formal BPA training; they either received informal on-the-job training or attended courses and workshops. Nearly a third of participants (31%) did not do supervised bloodstain pattern analysis casework before performing bloodstain pattern analysis on their own. Most analysts perform BPA only as one of their responsibilities, which was reflected in the survey results: nearly half (47%) of the participants perform fewer than five BPA cases per year. Because this study focuses on analyses of photographs, we asked participants the proportion of their casework that uses scene photographs: three participants reported that they never conduct examinations of scene photographs; 14 reported they examine scene photographs in less of 10% of cases; 28 reported they use photographs in at least half of their cases.

In addition to the background survey conducted prior to the study, participants who completed the study were asked to complete a brief post-study survey. The post-study survey provided a single question asking participants how the samples overall compared to actual casework, and provided an opportunity to make comments.* Out of the 75 participants, 47 participants completed the post-study survey, and 32 participants provided comments. Of the subset of participants who completed the post-study survey, 72% (34 participants) indicated the study was similar to casework in terms of difficulty; 23% (11 participants) indicated the study was harder or much harder than casework.

The comments are included verbatim in Appendix C. The general categories of comments are summarized here:

- 13 Compliments
- 6 Wanted ancillary information (not just images)
- 4 Limited opinion due to lack of crime scene visit
- 4 Issues with wording of questions
- 4 Issues with photo image quality
- 3 Classification definition semantics
- 3 Issues with "possible" and "included" categories

Upon completion of the study, each participant was given the option to obtain an anonymous identification number that would provide access to the individual results. Only participants who completed all (150) assigned samples were afforded this opportunity. The study team has no way

* *Although the participant instructions indicated that the post-study survey contained a question regarding the total time spent on the study, due to a software programming issue, that question was not included in the final survey.*

of knowing which (if any) of the participants who completed the study chose to obtain their anonymous identification number. Participants were cautioned:

> *Note that if you choose to access your personal results, that information may then be under legal discovery when you act as an expert witness in the future. In some legal systems, knowledge of your test score may create an obligation for you to disclose the test score in a criminal, civil or regulatory proceeding for which you are called to testify or provide evidence. If you elect to request your individual results, you are advised to consider first consulting with your agency's counsel or counsel of your choice.*

### 2.3 Prompts and Responses

The BPA discipline has developed a standard for terminology through the Organization of Scientific Area Committees for Forensic Science (OSAC) and the Academy Standards Board [20], which has been widely adopted by the BPA community [21]–[24]. However, there is no preexisting widely-used conclusion standard that could be adopted for use in this study. We developed three complementary methods to collect participants' assessments: classifications, questions, and short text summary conclusions. This approach provided three different tools to collect a range of information about every sample, and about every participant:

- Classification prompts were included to assess how accurately and consistently BPA analysts could classify patterns, as well as to evaluate the OSAC/ASB terminology standard. (Discussed further in Section 2.3.1)
- The short text summaries were included to assess how participants would classify patterns without prompting. (Section 2.3.3)
- The questions were included to evaluate and make effective use of the short text responses made by other participants, using multiple stages of assignments. (Section 2.3.2)

We developed these methods based on extensive testing and feedback during Alpha testing, resulting in the use of three-level responses, and the specific wording for the responses and their definitions. Alpha testing initially was based on five-level Likert scales, but Alpha testing feedback from BPA analysts strongly opposed that approach, because we could not control or unambiguously define the thresholds between the middle categories.

Each participant who completed the study received 30 samples assigned for short text summary conclusions, and 120 samples assigned with classification prompts and/or questions. We emphasized to participants that classification terminology was explicitly based on the OSAC/ASB standard terminology [20]: the introductory emails and instructions stated that they needed to be familiar with the ASB terminology, and the website software provided popup definitions for all ASB terms used.

Participants were provided detailed instructions (Appendix A). During the testing period, any communications with participants (other than resolving software access issues) were provided to all participants in a Frequently Asked Questions (FAQs) file (Appendix A). No changes to the instructions were made after the start of the study period. The instructions, FAQs, and the ASB recommended BPA terms and definitions were emailed to interested parties and were available via the study website.

When making conclusions, participants were told to assume that these were the best images available, no other evidence was available, and the substance in the images had been ascertained to be human blood. Participants were instructed to base conclusions on what could be seen in the images provided, and that generally there were no other relevant bloodstains in the vicinity, but in some cases images deliberately focused on a specific area in order to obtain assessments independent of the surroundings.

Participants were permitted to access up to five unsubmitted samples at a time, but responses to all five samples needed to be submitted before being presented with the next five samples. After submitting the response for a bloodstain, participants were asked to confirm the selections made. After submission, the website did not allow participants to return to completed samples or revise previous answers.

Not all samples were made available at the start of the study; rather they were released in three stages: the first stage (50 samples for each participant) was available at the start of the study period (2 October 2019), the second stage (the next 50 samples) was released 6 November 2019, and the third stage (the last 50 samples) was released 6 January 2020; the study was closed 2 February 2020. All of the short text responses were collected in the first two stages for each participant, comprising $^{1}/_{3}$ of the first 90 samples, but $^{1}/_{5}$ of all samples. This allowed the study design team to add questions (and a few classification prompts) based on the content of the short text responses received for each sample — thereby allowing the participants to define most of the questions asked, so that the questions reflected the participants' observations, and so that the reproducibility of statements made in short text responses could be evaluated. Most of the questions were added in the second and third stages (181 of the 234 questions), as well as some classification prompts (70 of 822 classification prompts). During analysis, classification prompts and questions that did not receive at least five responses each were ignored, so for analysis the total number of classification prompts used was 815 (omitting 7), and the total number of questions used was 223 (omitting 11).

### *2.3.1 Classification Prompts*

For the samples assigned with classification prompts, participants were presented with up to six BPA classifications for each sample. Classification prompts used terms explicitly from the prevailing ASB standard [20], which were shown in square brackets in the BPA-BB website. The study website showed the definition for OSAC/ASB terms (hovering the mouse over the bracketed term showed the ASB definition).

Participants were instructed to indicate the mechanisms that described the bloodstain pattern, based upon the following 3-level response scale:

- DEFINITIVE — This classification/mechanism is correct and accurate for this bloodstain pattern. You are certain that the given mechanism was involved in the creation of the pattern.
- EXCLUDED — This classification/mechanism cannot be responsible for creating this bloodstain pattern. You are certain that the given mechanism could NOT have been involved in the creation of the pattern.
- INCLUDED — This classification/mechanism is neither Definitive nor Excluded.

Participants were instructed that for each sample, zero, one, or more of the classifications may be valid, and that multiple mechanisms may be *definitive* (for example, if both broad and specific classifications were listed as options, or if multiple mechanisms were involved in the creation of the bloodstain pattern).

Each sample had 0-6 classification prompts (mean 4.2), for a total of 815 classification prompts. See *Supplemental Materials Data S2* for the classification prompts for each sample.

### *2.3.2 Questions*

Questions provided a means to address mechanisms that could not be described solely by classifications, often in areas that touch on reconstruction, such as the order or direction of causal mechanisms. For the samples assigned with questions, participants were asked to respond to a series of up to five questions per sample regarding possible mechanism(s) that may have caused the pattern(s), and to consider each question as if they were being asked the question as an expert witness in court.

Participants were instructed to answer the questions, based upon the following 3-level response scale:

- YES — Indicates that you agree with the entire question as stated
- NO — Indicates that you disagree with the entire question as stated
- POSSIBLE — Indicates either that you are not certain, or that the question is partly true and partly false.

In general, questions were stated as definite causes — "Was this bloodstain pattern created by [x]?" or "Did [x] cause this bloodstain pattern?" For these questions, participants were instructed that answering YES indicated certainty that [x] caused the bloodstain pattern (i.e., *definitive*), answering POSSIBLE indicated that [x] may have been a possible cause of the bloodstain pattern(s) (i.e., *included*), and answering NO indicated certainty that [x] did not cause the bloodstain pattern(s) (i.e., *excluded*).

Each sample had 0-5 questions (mean 1.2), for a total of 223 question prompts. Note that many questions were added over the course of the study as a means of evaluating statements made in the short text responses (Table 2).

Additional examples of questions:

- Was blood deposited upon the toilet while the toilet was in more than one position?
- Was label D created by the muzzle imprint of a handgun?
- Was the blood source moving from A4 to A1?
- Was the revolver on the floor prior to bloodstain pattern deposition?
- Was this caused by a gunshot?
- Was this the result of at least three [cast-off] events?
- Were these bloodstains caused by multiple events at different times?

| *SampleID* | *Short text response 2 (Mechanism)* | *Question created based on short text response* |
|---|---|---|
| 148 | The presence of mist type indicates very high Velocity which is consistent with forward spatter or impact pattern with very high force. | Does the size of the bloodstains indicate a very high velocity? |
| 244 | The spurt was likely caused by a major artery being cut causing the high velocity and volume of spatter present on the surface. | Was this pattern caused by a major artery being cut? |
| 258 | Drip stain. Altered stain. The drip stain was possibly altered by something blowing air from the right to the left of the stain. | Was the bloodstain altered by something blowing air from the right to the left of the stain? |
| 585 | The higher defect in the paint is consistent with an impact from a serrated blade. The fine linear transfers are from the points of the blade. Transfer on the inferior edge is from the side of the blade. Spatter and directional stains above comprise a cessation pattern caused by impact of the blade against the door. Minimal blood about the lower defect indicates it occurred subsequent to the higher one. | Does the minimal blood about the lower defect (to the right and below U) indicates it occurred subsequent to the higher one (right of U)? |
| 913 | Bloodstain spatter, created with two or more impact mechanisms. | Was this the result of two or more impact mechanisms? |
| 915 | Several impact spatter patterns, with similar area of convergence. | Was this the result of several impact events, with similar areas of convergence? |
| 937 | Blood dripping, due to gravity, from an object. The lack of directionality of the blood drops indicates the object was moving at a slow speed. | Does the lack of directionality of the blood drops indicate that the source (from which blood was dripping) was moving at a slow speed? |

*Table 2. Examples of questions developed based on short text responses.*

### 2.3.3 Short Text Summary Conclusions

For these samples, participants were instructed to write two concise statements (each less than about 75 words), briefly stating their observations of the depicted bloodstain patterns(s), and briefly discussing the mechanism(s) responsible for the depicted bloodstain pattern(s) — i.e., make a pattern classification.

The instructions provided the following guidance on writing short text summary conclusions (nicknamed "tweets") — based on review of Beta testing results, which showed notable variability in how people used terms such as "indicates," "consistent with," and "support for."

> *In writing your tweets, we want to give you freedom to express your conclusions as you want (within the length requirements) — but at the same time we need to clearly interpret what you mean. Therefore, phrase your conclusions as you wish, but note that this is how we will interpret what you say. If you disagree with us regarding these interpretations of wording, please be more explicit or reword.*
>
> - *We treat any of these as synonymous: ["is" "was" "was caused by" "was produced by" "definite" "certain"] — OR if you list a mechanism by itself (e.g. "Drip stain")*
> - *We treat any of these as synonymous: ["strong support for" "presumably" "the most likely cause" "probably" "likely" "most likely" "indicates"]*
> - *We treat any of these as synonymous: ["included" "cannot exclude" "possible" "consistent with" "support for"]*
> - *We treat any of these as synonymous: ["weak support for" "unlikely" "could conceivably be" "improbable"]*
> - *We treat any of these as synonymous: ["excluded" "impossible" "was NOT" "was NOT caused by" "was NOT produced by" "definitely NOT" "certainly NOT"]*

| *Scale* | *Qualifying terms and phrases* | | |
|---|---|---|---|
| Definitive | Is<br>Was<br>No/none | Was produced by<br>Was caused by | Definite<br>Certain |
| Included | Strong support for<br>Presumably<br>Most likely cause(d)<br>Characteristic of<br>Included<br>Cannot exclude<br>Possible<br>Consistent with<br>Weak support for | Likely<br>Classified as<br>Most likely<br>Exhibiting<br>Support for<br>Looks like<br>Might<br>Seems<br>Unlikely | Indicates<br>Appears<br>Suggests<br>Probably<br>Could conceivably be<br>Or<br>Maybe<br>Unknown if<br>Improbable |
| Excluded | Exclude<br>Impossible<br>Was NOT | Was NOT caused by<br>Was NOT produced by<br>Definitely NOT | Certainly NOT<br>No |

*Table 3. Qualifying language rubric used for assigning decisions to extracted classifications.*

Short text responses collected as part of this study served two primary purposes—to generate questions (which then fed into the primary study results) and to facilitate analyses specific to these open-ended responses (discussed in 3.9).

## 2.4 Collection, Selection, and Preparation of Bloodstain Samples and Images

The bloodstain samples used in the study were selected from a broader pool of bloodstains that was comprised of samples collected under controlled conditions ("controlled samples") and bloodstains from actual casework ("casework samples"). In the post-study survey (Appendix C), 72% of the respondents said the difficulty of the samples in the study was similar to casework; 23% said it was harder or much harder than casework.

See Supplemental Data S2 for low-resolution summary images for all samples; see [25] for full-resolution imagery for all samples.

### *2.4.1 Controlled Samples*

The controlled creation of bloodstains followed common practices used in development of training materials for BPA analysts, in order to replicate characteristics that are routinely found in actual casework. A total of 231 samples were produced under controlled laboratory conditions at the Kansas City Police Crime Laboratory (KCPCL, Kansas City, Missouri); 123 of these samples were selected for use in the study. Such controlled collection samples are commonly used in BPA training, and are necessary in order to not harm individuals or animals. The human blood employed to create these samples was either donated by study team members or purchased from the Community Blood Center (4040 Main Street, Kansas City, Missouri). A KCPCL staff phlebotomist drew the blood collected from team members. KCPCL personal safety and hazardous waste disposal protocols were adhered to throughout data collection. The controlled collection bloodstains were created inside the KCPCL laboratory at typical indoor office temperatures, except for a few (3%) bloodstains created in uncontrolled outdoor conditions and sub-freezing temperatures (about -3°C). The bloodstains were photographed by the BPA analysts on the study team using the same procedures as in actual casework (using a Canon EOS 70D, Canon EF-S 18-55mm f/3.5-5.6 IS II lens). The creation of the controlled collection bloodstains was videotaped for almost all samples: 110 of the 123 controlled samples used in the study were videotaped (see Supplemental Data S4 for example videos). For quality control, sample identifiers were visually captured in all photographs and both verbally and visually documented in all videos. Table 4 summarizes the pattern types and substrates for the controlled collection samples.

Additional details on the collection of some bloodstain pattern types:

- Gunshot patterns were created by firing at either a sponge soaked in blood or a pocket filled with blood made of foam core (using the method described in [26]; see Figure 1). Firearms used were 9mm Beretta semi-automatic pistol and .357 Magnum Ruger revolver, both shooting full metal jacket rounds (9mm: Federal #WM5199, 115 grain; .357: Magtech #357D, 158 grain). The muzzle-target distance was sufficient to minimize the effects of muzzle gases (approximately 3 meters between the muzzle and blood source). See Supplemental Data S4 for an example video.
- Projected patterns used a manual pump, syringe, or pipette. Mr. Kish has used this approach frequently in creating bloodstains simulating arterial projection patterns in BPA training. See Supplemental Data S4 for an example video.
- Expiration patterns were caused by expiration/coughing by a team member using his own freshly drawn blood. See Supplemental Data S4 for an example video.

*Figure 1. Foam core pockets used as bullet targets*

| *Pattern type* | *Samples collected* | *Samples used in study* |
|---|---|---|
| Altered flow | 4 | 2 |
| Castoff | 10 | 5 |
| Cessation | 6 | 4 |
| Dragged body | 3 | 2 |
| Drip pattern | 16 | 14 |
| Drip pattern (running) | 1 | 1 |
| Drip stain | 33 | 6 |
| Drip stain, diluted blood | 3 | 0 |
| Drip trail, running | 3 | 3 |
| Drip trail, walking | 3 | 3 |
| Expiration | 16 | 9 |
| Freefall volume of blood | 15 | 9 |
| Fresh whole blood pour | 1 | 1 |
| Gunshot: backspatter | 11 | 3 |
| Gunshot: backspatter and forward spatter | 2 | 0 |
| Gunshot: forward spatter | 16 | 5 |
| Gunshot: horizontal | 6 | 5 |
| Impact | 9 | 8 |
| Impact with castoff | 1 | 0 |
| Impact, clotted blood | 2 | 1 |
| Microflow: Spatter, then rapid movement of object | 4 | 3 |
| Pool: poured blood | 2 | 2 |
| Projected down | 13 | 5 |
| Projected down diagonal | 3 | 2 |
| Projected horizontally | 2 | 0 |
| Satellite stains from drip pattern | 2 | 2 |
| Saturation: poured blood | 2 | 2 |
| Saturation, then alteration | 2 | 2 |
| Spatter (Satellite stains from drip pattern) applied to existing transfer | 9 | 4 |
| Transfer | 12 | 7 |
| Transfer of clotted blood | 3 | 2 |
| Transfer of diluted blood | 3 | 2 |
| Transfer with clot | 1 | 1 |
| Wipe of clotted blood pool | 1 | 1 |
| Wipe of dried blood pool | 2 | 2 |
| Wipe of dry drip stain | 3 | 3 |
| Wipe of partially dry drip stain | 6 | 2 |

| *Substrate type* | *Samples collected* | *Samples used in study* |
|---|---|---|
| 110# cardstock | 22 | 9 |
| Asphalt | 4 | 2 |
| Asphalt + Van | 1 | 1 |
| Cardboard | 19 | 7 |
| Carpet | 32 | 23 |
| Ceiling | 1 | 1 |
| Countertop | 2 | 2 |
| Drywall | 36 | 17 |
| Drywall + Cardboard | 1 | 1 |
| Exterior siding | 3 | 2 |
| Floor (epoxy) | 2 | 2 |
| Flooring (hardwood) | 23 | 10 |
| Flooring (hardwood) + cardboard | 1 | 1 |
| Flooring (Linoleum) | 3 | 2 |
| Flooring (tile) | 21 | 11 |
| Flooring (vinyl) | 13 | 6 |
| Flooring (vinyl) + paneling | 2 | 2 |
| Furniture | 4 | 3 |
| Hatchet | 1 | 1 |
| Interior paneling | 16 | 6 |
| Paper | 4 | 2 |
| Pillow | 2 | 2 |
| Plywood (construction grade) | 8 | 3 |
| Plywood (Hardwood) | 1 | 0 |
| Tools | 5 | 4 |
| Wood (2x4) | 3 | 2 |
| Wood molding | 1 | 1 |

*Table 4. Types of bloodstain samples collected and used in study. (231 controlled collection samples; 123 used in study)*

The human blood sourced from the Kansas City Community Blood Bank was in the form of heparinized packed red blood cells (pRBCs), as opposed to whole blood. pRBCs are a derivative of whole blood in which the majority of the plasma component has been removed via centrifugal action for use in transfusion; heparin is added to prevent it from coagulating. For use in this study, the pRBCs were diluted with a 9% saline solution to approximate the proportion of red blood cells in whole blood (using proportions of 55% pRBC and 45% saline). Use of such reconstituted blood is typical for use in BPA training, and in previous studies [16], [17].

To evaluate whether the bloodstains created using reconstituted blood differed from bloodstains created using whole blood, an internal validation was conducted using both diluted packed cells and fresh, whole blood drawn from a study team member. The team member's blood was utilized immediately upon draw. The reconstituted blood was room temperature, approximately 20-22°C. As a validation test, three single drip stains were created on 110-pound card stock from each of three distinct drop heights (10, 50 and 100 cm), for both the reconstituted blood and the freshly drawn blood, shown in Figure 2 through Figure 4. We found no notable differences in the physical characteristics of the bloodstains between the whole blood and reconstituted blood.

The controlled collected samples in the study used reconstituted blood with these exceptions:

- The controlled collected samples that were expirated (coughed) used that team member's own (freshly drawn) blood.
- Whole blood was used for examples of clotting and serum separation. Note that due to the heparin, reconstituted blood does not clot and the serum does not separate as with whole blood.
- Whole blood was used for clots used in transfers or impact patterns.

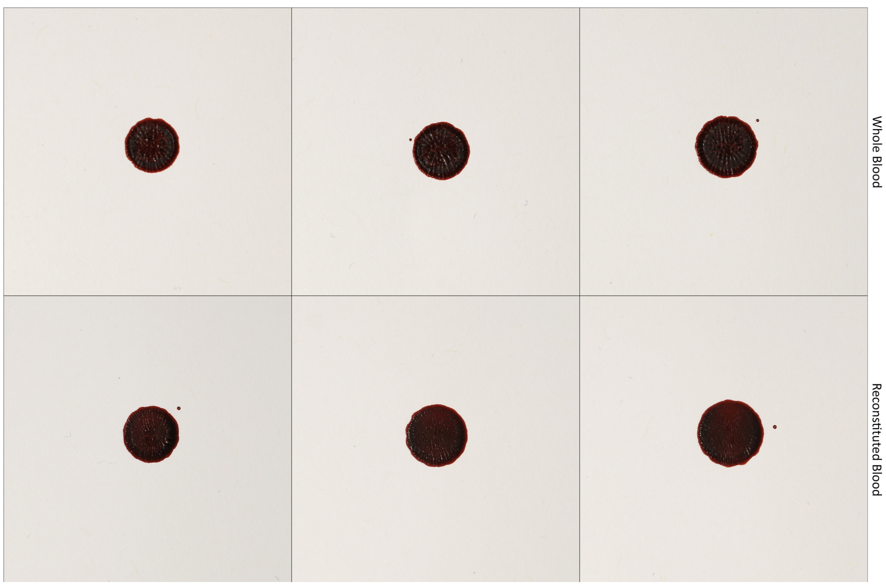


*Figure 2. Comparison of whole blood vs. reconstituted blood. Spatter of single drop from 10 cm onto 110# cardstock. Each image is 5x5 cm.*

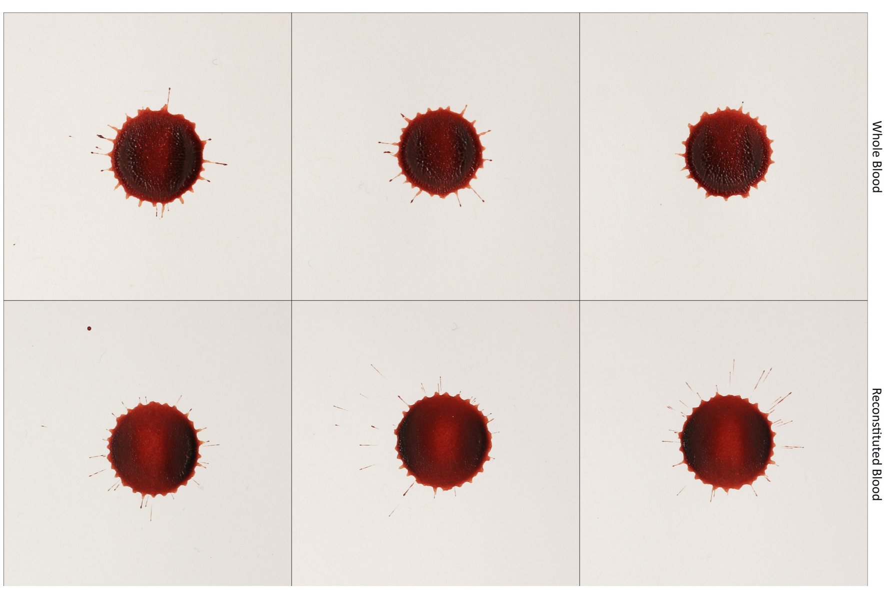


*Figure 3. Comparison of whole blood vs. reconstituted blood. Spatter of single drop from 50 cm onto 110# cardstock. Each image is 5x5 cm.*

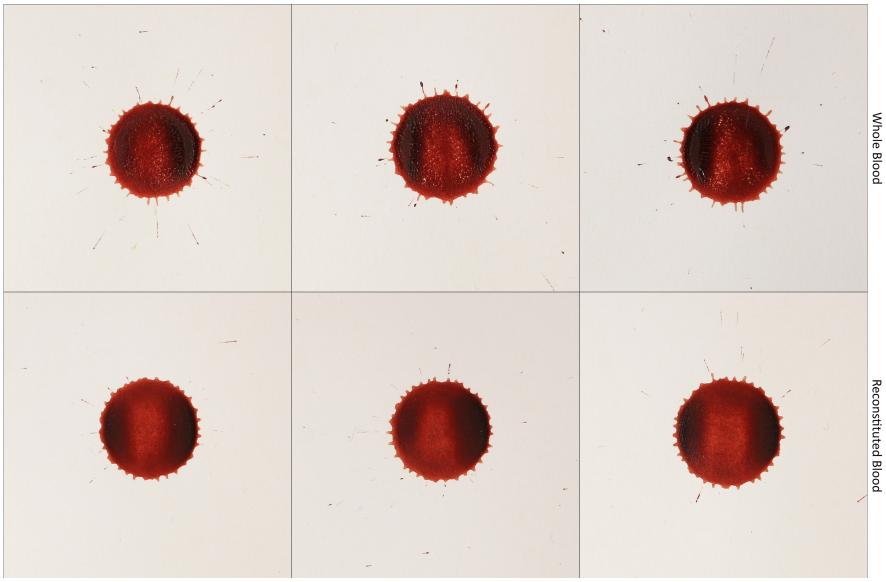


*Figure 4. Comparison of whole blood vs. reconstituted blood: spatter of single drop from 100 cm onto 110# cardstock. Each image is 5x5 cm.*

#### 2.4.2 Casework Samples

A total of 90 samples were collected from operational casework, 69 of which were selected for use in the study. The operational samples utilized in this study included closed cases under the jurisdiction of the Kansas City Police Department. Casework samples were reviewed and redacted as necessary to ensure anonymity. Casework images were acquired at crime scenes using a variety of cameras and lenses (Canon EOS 30D, 40D, 50D, 70D; Canon Digital Rebel; Canon PowerShot G12, G15; Nikon E5000, E5700, E950, E990; Fujifilm FinePix S1 Pro). Many of the casework samples were photographed by the same BPA analyst who photographed the controlled collection samples.

The casework samples were included to assure that the test samples are representative of actual casework, to add an additional level of realism, and so that the participants would not be able to make limiting assumptions regarding the types of mechanisms that may be included in the study. For example, if the study were limited to controlled collection, participants would be able to "game the test" by assuming that no one was killed or injured in the creation of the patterns. The short text responses showed a variety of instances in which participants assumed that controlled collection images came from casework — for example, a short text response to a controlled pattern (forward spatter from a gunshot) stated "Spatter impact pattern. Probably a forward impact. Stains of small size with different orientations and directionality. Some pieces of projected flesh, too."

#### 2.4.3 Sample Selection

From the overall dataset, 192 samples were selected for use in this study (see Supplemental Data S2). Samples were selected and assigned so that all participants had a broad range of bloodstain pattern types: 108 samples were assigned to every participant (55 Casework and 53 Controlled), and 84 samples were each assigned to half of participants (14 Casework and 70 Controlled). Casework samples were assigned to only half of the participants if the images were of the same area of the same crime scene. Controlled samples were assigned to half of the participants if the mechanism and content of the samples was approximately the same. The order of assignments was randomized among participants.

#### *2.4.4 Image Preparation*

The majority (65%) of the bloodstain patterns were represented by multiple images (range 1-9, mean 2.8). Multiple images were included to provide different perspectives and levels of detail (as would be done with casework). All controlled collection samples and most casework samples contained at least one scale, US customary and/or metric (22 casework samples did not include a scale). A portion of the samples (23%) contained an image with labels for specific area(s) of the bloodstain, particularly for complex bloodstain patterns. Some classifications and questions referred explicitly to these labels. For short text prompts, labels were there as a convenience to provide locations to optionally refer to in responses.

Original image dimensions for the controlled collection images were all 5472×3648 pixels. Original image dimensions for the casework samples varied (median 2736×3648 pixels; range of width and height 505-5472 pixels; 9 of the original images had one or both dimensions under 1000 pixels). The images distributed to participants were often cropped to eliminate extraneous areas or potentially biasing content. For consistent display in the software, images required a minimum width of 2500 pixels, and smaller images were rescaled. All image metadata (EXIF information) was removed, and captions were programmatically added with sample and image numbers. Four comments in the post-study survey referred to poor image quality, but note that the (relatively few) low-quality images were from actual casework.

#### *2.4.5 Sample Assignments*

Each participant was assigned 150 samples out of the pool of 192 samples: 108 samples were assigned to all participants, and 84 were assigned to half of the participants. Samples were assigned to half of participants for these reasons:

- Some casework samples were collected from the same crime scenes. To ensure against the possibility that participants could glean information about other samples, samples that shared crime scenes were each assigned to half of the participants.
- Some controlled samples that were collected in very similar manners were each assigned to half of the participants. This limited the number of redundant samples each participant received, while providing information on as many samples as possible.

#### *2.4.6 Examples*

Figure 5 depicts examples of bloodstain patterns from the study that resulted in substantive disagreements among analysts. Note that each sample is shown in multiple images. Sample #240 was caused by *expiration*, created by a team member coughing blood. One of the prompts that participants were provided for this sample was *impact* ("a bloodstain pattern resulting from an object striking liquid blood" [20], such as from a gunshot or club): 16 participants responded *definitive*, 23 *included*, and 4 *excluded* (43 total). Because this sample was known to be caused by *expiration*, the *impact* prompt is false, and those 16 *definitive* responses can be assessed as errors (i.e., contradicting a known cause).

The second and third samples in Figure 5 are from casework, showing examples of prompts that do not have known causes, and therefore are assessed in terms of reproducibility, not error or correctness. For sample #188, participants were provided *impact* as a prompt: 15 responded *definitive*, 11 *included*, and 17 *excluded* (43 total). For Sample #975, in response to the question "Did Pattern A occur after Pattern B?" 8 participants responded *no*, 10 *possible*, and 25 *yes* (43 total).

Note: See Supplemental Data S2 for low-resolution summary images and responses for all samples; see [25] for full-resolution imagery for all samples. Supplemental Data S3 contains all prompt and response data.


*Figure 5. Examples of bloodstain patterns used in the study. Sample #240 (controlled collection): expiration on cardboard, created by a team member coughing blood. Sample #188 (casework): drip pattern in basement; victim shot and killed in kitchen; blood flow through floor, down a rafter and dripped onto the basement floor. Sample #975 (casework): Three homicide victims within this room; pattern shows characteristics of spatter stains altered by a swipe resulting in multiple perimeter stains. (Descriptions were not provided to participants)*

## 2.5 Assessing Known Cause for Classification Prompts and Questions

For a specific classification prompt or question, a definitively correct ("known cause") response was generally available for controlled collection samples, but was not asserted for prompts associated with casework samples (Table 5). We used a high threshold in determining known cause, and if the cause was considered at all debatable or semantic, we marked it as "unknown cause."

| | *Classifications* | | | | | | *Questions* | | | | | | *All prompts* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Casework* | | *Controlled* | | *All samples* | | *Casework* | | *Controlled* | | *All samples* | | *Casework* | | *Controlled* | | *All samples* | |
| Known Cause | 0 | 0% | 392 | 82% | 392 | 48% | 0 | 0% | 96 | 77% | 96 | 43% | 0 | 0% | 488 | 81% | 488 | 47% |
| Unknown Cause | 339 | 100% | 84 | 18% | 423 | 52% | 98 | 100% | 29 | 23% | 127 | 57% | 437 | 100% | 113 | 19% | 550 | 53% |
| Total | 339 | | 476 | | 815 | | 98 | | 125 | | 223 | | 437 | | 601 | | 1,038 | |

*Table 5. Proportion of classifications and questions for which known cause was asserted.*

The causal mechanism for a given bloodstain sample was known definitively for the controlled collection bloodstains (and usually videotaped). An unexpected finding was that semantic issues sometimes limited our ability to define known cause for a specific classification or question, even when video of the event was available. For example:

- The definitions of "cessation" and "cast-off" do not indicate whether cessation is a subcategory of cast-off, or if they are disjoint.
- "Pool" and "saturation stain" refer to "accumulation of blood" without any indication of what quantity is necessary to be considered an accumulation, resulting in inconsistent usage (some BPA analysts consider a single drop on an absorbent surface a saturation stain, but most do not).

We also did not claim known cause for classifications based solely on observations (such as "air bubble" or "perimeter stain"). We did not assert known cause for casework samples because we cannot claim absolute knowledge of the cause of each bloodstain.

For the 192 samples, the classification prompts were associated with known cause as follows:

- 21 samples had at least one False and no True classification prompts (nine of which also had classification prompts with unknown causes);
- Six samples had at least one True and no False (all of which also had prompts with unknown causes);
- 94 samples had at least one True and at least one False;
- 70 samples had no True and no False (all unknown cause);
- One sample had no classification prompts.

Each sample had 0-6 classification prompts (mean 4.2), for a total of 815 classification prompts. Table 7 shows the classification prompts used.

## 2.6 Assessing Operational Consequence for Classification Prompts and Questions

Different classifications and questions vary as to the operational consequences of errors and disagreements thereon. Some disagreements are semantic or trivial (such as if analysts disagree whether a single drop of blood constitutes spatter) and would not be likely to have any meaningful impact on actual casework. However, other disagreements (such as whether a given pattern was caused by expiration (e.g., coughing) or by forward spatter from a gunshot) would be likely to have very serious consequences in casework or in court. (Note that any errors made by an expert could negatively impact the expert's credibility, however this broader consequence was not considered when assessing operational consequence.)

To enable our analyses to account for these differences, the BPA analysts on the study team reviewed all of the classification prompts and questions and flagged classification prompts and questions in response to this question:

> *If this sample were in actual casework and a BPA analyst was asked this classification or question on the stand and made an error, would it be highly consequential to the case? (or if two BPA experts were given this prompt and gave opposite conclusions in testimony)*

Classification prompts and questions that were assessed by both of the BPA analysts on the study team as likely to have a high impact on casework under this scenario were labelled as "most consequential." Of the classification prompts, 24.2% (197 of 815) were assessed as most consequential (see Table 6); of the questions, 16.6% (37 of 223) were assessed as most consequential; of all prompts, 22.5% were most consequential (234 of 1,038). Note that the prompts assessed as most consequential were generally associated with the type of classification (Table 7).

| | *All* | | | | *Most consequential* | | | |
|---|---|---|---|---|---|---|---|---|
| | *False* | *Unknown* | *True* | *Total* | *False* | *Unknown* | *True* | *Total* |
| *Classification prompts* | 235 | 423 | 157 | 815 | 62 | 80 | 55 | 197 |
| *Questions* | 62 | 127 | 34 | 223 | 8 | 21 | 8 | 37 |
| *Total prompts* | 297 | 550 | 191 | 1,038 | 70 | 101 | 63 | 234 |

*Table 6. Counts of classification and question prompts, by known cause and operational consequence.*

Examples of questions that were assessed as most consequential:

- Were these bloodstains created by a gunshot?
- Is the [spatter] on the gun caused by [backspatter]?
- Was the bloodstain caused by blood released from an object moving upward and from left to right?

| *Classification prompt* | *Uses* | *MostConseq* |
|---|---|---|
| Altered | 38 | 0 |
| Backspatter | 36 | 0 |
| Blood clot | 20 | 0 |
| Bubble ring | 8 | 0 |
| Cast-off | 28 | 0 |
| Cessation | 16 | 0 |
| Drip pattern | 36 | 0 |
| Drip stain | 35 | 0 |
| Drip trail | 16 | 0 |
| Expiration | 45 | 45 |
| Flow | 11 | 0 |
| Forward spatter | 29 | 0 |
| Impact | 69 | 64 |
| Insect stain | 16 | 0 |
| Perimeter stain | 8 | 0 |
| Pool | 23 | 0 |
| Projected | 71 | 0 |
| Satellite stain | 31 | 12 |
| Saturation stain | 26 | 0 |
| Serum stain | 13 | 0 |
| Spatter | 64 | 51 |
| Splash | 37 | 0 |
| Swipe | 30 | 0 |
| Transfer | 57 | 25 |
| Void | 15 | 0 |
| Wipe | 37 | 0 |

*Table 7. Classification prompts, with number of uses (total 815) and number that were assessed as "most consequential" (total 197; See 1AAppendix D.1 for ASB definitions.*

## 2.7 Evaluating Short Text Responses

Short text responses collected as part of this study served two primary purposes—to generate questions and to facilitate analyses specific to these open-ended responses. The questions that were added to the study based upon these responses were used in all analyses and allowed for a large-scale broad evaluation of statements that extend beyond classifications. And, in contrast with explicit prompts, short text responses provided a means to assess accuracy in a manner comparable to how pattern classifications are reported in operational casework. Although this assessment was necessarily smaller in scale and required more interpretation, it was important to determine whether performance was notably different as a function of the reporting mechanism, or whether accuracy remained stable irrespective of whether decisions were provided through prompted items or open-ended summaries. Short text responses were evaluated in two phases: classifications were extracted and associated with specific known cause(s) when possible[*], and responses were assessed holistically for quality and thoroughness.

### *2.7.1 Extracting Classification Decisions from Short Text Responses*

In order to evaluate the short text responses with respect to known cause, we developed a multi-phasic method of review to extract the pattern classifications of interest from the text and assign a decision corresponding to the classification prompt responses (Section 3.9). First, the short text

[*] *Note that short text responses were also associated with prompted questions and assigned responses, when possible. However, this required notable subjective interpretation by the research team, so the associations of short text responses with specific questions were not used for analyses.*

responses were searched algorithmically to locate any uses of the 26 classification terms and their variants. Usually, classifications were explicit in the short text responses. However, given the highly variable descriptions of observations and classifications in these responses, additional words and phrases were included in this automated search, as described in Table 8. Subsequently, a manual review was also conducted to identify and remove any inappropriately flagged classification terms (e.g., the term *impact* used as a verb, rather than a pattern classification) as well as classifications not flagged due to misspelling. Each classification was then assigned a decision of either *definitive, included,* or *excluded* based upon the criteria provided to the participants (outlined in Section 2.3.1) as well as a qualifying language rubric developed for this purpose (Table 8)). When assigning decisions to each classification, the context of the short text response(s) was reviewed in its entirety by multiple members of the research team, which included subject matter experts. A final decision was assigned based upon consensus within the team and a final verification review conducted by the subject matter experts. Overall, the short text responses yielded 4,648 extracted classification decisions, of which 1,976 were on casework samples and 2,672 were on controlled samples.

| *If written response contained:* | *Response was coded for:* |
|---|---|
| ASB definition (verbatim, or nearly verbatim) | Corresponding ASB classification |
| Airborne drops/droplets | Spatter |
| Impact spatter | Impact and spatter |
| Secondary stains | Satellite |
| Beat, beating, beaten | Impact |
| Strike, striking, struck blood or blood source | Impact |
| Firearm, gun, gunshot | Forward spatter, back spatter, spatter, and impact |

*Table 8. Rubric for coding of short text responses when classifications were not explicitly mentioned.*

As an example of this process, consider the following short text response provided for Sample 632:

> *Blood has **pooled** on the surface of the carpet but not yet soaked into the carpet to **saturate** it. One component of the stained area is either a **transfer** stain or an **altered** stain.*

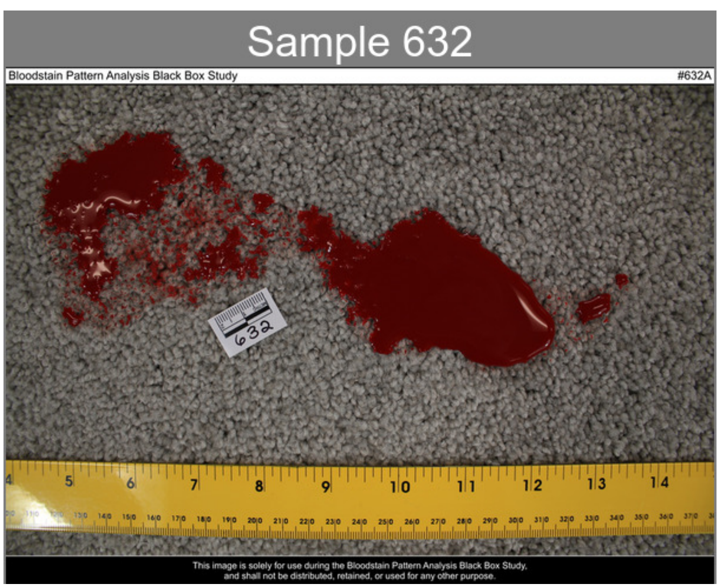


*Figure 6. Sample #632. Controlled. Pour 50ml of blood on carpet. (Description was not provided to participants)*

The bolded terms indicate the extracted pattern classifications (or derivatives) for this response: *pool*, *saturation*, *transfer*, *altered*. The underlined words indicate the qualifying language considered for assigning decisions to each classification. Therefore, for this response *pool* was assigned a response of *definitive*, *saturation* was assigned a response of *excluded*, and *transfer* and *altered* were both assigned a response of *included*.

Each extracted classification was then assigned one of three types: physical characteristic observation, super-classification, or classification (see Table 9). Physical characteristic observations (Obs) are those terms that can be used to directly describe a bloodstain, but not a causal mechanism. Super-classifications (Super) and classifications (Class) are those terms that can be used to describe the causal mechanisms of bloodstains—these cannot be directly observed, but rather are inferred based upon observations. Super-classifications represent the larger mechanism that encompasses a number of classification mechanisms as sub-categories (e.g., *Altered* is a super-classification that includes the classification *Wipe* as a sub-category). Classifications are more specific causal mechanisms (e.g., *Expiration* indicates that the *Spatter* super-classification was caused by blood forced by airflow). Classifications can be further broken down into Spatter (Class:S) and Non-Spatter (Class:NS) groups [2], [3]. These type categorizations are intended to represent the hierarchical nature of bloodstain pattern classifications [2], [3] and allow for an assessment of responses as a function of these types.

| *Extracted Classification* | *Type* | *Extracted Classification* | *Type* |
|---|---|---|---|
| Altered | Super | Insect Stain | Class:NS |
| Backspatter | Class:S | Perimeter | Obs |
| Blood Clot | Obs | Pool | Obs |
| Bubble Ring | Obs | Projected | Class:S |
| Castoff | Class:S | Satellite | Class:S |
| Cessation | Class:S | Saturation | Obs |
| Drip Pattern | Class:NS | Serum | Obs |
| Drip Stain | Class:NS | Spatter | Super |
| Drip Trail | Class:NS | Splash | Class:NS |
| Expiration | Class:S | Swipe | Class:NS |
| Flow | Class:NS | Transfer | Super |
| Forward Spatter | Class:S | Void | Class:NS |
| Impact | Class:S | Wipe | Class:NS |

*Table 9. Assigned types for each classification term used is this study.*

Following extraction and response and type assignment, the set of classifications for each short text response was then evaluated with respect to the known cause(s), when possible. Samples from casework were marked as having unknown cause. Extracted classifications that corresponded exactly to classification prompts were assigned that prompt's known cause (if any). Other extracted classifications on controlled collection samples were reviewed by the team and assigned known cause, if it could be determined unambiguously. The resulting coded classifications, responses, and known causes were used in analyses of the short text responses.

### *2.7.2 Short Text Response Categorization*

To further characterize the text responses, the BPA experts on the team evaluated the quality and thoroughness of each, while also assessing whether the analysts' observations and conclusions were adequately supported. Each response was assessed with respect to the categories listed in Table 10.

| *IssueType* | | *Definition* |
|---|---|---|
| Error | Inaccurate reconstruction | Reconstruction statements* that were assessed as inaccurate, or mostly inaccurate |
| | Inaccurate observations | Participant described something that was clearly not present or inaccurate. For example, incorrect measurements of stains, or observing clots that were not present. |
| | Widely unsupported | Participant provided observations and/or classifications that were clearly not supported by the imagery presented. For example, one participant suggested a specific object was used to create depicted spatter, even though only images of the spatter stains were provided with no additional context images or information. |
| | Observations do not support pattern classification | Participant's descriptive observations did not support the stated classification. For example, one participant described a splash pattern, but then classified the bloodstain as a wipe pattern. |
| Clerical Error | Clerical error | Clerical error. Participant's observations suggest incorrect image(s) was characterized. For example, one participant described a bloodstain “on the chair and rug”, however, the presented bloodstain imagery clearly depicted no such objects. |
| Not Thorough | Not thorough or no observations | Key observations were missed that may have influenced classification, or participant stated a classification but offered no, or minimal, descriptive observations. Both components were required in this study. |
| None | Accurate Reconstruction | Reconstruction statements‡ that were assessed as accurate, or mostly accurate. |
| | No issues noted | |

*Table 10. Categories used to evaluate short text responses.*

As an example of a “widely unsupported” response, that designation was given to the following response to Sample 763 (Figure 7), which was created as a drip pattern: “*The smaller stain pattern is consistent with backspatter from a gunshot entry with the drip pattern resulting from drainage from the target entry site*.”


*Figure 7. Sample #763. Controlled. Drip pattern, 1.3 meters (48") Carpet; 50 drops. (Description was not provided to participants)*

Of these categories (entries in the IssueType column in Table 10), four were classified as “Procedural Errors”: *widely unsupported*, *error in reconstruction*, *inaccurate observations*, and *observations do not support pattern classification*. The category of “Procedural Error” was then used in the analyses.

## 2.8 Test Yield

Table 1 above summarizes the test yield in terms of participants; **Error! Reference source not found.** Table 11 and Table 12 provide prompt and response counts, and **Error! Reference source**

* *Statements were considered reconstruction statements if they assessed details of the cause of a bloodstain that went beyond bloodstain pattern classifications, such as order of events or directionality.*

**not found.** summarize the yield in terms of responses by prompt type, known cause, and operational consequence. In addition to the classification and question responses, the study yielded 1,760 short text responses (each response containing one observation text response and one mechanism text response), with a median of 9 text responses per sample (mean 9.2; minimum 5, maximum 13).

| | | | |
|---|---|---|---|
| *Prompts* | *Total* | 822 | 234 |
| | *Omitted (prompts with fewer than 5 responses)* | -7 | -11 |
| | *Used for analyses* | 815 | 223 |
| *Responses* | *Total* | 27,063 | 5,995 |
| | *Omitted (prompts with fewer than 5 responses)* | -25 | -28 |
| | *Used for analyses* | 27,038 | 5,967 |
| *Responses per prompt* | *Mean* | 33.2 | 26.8 |
| | *Median* | 42 | 28 |
| | *Range* | 5-49 | 5-48 |
| *Prompts per sample* | *Mean* | 4.2 | 1.2 |
| | *Median* | 4 | 1 |
| | *Range* | 0-6 | 0-5 |

Table 11. Counts of classification and question prompts and responses.

| | *All* | | | | *Most consequential* | | | |
|---|---|---|---|---|---|---|---|---|
| | *False* | *Unknown* | *True* | *Total* | *False* | *Unknown* | *True* | *Total* |
| *Classification prompts* | 235 | 423 | 157 | 815 | 62 | 80 | 55 | 197 |
| *Questions* | 62 | 127 | 34 | 223 | 8 | 21 | 8 | 37 |
| *Total prompts* | 297 | 550 | 191 | 1,038 | 70 | 101 | 63 | 234 |

Table 12. Counts of classification and question prompts, by known cause and operational consequence.

# 3 Results

Note: Supplemental Data S3 includes details for all response data: classification prompts and questions with responses, short text responses, de-identified survey responses, and summary results by participant.

## 3.1 Accuracy

Figure 8 shows the distribution of responses for classifications and questions. For example, of the classification prompts for which the prompt represented the known cause (*true*), 52.8% of responses were *definitive* (agreeing with known cause), but 15.2% of responses were *excluded* (erroneous exclusions, contradicting known cause). Overall, 11,634 classification prompts for which there was a known cause, 11.2% of responses contradicted the known cause and therefore were erroneous (weighted average of erroneous *definitive* and erroneous *excluded* responses). For the 2,163 questions for which there was a known cause, 11.0% of responses were erroneous (weighted average of erroneous *yes* and erroneous *no* responses). These results are similar to [16], which reported that 13% of classifications were erroneous. Responses were indeterminate for 30.1% of classifications and 43.8% of questions. If we consider only determinate responses (*definitive*, *excluded*, *yes*, *no*) on prompts with known cause, 83.0% of responses were correct with respect to known cause ("overall predictive value").

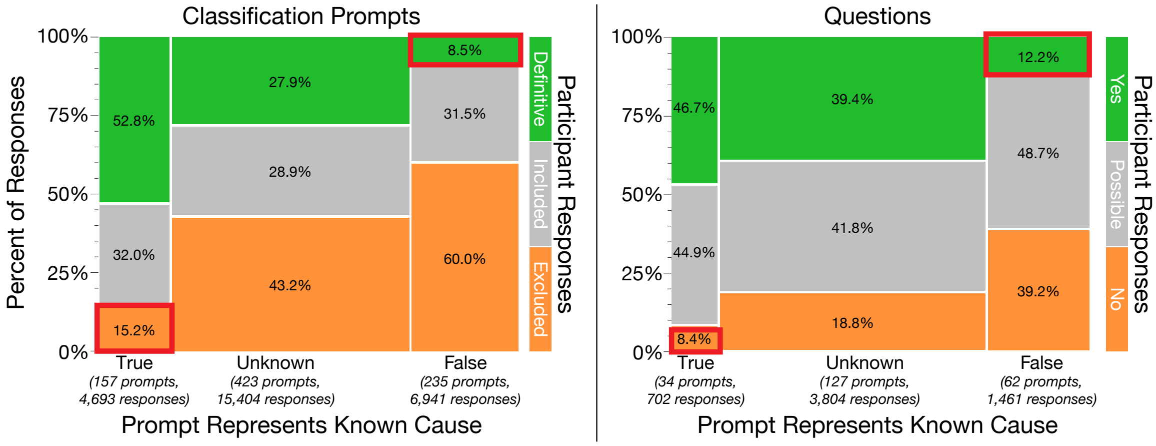


*Figure 8. Mosaic plots of responses. The columns indicate whether the prompts represent known cause of the bloodstain: prompts that are consistent with known cause are labelled* true*; prompts contrary to known cause are labelled* false*; prompts for which the cause is unknown are labelled* unknown*, which we do not assess in terms of accuracy or error. Responses are color-coded, with proportions shown in the y-axis. Erroneous responses (i.e., contradicting known causes) are outlined in red.*

In order to limit the effects of prompts that could be seen as minor or semantic, the BPA analysts on the study team evaluated each prompt to determine whether an error on that prompt would be highly consequential in an actual case. These are labeled "most consequential", and 22.5% of the prompts were labeled as such. If limited to the most consequential prompts, the error rate for classifications was 9.0% and the error rate for questions was 5.8%; the overall predictive value was 86.6% (see Table 15).

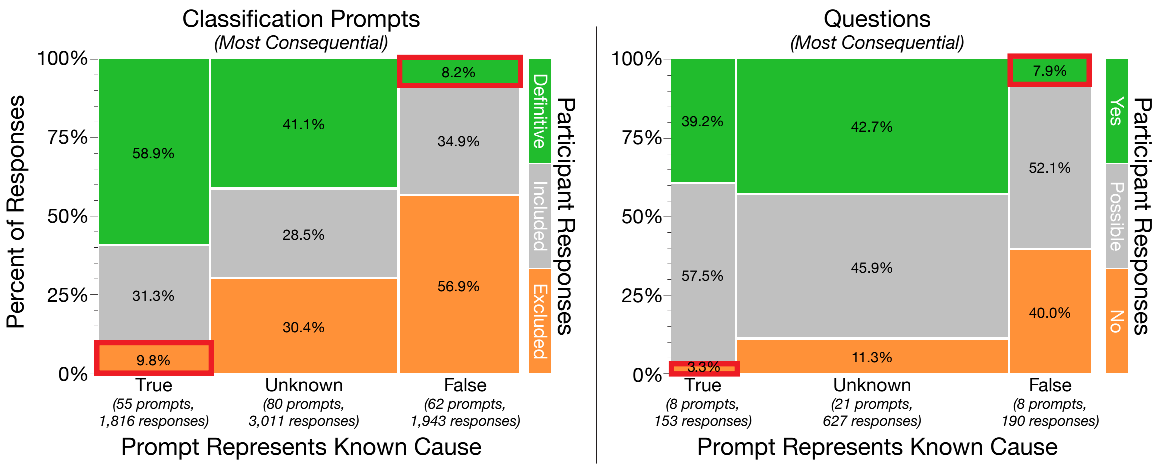


*Figure 9. Mosaic plots of responses for prompts assessed as "most consequential". Compare to Figure 8, which shows all results.*

Table 13 shows details for the data shown in Figure 8 and Figure 9.

| | | *All* | | | | | | | | *Most consequential* | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *False* | | *Unknown* | | *True* | | *Total* | | *False* | | *Unknown* | | *True* | | *Total* | |
| *Classificati* | *Definitive* | 592 | 8.5% | 4,302 | 27.9% | 2,478 | 52.8% | 7,372 | 27.3% | 160 | 8.2% | 1,238 | 41.1% | 1,070 | 58.9% | 2,468 | 36.5% |
| | *Included* | 2,183 | 31.5% | 4,450 | 28.9% | 1,504 | 32.0% | 8,137 | 30.1% | 678 | 34.9% | 857 | 28.5% | 568 | 31.3% | 2,103 | 31.1% |
| | *Excluded* | 4,166 | 60.0% | 6,652 | 43.2% | 711 | 15.2% | 11,529 | 42.6% | 1,105 | 56.9% | 916 | 30.4% | 178 | 9.8% | 2,199 | 32.5% |
| | *Total responses* | 6,941 | | 15,404 | | 4,693 | | 27,038 | | 1,943 | | 3,011 | | 1,816 | | 6,770 | |
| *Questions* | *Yes* | 178 | 12.2% | 1,499 | 39.4% | 328 | 46.7% | 2,005 | 33.6% | 15 | 7.9% | 268 | 42.7% | 60 | 39.2% | 343 | 35.4% |
| | *Possible* | 711 | 48.7% | 1,589 | 41.8% | 315 | 44.9% | 2,615 | 43.8% | 99 | 52.1% | 288 | 45.9% | 88 | 57.5% | 475 | 49.0% |
| | *No* | 572 | 39.2% | 716 | 18.8% | 59 | 8.4% | 1,347 | 22.6% | 76 | 40.0% | 71 | 11.3% | 5 | 3.3% | 152 | 15.7% |
| | *Total responses* | 1,461 | | 3,804 | | 702 | | 5,967 | | 190 | | 627 | | 153 | | 970 | |
| *All prompts* | *Positive* | 770 | 9.2% | 5,801 | 30.2% | 2,806 | 52.0% | 9,377 | 28.4% | 175 | 8.2% | 1,506 | 41.4% | 1,130 | 57.4% | 2,811 | 36.3% |
| | *Indeterminate* | 2,894 | 34.4% | 6,039 | 31.4% | 1,819 | 33.7% | 10,752 | 32.6% | 777 | 36.4% | 1,145 | 31.5% | 656 | 33.3% | 2,578 | 33.3% |
| | *Negative* | 4,738 | 56.4% | 7,368 | 38.4% | 770 | 14.3% | 12,876 | 39.0% | 1,181 | 55.4% | 987 | 27.1% | 183 | 9.3% | 2,351 | 30.4% |
| | *Total responses* | 8,402 | | 19,208 | | 5,395 | | 33,005 | | 2,133 | | 3,638 | | 1,969 | | 7,740 | |

*Table 13. Counts of responses by known cause and operational consequence. Errors are highlighted.*

Accuracy can be evaluated using a variety of metrics, shown for all prompts in Table 14 and for the most consequential prompts in Table 15. Definitions for each accuracy metric as implemented here are shown in Table 16. Although measures of accuracy, error rates, and predictive values are longstanding and widely incorporated in many fields, they are most commonly applied to binary classification/decision tasks with explicit "positive" and "negative" states. Their usage becomes more ambiguous for a decision task including more than two levels, as is the case for this study — in which there are three levels of responses (indeterminate as well as positive and negative) and three levels of known values (unknown as well as true and false). In the forensic literature, there is disagreement in how to handle indeterminate responses and whether they should be included in the denominator for the computation of accuracy and error rate [11], [27], [28]. For completeness, we have provided both computations in Table 14 and Table 15; subscripts of "all" indicate that all responses on prompts with defined known cause are included in the denominator, while subscripts of "det" indicate that only determinate responses on prompts with defined known cause are included in the denominator.

| | *All responses* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Classifications* | | | *Questions* | | | *Overall* | | |
| | *Rates* | *C.I.* | *Counts* | *Rates* | *C.I.* | *Counts* | *Rates* | *C.I.* | *Counts* |
| OPV | 83.6% | 82.8%-84.4% | (6644/7947) | 79.2% | 76.7%-81.5% | (900/1137) | 83.0% | 82.3%-83.8% | (7544/9084) |
| PPV | 80.7% | 79.3%-82.1% | (2478/3070) | 64.8% | 60.5%-69.0% | (328/506) | 78.5% | 77.1%-79.8% | (2806/3576) |
| NPV | 85.4% | 84.4%-86.4% | (4166/4877) | 90.6% | 88.1%-92.8% | (572/631) | 86.0% | 85.1%-86.9% | (4738/5508) |
| $TPR_{all}$ | 52.8% | 51.4%-54.2% | (2478/4693) | 46.7% | 43.0%-50.5% | (328/702) | 52.0% | 50.7%-53.4% | (2806/5395) |
| $TPR_{det}$ | 77.7% | 76.2%-79.1% | (2478/3189) | 84.8% | 80.8%-88.2% | (328/387) | 78.5% | 77.1%-79.8% | (2806/3576) |
| $TNR_{all}$ | 60.0% | 58.9%-61.2% | (4166/6941) | 39.2% | 36.6%-41.7% | (572/1461) | 56.4% | 55.3%-57.5% | (4738/8402) |
| $TNR_{det}$ | 87.6% | 86.6%-88.5% | (4166/4758) | 76.3% | 73.1%-79.3% | (572/750) | 86.0% | 85.1%-86.9% | (4738/5508) |
| $FPR_{all}$ | 8.5% | 7.9%-9.2% | (592/6941) | 12.2% | 10.5%-14.0% | (178/1461) | 9.2% | 8.6%-9.8% | (770/8402) |
| $FPR_{det}$ | 12.4% | 11.5%-13.4% | (592/4758) | 23.7% | 20.7%-26.9% | (178/750) | 14.0% | 13.1%-14.9% | (770/5508) |
| $FNR_{all}$ | 15.2% | 14.1%-16.2% | (711/4693) | 8.4% | 6.5%-10.7% | (59/702) | 14.3% | 13.3%-15.2% | (770/5395) |
| $FNR_{det}$ | 22.3% | 20.9%-23.8% | (711/3189) | 15.2% | 11.8%-19.2% | (59/387) | 21.5% | 20.2%-22.9% | (770/3576) |
| $ER_{all}$ | 11.2% | 10.6%-11.8% | (1303/11634) | 11.0% | 9.7%-12.3% | (237/2163) | 11.2% | 10.6%-11.7% | (1540/13797) |
| $ER_{det}$ | 16.4% | 15.6%-17.2% | (1303/7947) | 20.8% | 18.5%-23.3% | (237/1137) | 17.0% | 16.2%-17.7% | (1540/9084) |
| DR | 69.9% | 69.4%-70.5% | (18901/27038) | 56.2% | 54.9%-57.4% | (3352/5967) | 67.4% | 66.9%-67.9% | (22253/33005) |
| IR | 30.1% | 29.5%-30.6% | (8137/27038) | 43.8% | 42.6%-45.1% | (2615/5967) | 32.6% | 32.1%-33.1% | (10752/33005) |

*Table 14. Accuracy rates. See Table 16 for definitions. All rates are calculated from values in Table 13. Clopper–Pearson 95% confidence intervals are shown* [29]*. Counts for each rate are included to avoid ambiguity.*

| | *Most consequential* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Classifications* | | | *Questions* | | | *Overall* | | |
| | *Rates* | *C.I.* | *Counts* | *Rates* | *C.I.* | *Counts* | *Rates* | *C.I.* | *Counts* |
| OPV | 86.5% | 85.2%-87.9% | (2175/2513) | 87.2% | 80.9%-92.0% | (136/156) | 86.6% | 85.2%-87.9% | (2311/2669) |
| PPV | 87.0% | 85.0%-88.8% | (1070/1230) | 80.0% | 69.2%-88.4% | (60/75) | 86.6% | 84.6%-88.4% | (1130/1305) |
| NPV | 86.1% | 84.1%-88.0% | (1105/1283) | 93.8% | 86.2%-98.0% | (76/81) | 86.6% | 84.7%-88.3% | (1181/1364) |
| $TPR_{all}$ | 58.9% | 56.6%-61.2% | (1070/1816) | 39.2% | 31.4%-47.4% | (60/153) | 57.4% | 55.2%-59.6% | (1130/1969) |
| $TPR_{det}$ | 85.7% | 83.7%-87.6% | (1070/1248) | 92.3% | 83.0%-97.5% | (60/65) | 86.1% | 84.1%-87.9% | (1130/1313) |
| $TNR_{all}$ | 56.9% | 54.6%-59.1% | (1105/1943) | 40.0% | 33.0%-47.3% | (76/190) | 55.4% | 53.2%-57.5% | (1181/2133) |
| $TNR_{det}$ | 87.4% | 85.4%-89.1% | (1105/1265) | 83.5% | 74.3%-90.5% | (76/91) | 87.1% | 85.2%-88.8% | (1181/1356) |
| $FPR_{all}$ | 8.2% | 7.1%-9.5% | (160/1943) | 7.9% | 4.5%-12.7% | (15/190) | 8.2% | 7.1%-9.5% | (175/2133) |
| $FPR_{det}$ | 12.6% | 10.9%-14.6% | (160/1265) | 16.5% | 9.5%-25.7% | (15/91) | 12.9% | 11.2%-14.8% | (175/1356) |
| $FNR_{all}$ | 9.8% | 8.5%-11.3% | (178/1816) | 3.3% | 1.1%-7.5% | (5/153) | 9.3% | 8.0%-10.7% | (183/1969) |
| $FNR_{det}$ | 14.3% | 12.4%-16.3% | (178/1248) | 7.7% | 2.5%-17.0% | (5/65) | 13.9% | 12.1%-15.9% | (183/1313) |
| $ER_{all}$ | 9.0% | 8.1%-10.0% | (338/3759) | 5.8% | 3.6%-8.9% | (20/343) | 8.7% | 7.9%-9.6% | (358/4102) |
| $ER_{det}$ | 13.5% | 12.1%-14.8% | (338/2513) | 12.8% | 8.0%-19.1% | (20/156) | 13.4% | 12.1%-14.8% | (358/2669) |
| DR | 68.9% | 67.8%-70.0% | (4667/6770) | 51.0% | 47.8%-54.2% | (495/970) | 66.7% | 65.6%-67.7% | (5162/7740) |
| IR | 31.1% | 30.0%-32.2% | (2103/6770) | 49.0% | 45.8%-52.2% | (475/970) | 33.3% | 32.3%-34.4% | (2578/7740) |

*Table 15. Accuracy rates for classification prompts and questions assessed as most consequential. See Table 16 for definitions. All rates are calculated from values in Table 13. Clopper–Pearson 95% confidence intervals are shown. Counts for each rate are included to avoid ambiguity.*

| | | |
|---|---|---|
| Overall predictive value | OPV | Percent of responses that are correct (omitting unknown cause and indeterminate responses) |
| Positive predictive value | PPV | Percent of positive responses (yes or definitive) that are correct (omitting indeterminate known cause) |
| Negative predictive value | NPV | Percent of negative responses (no or excluded) that are correct (omitting indeterminate known cause) |
| True positive rate (sensitivity) | | Percent of actual positives (true known cause) that result in positive responses (yes or definitive) |
| | $TPR_{all}$ | (including all responses in the denominator) |
| | $TPR_{det}$ | (omitting indeterminate responses (possible or included) from the denominator) |
| True negative rate (specificity) | | Percent of actual negatives (false known cause) that result in negative responses (no or excluded) |
| | $TNR_{all}$ | (including all responses in the denominator) |
| | $TNR_{det}$ | (omitting indeterminate responses (possible or included) from the denominator) |
| False positive rate | | Percent of actual negatives (false known cause) that result in positive responses (yes or definitive) |
| | $FPR_{all}$ | (including all responses in the denominator) |
| | $FPR_{det}$ | (omitting indeterminate responses (possible or included) from the denominator) |
| False negative rate | | Percent of actual positives (true known cause) that result in negative responses (no or excluded) |
| | $FNR_{all}$ | (including all responses in the denominator) |
| | $FNR_{det}$ | (omitting indeterminate responses (possible or included) from the denominator) |
| Error rate | | Percent of responses that contradict known cause |
| | $ER_{all}$ | (including all responses in the denominator) |
| | $ER_{det}$ | (omitting indeterminate responses (possible or included) from the denominator) |
| Determinate rate | DR | Percent of responses that are not indeterminate (definitive, excluded, yes, or no) (1-IR) |
| Indeterminate rate | IR | Percent of responses that are indeterminate (included or possible) (1-DR) |

*Table 16. Definitions of rates reported in Table 14 and Table 15.*

As an alternative to explicit prompts, short text responses provided a means to assess accuracy in a manner comparable to how pattern classifications are reported in operational casework: these results are discussed in Section 3.9.

## 3.2 Consensus

Between five and 49 participants responded to each prompt (mean 33.2 participants per classification prompt; 26.8 per question). The responses for each prompt can be seen as votes in a decision space, as shown in Figure 10, which plots each prompt in terms of the proportion of each response type. For example, a classification prompt of *spatter* (with a true known cause) with 46 responses (22 *definitive*, 12 *included*, 12 *excluded*) is plotted as a blue point in the classifications chart at (48%, 26%). This provides a means to evaluate results in terms of consensus, which serves multiple purposes: it provides an understanding of the collective behavior of analysts, it serves as a proxy for the known cause when the cause for a bloodstain is unknown, and it indicates the collective judgement of analysts regarding whether the samples contain sufficient information to make a given decision.

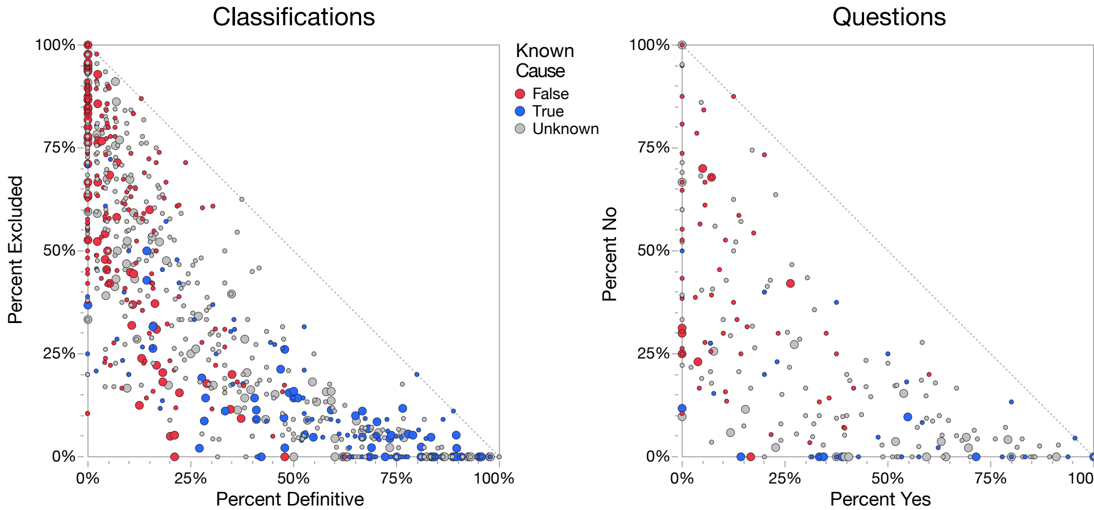


*Figure 10. Consensus on classifications and questions. Each prompt (classification or question) is shown as a circle, positioned based on proportion of responses. Prompts assessed as "most consequential" are shown as larger circles. Responses were unanimous (i.e., superimposed on the 100% corners) on 26 classification prompts and 5 questions. (27,038 responses on 815 classification prompts; 5,967 responses on 223 questions)*

These results show that consensus was limited, and errors were widely distributed across prompts: only 3% of prompts received unanimous responses (i.e., 100% on the x- or y-axis), 33% of prompts had at least 75% consensus (293 classification prompts and 54 questions), and 81% of prompts had a majority consensus (649 classification prompts and 190 questions). If there were strong consensus, we would see clumps at the top left and bottom right (and potentially bottom left) of Figure 10 with few points in between. If we assess just the *excluded* vs. *not excluded* decision (combining *definitive* and *included* as *not excluded*), 13.9% of responses to classification prompts had unanimous consensus (i.e., were at 0% or 100% on the y-axis), and 63.2% had at least a 75% supermajority consensus (≥ 75% or ≤25% on the y-axis).

Majority responses were rarely incorrect. For classification prompts, responses with a 95% supermajority never contradicted known cause; 75% supermajority responses contradicted known cause on four prompts (1.0% of 392 classification prompts with known cause); a majority contradicted known cause on an additional five prompts (nine total, 2.3%); and a plurality contradicted known cause on an additional eight prompts (17 total, 4.3%). On questions, a 75% supermajority never contradicted known cause, a majority (or plurality) contradicted known cause on one question (1.0% of 96 questions with known cause). When limited to the most consequential prompts, the majority was always correct.

There was one sample on which the majority of class responses were errors. Sample 885 (Figure 11) had three classification prompts, resulting in 70 of the 1303 class errors (5.4% of all classification errors). Sample 885 had one question, which resulted in no errors. Note that none of these prompts was most consequential.

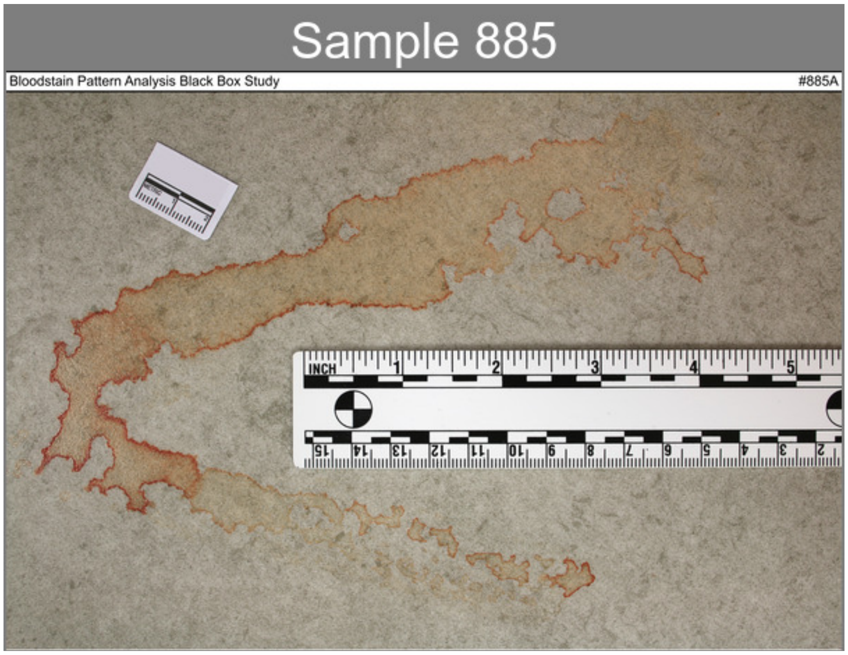


*Figure 11. Sample #885. Controlled. Transfer and swipe of towel soaked with diluted blood. Vinyl flooring. (Description was not provided to participants). Classification prompts: flow (false: 15 def, 18 inc, 12 excl, 45 total); swipe (true: 2 def, 7 inc, 36 excl, 45 total); transfer (true: 10 def, 16 inc, 19 excl, 45 total).*

Errors were widely distributed over samples:

- 121 samples had known cause for at least one classification prompt, 109 of which had at least one error.
- 44 samples had known cause for at least one question: all of those samples had at least one error.

A few prompts received a disproportionate number of errors:

- There were nine classification prompts on which a majority of responses were errors.
- There was one question on which a majority of responses were errors (small N: 3 errors out of 5 responses)

### 3.2.1 Exclusion vs Inclusion

In general, our analyses consider that the response to a classification prompt is a single decision, selecting among the three categories [*excluded* | *included* | *definitive*]. In this section we briefly address an alternative approach, recognizing that the response may be seen as a series of two decisions, first deciding "Exclusion" vs "Inclusion" [*excluded* | *not excluded*] and then differentiating within *not excluded* [*included* | *definitive*].

In terms of consensus, 13.9% (113/815) of classification prompts had unanimous consensus on the [*excluded* | *not excluded*] decision: 11.5% (94/815) were 0% *excluded* (unanimous *not excluded*, shown as 0% on the y-axis in Figure 10 Classifications); 2.3% (19/815) were 100% *excluded* (shown as 100% on the y-axis in Figure 10 Classifications). Such unanimity is (obviously) easier to achieve than in the three-category [*excluded* | *included* | *definitive*] results, in which 3.2% (26/815) of classification prompts had unanimous consensus. Similarly, 63.2% (515/815) of classification prompts had at least a 75% supermajority consensus on the [*excluded* | *not excluded*] decision: 39.0% (318/815) were at least 75% of responses as *not excluded*; 22.9% (187/815) were at least 75% *excluded*. This compares to 36.0% (293/815) of classification prompts had at least a 75% supermajority consensus on the three-category [*excluded* | *included* | *definitive*] results.

In terms of accuracy, the effect of a [*excluded* | *not excluded*] decision can be read directly from Figure 10 Classifications: one could argue that the proportion of "correct" responses would include *definitive* or *included* responses when known cause is *True* (84.8% of responses when known cause is *True*), but then the proportion of "incorrect" responses would include *definitive* or *included* responses when known cause is *False* (40.0% of responses when known cause is *False*).

## 3.3 Reproducibility

In some casework situations, two or more BPA analysts render classification opinions on the same bloodstain evidence, such as during technical review or in court. For this reason, we evaluated the reproducibility (inter-analyst variability) of responses: how frequently they reproduce one another's decisions when each response for a given prompt is compared to all other responses for that prompt. One advantage of reproducibility—and, for that matter, consensus—over accuracy is that it can be assessed for all samples, including those from operational casework. Accuracy, by contrast, requires known cause. Both reproducibility and consensus reflect on the overall reliability of BPA: imperfect reproducibility or consensus limits precision and places an upper bound on accuracy.

Figure 12 summarizes the reproducibility of responses. For example, for every participant who responded *excluded* to a classification prompt, 65.3% of other participants also responded *excluded* to that prompt (agreement rate), 25.3% responded *included*, and 9.4% responded *definitive* (contradiction rate). Across all classification prompts and questions, the agreement rate (the proportion of other participants who had an identical response to a given prompt) was 54.6%; the contradiction rate (the proportion of other participants who had a diametrically opposed response) was 7.8%. If limited to the most consequential prompts, the agreement rate was 56.3% and contradiction rate was 6.2%. Contradictions were distributed broadly across prompts: 549 of the 815 classification prompts and 146 of the 223 questions resulted in contradictions.

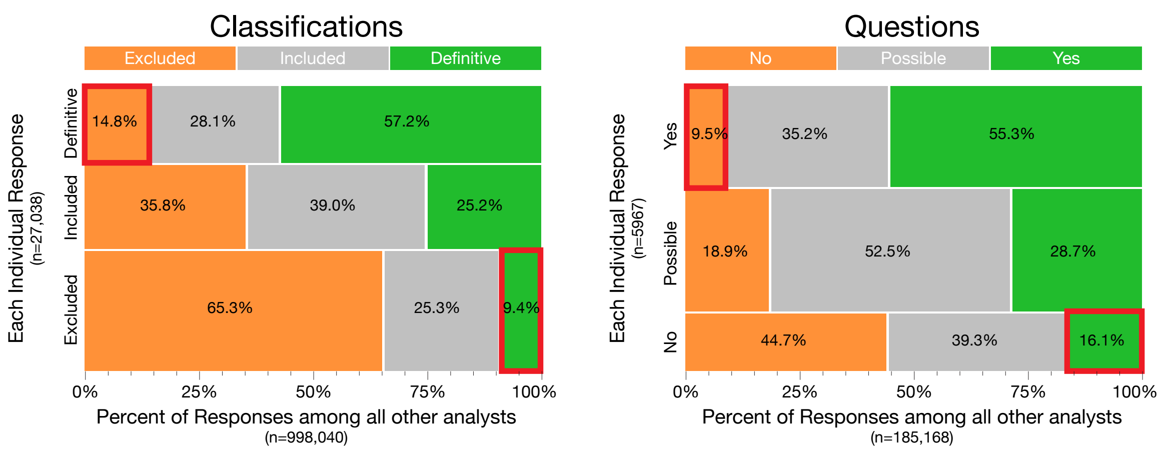


*Figure 12. Reproducibility of responses. Each percentage represents the probability of a second analyst providing a given response, conditioned on the first analyst's response. Contradictory responses are outlined in red. We calculate these probabilities by comparing each response for a given prompt to all other responses for that prompt. Counts are of all pair-wise combinations of responses from different analysts on each prompt.*

Table 17 through Table 19 and Figure 13 and Figure 14 provide additional information supporting the reproducibility results. Contradictions were distributed broadly across prompts: 549 of the 815 classification prompts and 146 of the 223 questions resulted in contradictory responses from participants.

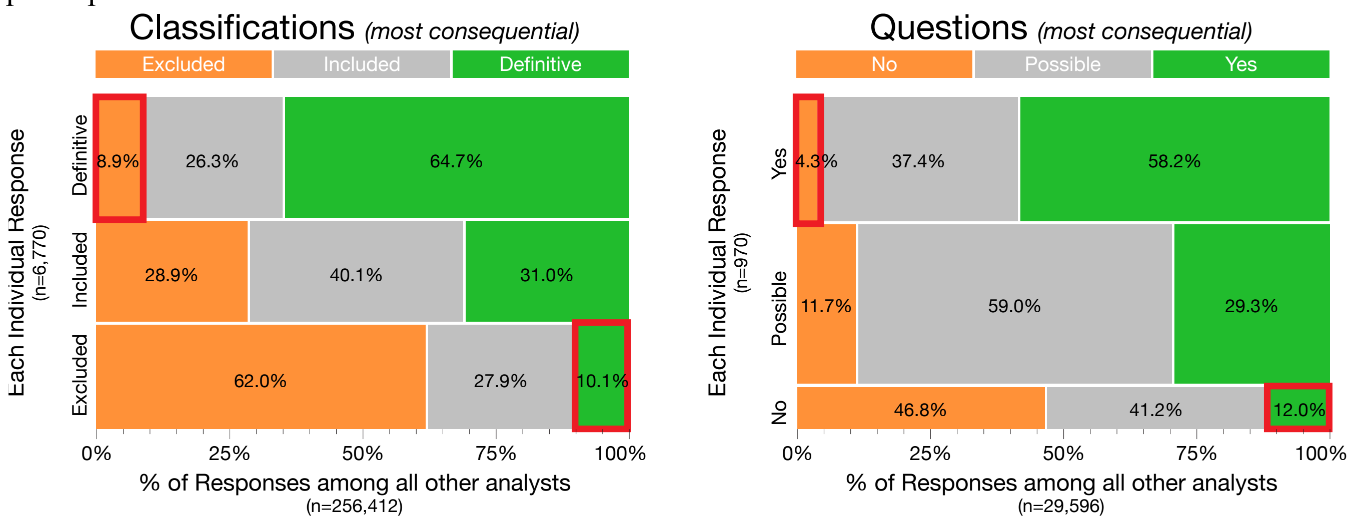


*Figure 13. Reproducibility of responses on prompts assessed as "most consequential". Compare to Figure 12 which shows results for all prompts. Each percentage represents the probability of a second analyst providing a given response, conditioned on the first analyst's response. Contradictory responses are outlined in red. We calculate these probabilities by comparing each response for a given prompt to all other responses for that prompt. Counts are of all pair-wise combinations of responses from different analysts on each prompt.*

When reproducibility of responses was conditioned on known cause (Figure 14), erroneous responses were often reproduced: 17.7% of erroneous *definitives* were reproduced, as were 34.1% of erroneous *excluded* responses, 24.2% of erroneous *yes* responses, and 22.5% of erroneous *no* responses. These results suggest that if two BPA analysts both analyze a pattern (such as occurs operationally during technical review) they cannot always be expected to agree, and if they do agree they may both be wrong.

Figure 14 shows the reproducibility of responses conditioned on known cause, providing a means of assessing the reproducibility of errors. Reproduced errors were distributed broadly across prompts (247 of the 392 classification prompts and 55 of the 96 questions with known cause). There were no prompts that were responsible for a disproportionate number of reproduced errors. If limited to "most consequential" prompts, 18.6% (vs 17.7%) of erroneous *definitives* were reproduced, as were 20.7% (vs 34.1%) of erroneous *exclusions*, 19.5% (vs 24.2%) of erroneous *yes* responses, and 6.6% (vs 22.5%) of erroneous *no* responses.

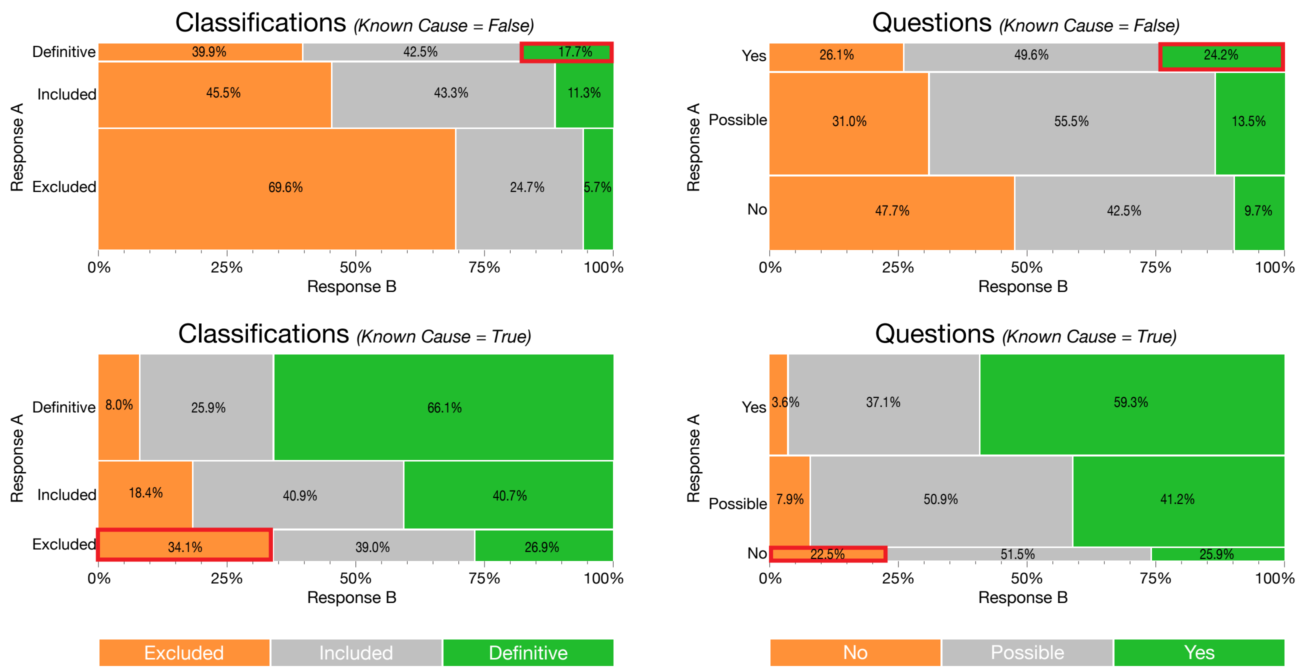


*Figure 14. Reproducibility of responses conditioned on known cause. Compare to Figure 12, which shows overall results. Reproduced errors are outlined in red. (N: Classifications(False): 6,941 individual responses, 236,120 pairwise combinations of responses; Classifications(True): 4693 individual responses, 161,598 pairwise combinations of responses; Questions(False): 1461 individual responses, 40,142 pairwise combinations of responses; Questions(True): 702 individual responses, 18,684 pairwise combinations of responses.)*

| *CLASSIFICATIONS* | | *Individual responses* | *Response B (all pairwise combinations)* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | *Excluded* | | *Included* | | *Definitive* | | *Total* |
| *Response A* | *Excluded* | 11,529 | 278,540 | 65.3% | 107,833 | 35.8% | 39,946 | 14.8% | 426,319 |
| | *Included* | 8,137 | 107,833 | 25.3% | 117,464 | 39.0% | 75,947 | 28.1% | 301,244 |
| | *Definitive* | 7,372 | 39,946 | 9.4% | 75,947 | 25.2% | 154,584 | 57.2% | 270,477 |
| | *Total* | 27,038 | 426,319 | | 301,244 | | 270,477 | | |
| *QUESTIONS* | | *Individual responses* | *Response B (all pairwise combinations)* | | | | | | |
| | | | *No* | | *Possible* | | *Yes* | | *Total* |
| *Response A* | *No* | 1,347 | 17,306 | 44.7% | 15,228 | 18.9% | 6,225 | 9.5% | 38,759 |
| | *Possible* | 2,615 | 15,228 | 39.3% | 42,338 | 52.5% | 23,151 | 35.2% | 80,717 |
| | *Yes* | 2,005 | 6,225 | 16.1% | 23,151 | 28.7% | 36,316 | 55.3% | 65,692 |
| | *Total* | 5,967 | 38,759 | | 80,717 | | 65,692 | | |

*Table 17. Reproducibility of responses. Counts for reproducibility results shown in Figure 9.*

### *3.3.1 Overall Agreement Rate and Overall Contradiction Rate*

Reproducibility can be measured using the overall agreement rate ($OAR$) and overall contradiction rate ($OCR$). These rates are defined as follows, using $A$ for agreement and $C$ for contradiction:

$$OAR = \frac{\sum_{l=1}^{m} \chi_A(l)}{m} \quad OCR = \frac{\sum_{l=1}^{m} \chi_C(l)}{m}$$

Here $m$ is the number of distinct pairwise combinations of responses (on any given prompt) including both forward and backward couplings. $\chi_A(l) = 1$ if the responses are identical, otherwise 0; by contrast, $\chi_C(l) = 1$ if the responses contradict, otherwise 0. OAR is similar but not identical to *percent agreement*, used in [30]. Percent agreement is computed separately for each prompt and then aggregated, whereas OAR is calculated in aggregate from the beginning.

| | All | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Classifications | | | Questions | | | Total | | |
| | Counts | Rates | C.I. | Counts | Rates | C.I. | Counts | Rates | C.I. |
| Contradict (OCR) | 79,892 | 8.0% | 7.1%-9.0% | 12,450 | 6.7% | 5.9%-7.6% | 92,342 | 7.8% | 7.0%-8.7% |
| Same (OAR) | 550,588 | 55.2% | 54.2%-56.1% | 95,960 | 51.8% | 50.8%-52.8% | 646,548 | 54.6% | 53.7%-55.5% |
| (1Indeterminate) | 367,560 | 36.8% | 35.7%-38.0% | 76,758 | 41.5% | 40.5%-42.5% | 444,318 | 37.6% | 36.5%-38.7% |
| Total | 998,040 | | | 185,168 | | | 1,183,208 | | |

*Table 18. Reproducibility rates for all prompts. Counts are of all pair-wise combinations of responses from different analysts on each prompt. "(1Indeterminate)" indicates pairs of responses in which one participant made an indeterminate response.*

| | Most consequential | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Classifications | | | Questions | | | Total | | |
| | Counts | Rates | C.I. | Counts | Rates | C.I. | Counts | Rates | C.I. |
| Contradict (OCR) | 16,770 | 6.5% | 5.7%-7.5% | 972 | 3.3% | 2.5%-4.2% | 17,742 | 6.2% | 5.4%-7.1% |
| Same (OAR) | 144,092 | 56.2% | 55.0%-57.3% | 16,886 | 57.1% | 55.1%-59.0% | 160,978 | 56.3% | 55.2%-57.3% |
| (1Indeterminate) | 95,550 | 37.3% | 36.0%-38.6% | 11,738 | 39.7% | 38.0%-41.3% | 107,288 | 37.5% | 36.4%-38.7% |
| Total | 256,412 | | | 29,596 | | | 286,008 | | |

*Table 19. Reproducibility rates for most consequential prompts. Counts are of all pair-wise combinations of responses from different analysts on each prompt.*

Note that the study assessed reproducibility (inter-analyst) but not repeatability (intra-analyst). Although we considered repeatability in the study design, we determined that any non-trivial bloodstains would be too memorable for a meaningful test of repeatability rates.

### *3.3.2 Gwet's Agreement Coefficient*

An alternative method of assessing reproducibility is Gwet's agreement coefficient (Gwet's $AC_2$). In borderline cases in which it may be difficult for an analyst to decide which of two neighboring responses to select (e.g., *included* vs *definitive*), the choice of response may be more arbitrary than in non-borderline cases. These plus/minus-one category disagreements can be seen as decision noise that could result in the measured OAR or OCR rates either over- or under-representing the underlying variation. Gwet's $AC_2$ corrects *OAR* by adjusting for the possibility of such arbitrary borderline choices ("agreement by chance"). In addition, Gwet's $AC_2$ allows for weighting of decision categories according to their underlying structure, allowing for a more representative penalty structure for disagreements. In this study, the conclusion scale was ordinal, so it was appropriate to penalize analysts less for disagreements in adjacent categories (e.g., *definitive* vs *included*) than for disagreements in contradictory categories (e.g., *definitive* vs *excluded*). Of the many inter-rater reliability statistics discussed in the literature [31], Gwet's $AC_2$ is known to be especially robust because it is insensitive to decision category prevalence. Therefore, Table 20 and Table 21 provide $AC_2$ values for all prompts and the most consequential prompts, each benchmarked to the Landis-Koch verbal scale [32] to facilitate interpretation according to the method described in [31].

| | All prompts | | | | |
|---|---|---|---|---|---|
| | OAR | $C.I._{0.95}$ for OAR | Gwet's $AC_2$ | $C.I._{0.95}$ for $AC_2$ | Landis-Koch Scale |
| Classifications | 0.552 | [0.542,0.561] | 0.472 | [0.450,0.494] | Moderate |
| Questions | 0.518 | [0.508,0.528] | 0.460 | [0.424,0.495] | Fair |
| Both | 0.546 | [0.537,0.555] | 0.458 | [0.441,0.476] | Moderate |

*Table 20. Comparing OAR with Gwet's AC: all prompts. The Landis-Koch verbal scale categorizes agreement into the following verbal equivalent categories: [poor, slight, fair, moderate, substantial, almost perfect].*

| | Most Consequential prompts | | | | |
|---|---|---|---|---|---|
| | OAR | C.I.$_{0.95}$ for OAR | Gwet's $AC_2$ | C.I.$_{0.95}$ for $AC_2$ | Landis-Koch Scale |
| Classifications | 0.562 | [0.550,0.573] | 0.501 | [0.464,0.538] | Moderate |
| Questions | 0.571 | [0.551,0.590] | 0.614 | [0.536,0.691] | Moderate |
| Both | 0.563 | [0.552,0.573] | 0.511 | [0.479,0.543] | Moderate |

*Table 21. Comparing OAR with Gwet's AC: most consequential prompts.*

## 3.4 Comparing Participants

In order to compare the performance of the participants, we identified several metrics that collectively characterize an analyst's performance, as described above: accuracy (what percent of their responses contradict known cause), consensus (what percent of their responses contradict those of the majority of the other analysts), and reproducibility (what percent of their responses contradict those of other analysts when considered pairwise). Each of these metrics was in turn assessed relative to decisiveness (the proportion of analysts' responses that were indeterminate), as shown in Figure 15 below; low rates of error should not be achieved at the cost of extremely high rates of indeterminacy.

Although accuracy (especially on the most consequential prompts) is obviously the performance metric of greatest interest, accuracy results are calculated on a subset of the responses and prompts; we use the consensus and reproducibility metrics to provide supporting information on broader portions of the data. The number of prompts and responses used in calculating each performance metric varied notably, as shown in Table 22. Accuracy metrics are necessarily limited to prompts with known cause, and consensus metrics are limited to prompts that resulted in a majority consensus. Each metric was calculated both on all applicable prompts, as well as on only those prompts assessed as most consequential. Note that decisiveness and reproducibility(all) were calculated on all prompts and responses, regardless of whether the prompt had known cause.

Consensus can be seen as a proxy for accuracy, especially on casework samples where we did not assert known cause. The majority was usually correct: the majority contradicted known cause on nine classification prompts (out of 392 with known cause; 2.3%), and on one question (out of 96 with known cause; 1.0%). When limited to the most consequential prompts, the majority was always correct.

Reproducibility does not purport to assess accuracy, but rather provides a means of comparing participants against each other on all of the data. The participants who contradicted other analysts at rates higher than average usually had higher than average rates of error.

Because some participants completed as few as 5 samples (Table 1), we only calculated rates for individual participants if they completed at least 50 of the assigned 150 samples. Therefore, results comparing participants are limited to 31,803 responses from 56 participants (omitting 1,202 responses from the 19 participants who completed 45 or fewer samples).

Figure 15 compares the participants with respect to each metric. When calculating rates for each participant, we limited analyses to the 56 participants who completed at least 50 samples. Each chart shows three interrelated dimensions: accuracy / consensus / reproducibility (measured as the proportion of erroneous or contradicting responses, y-axis), decisiveness (measured as the proportion of indeterminate responses, x-axis), and effectiveness (color). We assessed the effectiveness of participants in terms of the receiver operating characteristic area under the curve (AUC), which uses the participants' responses as predictors of known cause: two decision thresholds are modeled by considering indeterminate responses as positive or negative (see 3.4.1

below for further details). High AUC values resulted not only from a high number of positive (*definitive* or *yes*) responses on true prompts and negative (*excluded* or *no*) responses on false prompts, but also from a high number of non-negative (*definitive*, *included*, *yes*, or *possible*) responses on true prompts and non-positive (*excluded*, *included*, *no*, or *possible*) responses on false prompts. Participants in the same color band in a given graph in Figure 15 showed similar effectiveness. The most effective analysts are shown in blue closest to the bottom-left corner of each graph. Variation within a color band can be seen as differences in risk aversion among participants, shown as an inverse relationship between accuracy and decisiveness within a given color band. Participants who tended to give more determinate responses (shifted left on the x-axis) were more likely to make mistakes (shifted up on the y-axis), whereas others in the same color band made fewer mistakes at the expense of being less definitive.

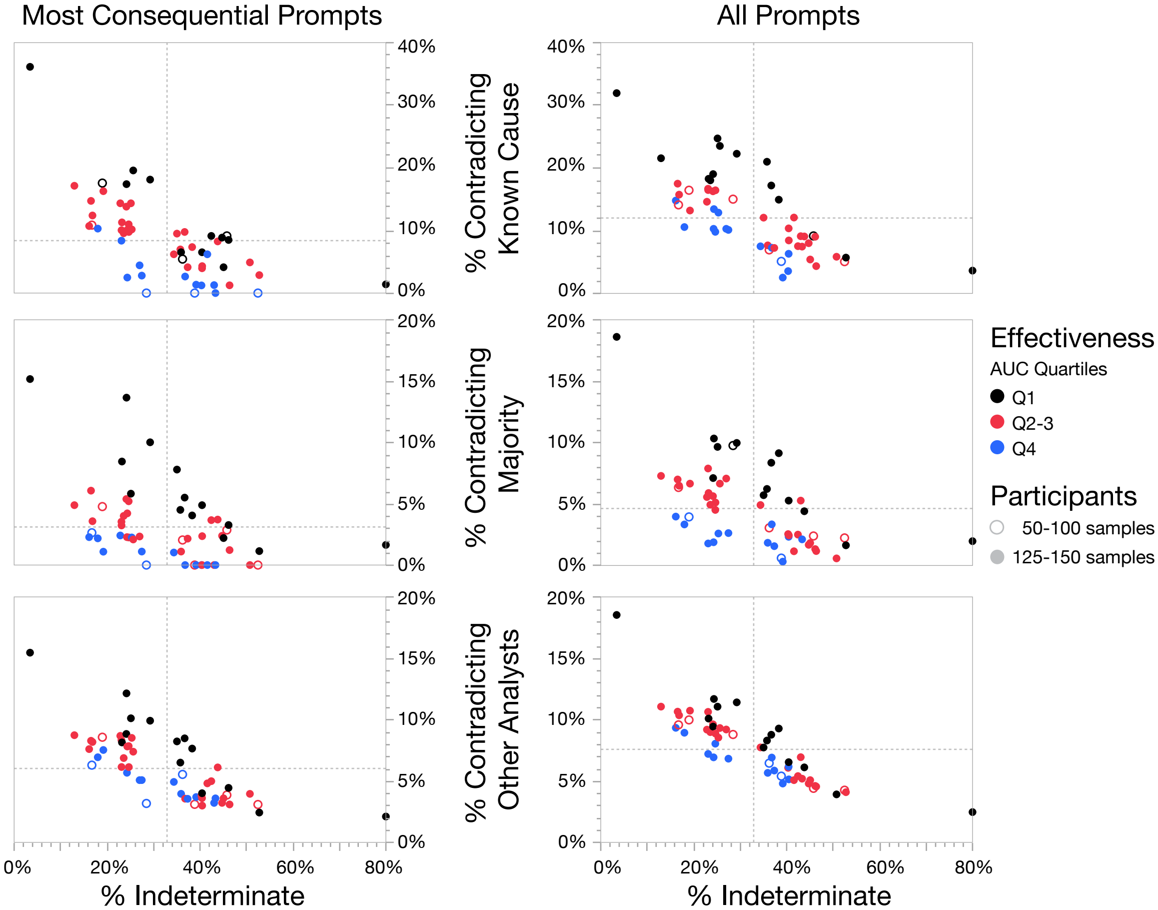


*Figure 15. Decisiveness plotted against six performance metrics: accuracy (% contradicting known cause; top), consensus (% contradicting majority; middle), and reproducibility (% contradicting other analysts; bottom). For each metric, the left panels show results limited to the most consequential prompts; the right panels show results from all prompts. The value on the x-axis for each participant is the same on all panels. Color-coding represents effectiveness for each participant based on the area under the curve (AUC) calculated from receiver operating characteristic (ROC) analysis for each performance measure, in which two decision thresholds are modeled by considering indeterminate responses as positive or negative. Higher quartiles of the AUC correspond to greater effectiveness. Dashed lines represent means. (N= 56 participants (44 with > 100 samples); omits 19 participants with < 50 samples. The 12 participants with fewer than 100 samples are shown as open circles. See Table 22 for number of prompts and responses used for each metric.)*

| Performance Type | Performance Metric | Column Header | # Prompts | Responses [Pairwise] | Mean Responses per analyst (sd) |
|---|---|---|---|---|---|
| Accuracy | % contradict known cause (all prompts) | KCAll | 488 | 11,810 | 211 ± 45 |
| Accuracy | % contradict known cause (MC prompts) | KCMC | 133 | 3,968 | 71 ± 15 |
| Consensus | % contradict majority (all prompts) | MAll | 839 | 19,602 | 350 ± 73 |
| Consensus | % contradict majority (MC prompts) | MMC | 197 | 4,575 | 82 ± 18 |
| Reproducibility | % contradict other analysts (all prompts) | RAll | 1,038 | [1,143,945] | [20,428 ± 4,681] |
| Reproducibility | % contradict other analysts (MC prompts) | RMC | 234 | [277,580] | [4,957 ± 1,154] |
| Decisiveness | % indeterminate | IND | 1,038 | 31,803 | 568 ± 120 |

*Table 22. Performance metrics used in comparing participants. Performance measures with number of prompts and responses for the 56 analysts who completed at least 50 comparisons. Reproducibility is based on comparison of all pairwise combinations of responses on each prompt.*

### 3.4.1 ROC Analysis – Detail

In order to provide a means of measuring the effectiveness of analysts' performance that accounts for both dimensions shown in each graph of Figure 15, we computed the area under the curve (AUC) of a receiver operating characteristic curve (ROC) [33]. For each participant, for each performance metric, the ROC has 2 measurable operating points, defined for positive responses and non-negative responses. (Positive = definitive or yes; negative = excluded or no; indeterminate = included or possible.) The operating points are defined for each performance metric as follows:

- Known cause:
  - P1 = (p(positive|false), p(positive|true))
  - P2 = (p(not negative|false), p(not negative |true))
- Majority consensus:
  - P1 = (p(positive|majority negative), p(positive|majority positive))
  - P2 = (p(not negative| majority negative), p(not negative | majority positive))
- Repro:
  - P1 = (p(positive|negative), p(positive|positive))
  - P2 = (p(not negative|negative), p(not negative|positive))

These points form a curve that originates at (0,0) and terminates at (1,1), and the measure of effectiveness is computed as the area under this curve, providing a measure of the analysts' performance that incorporates both the advantages and the disadvantages of giving indeterminate responses (as an example, see Figure 16). The P1 point plots the proportions of incorrect vs correct positive responses; the P2 point also plots the proportions of incorrect vs correct responses, but treats indeterminate responses as if they were positive. In this way the two points model two different decision thresholds.

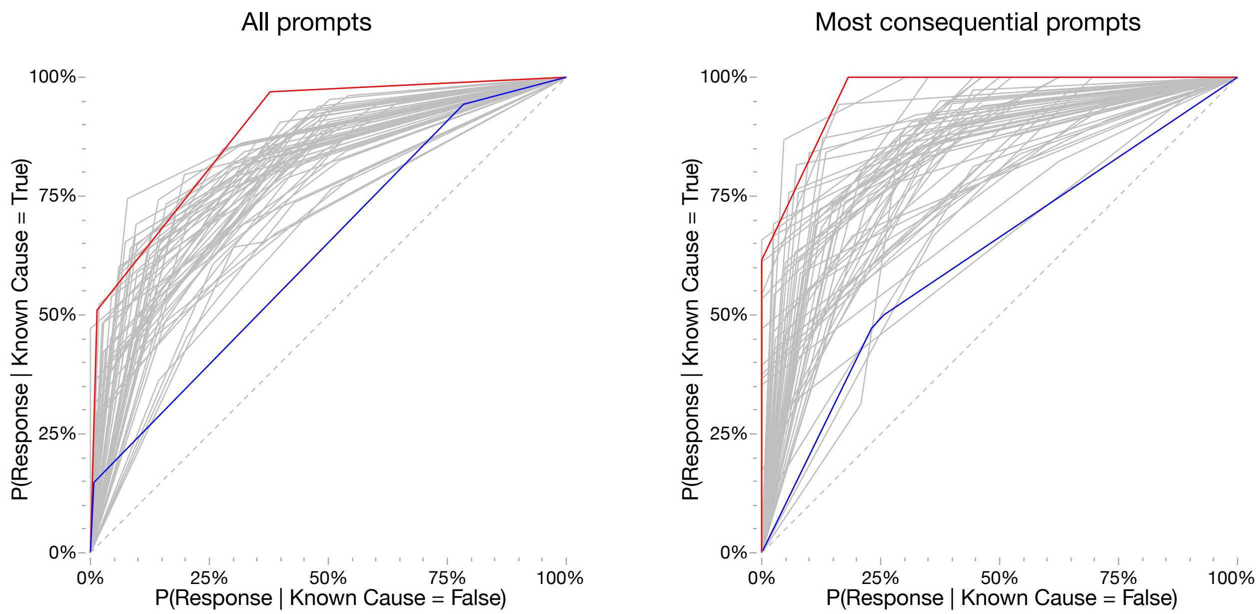


*Figure 16. ROC curves for analyst performance: known cause for all prompts (left) and known cause for most consequential prompts (right). The red line in each figure indicates the highest performance as a function of area under the curve (AUC(ALL) = 0.8854; AUC(MC) = 0.9650). In contrast, the blue line in each figure indicates the lowest performance as a function of area under the curve (AUC(ALL) = 0.6340; AUC(MC) = 0.6246). Note: the response of interest is positive for the first point on each curve (P1) and not negative for the second point on each curve (P2), as previously described. The dashed line on each plot represents an AUC of 0.5, which indicates decision performance based upon chance.*

## 3.5 Error by Participant

In general, the participants exhibited a continuum of performance: errors were widely distributed among participants, and all participants who completed more than 50 samples made multiple errors. However, two participants did show anomalous results. One participant (top left of all charts in Figure 15) contradicted known cause on 36% of responses on highly consequential prompts (32% on all prompts), but was indeterminate on only 4% of responses — that participant was responsible for 5.7% of all errors in the study and 5.0% of all contradictions of other analysts. Another participant (bottom right of all charts in Figure 15) contradicted known cause on only 1% of responses on highly consequential prompts (4% on all prompts), but was indeterminate on 80% of responses. Both of the anomalous participants currently conduct bloodstain pattern analysis as part of their employment, work in a laboratory environment, conduct fewer than 5 BPA cases per year, have testified in court as BPA experts, have at least a master's degree, did not complete a formal program of BPA instruction/supervision, and are not certified by the International Association for Identification (IAI); one is from the US. A total of five participants share these background attributes — note the others who share these attributes did not show problematic performance. Human subjects research protections do not permit revealing further information that could be used to identify these individuals.

No small group of participants was responsible for a sharply higher percentage of the total errors or unreproduced responses than the rest. However, for purposes of illustration, we note that the participants with the five highest error rates were responsible for 20.8% (309) of all errors, and those with the 10 highest error rates were responsible for 35.6% (529) of all errors. In terms of reproducibility, those with the five highest rates of contradictions of other analysts were responsible for 15.8% (14,552) of all contradictions of other analysts, and those with the 10 highest were responsible for 28.3% (26,141) of all contradictions. For comparison, the participants who completed the study with the five lowest error rates were responsible for 2.9% (44) of all errors; the participants who completed the study with the five lowest contradiction rates were responsible for 5.2% (4,793) of all contradictions of other analysts.

### 3.6 Associating Participants' Performance and Background

In order to further characterize the performance of the participants, we developed a novel procedure for detecting and reporting any associations between participants' performance and their background attributes. Performance on each of the seven metrics shown above in Table 22 was assessed for 54 of the participants (omitting the two outliers, and the participants who completed fewer than 50 samples each), with respect to 25 background attributes (shown below in Table 23).

| *Attribute* | *Survey Q#* | *Variable Type* | *Levels* |
|---|---|---|---|
| Highest level of education | 1 | Ordinal | < Bachelors, Bachelors, Graduate |
| Years of experience | 3 | Ordinal | < 5, 5-10, 11-15, 16+ |
| Cases per year | 4 | Ordinal | 0-1, 2-4, 5-10, 11-20, 21+ |
| # times testified | 5 | Ordinal | 0, <10, 10-20, 21+ |
| # times testified (admissibility) | 6 | Ordinal | 0, <10, 10-20, 21+ |
| % of time at crime scenes | 7A | Ordinal | <25, 25-49, 50-74, 75+ |
| # other disciplines | 8 | Ordinal | 1, 2, 3, 4, 5, 6+ |
| Training program | 9 | Ordinal | Informal, 6mo, 6-12mo, 1+ year |
| # training components | 10 | Ordinal | <15, 15-19, 20+ |
| Hours basic workshop training | 11 | Ordinal | 1-40, 41-80, 81+ |
| Hours advanced workshop training | 11 | Ordinal | 0, 1-40, 41-80, 81+ |
| Length of supervised casework | 12 | Ordinal | None, 1mo, 2-6mo, 6-12mo, 1+ year |
| Certification status | 13 | Binary | Have been certified, never certified |
| Employer | 17 | Categorical | Local, state, federal, other |
| Country of practice | 18 | Binary | US, not US |
| Country of practice | 18 | Binary | English, non-native English |
| Country of practice | 18 | Categorical | US, Canada, Italy, Other |
| Law enforcement status | 19 | Binary | Law enforcement, not law enforcement |
| Laboratory work status | 20 | Binary | Work in lab, do not work in lab |
| Laboratory accreditation status | 21 | Binary | Accredited, not accredited |
| # other BPA analysts | 22 | Ordinal | 0, 1, 2, 3 |
| Employer BPA policies | 23 | Binary | Written policies, no written policies |
| Technical review procedure | 24 | Ordinal | Yes (trained), yes (untrained), no |
| % of cases reviewed | 25 | Ordinal | <25, 25-49, 50-74, 75+ |
| Order of case material review | 26 | Binary | Police report first, photos first |

*Table 23. Background attributes of interest. Survey question numbers, variable types, and levels are provided for each attribute: see 2.2 for specifics. Note that in some cases response categories appearing on the survey have been combined in order to ensure sufficient sample sizes for comparisons (minimum of 5 for any category).*

Given the small sample size (54 participants*) with respect to the number of background attributes of interest, we used two complementary approaches for detecting associations with performance: variable importance analysis and attribute-specific significance testing. Variable importance analysis (VIA) was conducted by considering all attributes simultaneously and coupling linear regression and random forest analysis to yield importance scores. In addition, significance testing was conducted for each attribute individually using the Kruskal-Wallis test to yield $p$-values and $q$-statistics. Using these importance scores, $p$-values, and $q$-statistics, we set association thresholds and a reporting criteria hierarchy to determine which (if any) of these background attributes

* *In total, there were 56 participants who completed more than 50 examinations (one-third of the assigned samples) and were therefore deemed suitable for comparison with other analysts. Of these, two participants were anomalous in every performance measure. These anomalous analysts are excluded from this analysis in order to preserve anonymity with respect to background attribute combinations and prevent outliers from skewing the results.*

exhibited sufficient support to indicate an association with performance. For the majority of background attributes (including length of training, educational degree, certification, or length of experience), we found no support for associations with performance. The exceptions were country of practice and extent of advanced workshop training, for which we found limited support for an association with performance.

### *3.6.1 Variable Importance Analysis via Linear Regression & Random Forest*

Although linear regression is commonly used to associate various factors with an observation of interest, it is generally not a suitable technique when the ratio of observations to covariates is low (<10:1) and/or when there is high correlation between predictor variables [34], [35], both of which are the case for this study. Alternatively, random forest regression, a non-parametric technique, is robust even in the presence of small sample sizes and multicollinearity [35], [36]. However, the primary issues for random forest, particularly as related to variable importance, are selecting an unbiased estimator of importance and highlighting important variables even in the presence of redundancy [35], [36].

To overcome these limitations, linear regression was coupled with random forest analysis to conduct variable importance analysis (VIA). Linear regression was used to associate subsets of background attributes with a given performance measure such as % Contradict Known Cause, and random forest analysis was then used to determine the importance of each background attribute based on the goodness of fit measures from the linear regression models. More specifically, the following process was implemented to quantify attribute importance:

1. Randomly sample covariates (background attributes) and use these to build a linear regression model predicting the given performance measure. Retain the AIC* (Akaike Information Criterion, a common model assessment statistic [34]) and the sampled covariate information (which predictors were used to build the model). In order to maintain an observation-to-covariate ratio of approximately 10:1 [34], [35], five distinct predictors were bootstrap-sampled in each trial.
2. Repeat this process, randomly sampling covariates with replacement each time, to build 50,000 bootstrapped linear regression models.
    a. Create an S x M indicator matrix (P), where S represents the total number of bootstrapped models (50,000 in this case) and M represents the total number of predictors (25 in this case). The value in each cell of the matrix is 1 if the model of interest was built using the specific predictor, and 0 otherwise.
    b. Create an S-length vector (R) of the AIC values for each bootstrapped model.
3. Conduct random forest regression, now using vector (R) as the set of responses and indicator matrix (P), constructed in Step 2, as the set of predictors. Retain the

---

* *A variety of model assessment statistics (AIC, BIC (Bayesian information criteria), SSE (sum of squared errors), and Adjusted $R^2$) were initially retained in order to determine which was the most suitable for analyses moving forward. To do so, resulting importance rankings for each of the four model assessment statistics were compared to the boxplots of performance as a function of each background attribute. AIC most closely mirrored the observed effects, without missing nominations of extreme outliers (as was the case for BIC) or labeling too many attributes as extreme outliers wherein many did not exhibit significant associations upon further investigation (as was the case for SSE and Adjusted $R^2$).*

importance scores from the random forest model. For each background attribute, importance was computed as the percent increase in mean squared error (%IncMSE), which represents the relative error increase when values of a specific predictor—in this case, the binary indicator column corresponding to that attribute—have been permuted [35], [37]. Higher values of this metric indicate higher importance for the given variable because shuffling of its values yields larger prediction errors. This metric has been widely adopted as the norm given its low inherent bias and high robustness to various data structures and model fits [35].

a. This process was implemented using the randomForest package in R [37]. Random forest utilizes a bagging (bootstrap aggregation) procedure, which allows for an estimate of prediction error and subsequently the computation of attribute importance via %IncMSE, which is a permutation approach [35], [37]. For each tree constructed, a random sample of approximately 1/3 of the observations are selected and the regression model is trained on this subset [37]. The remaining unselected observations are then predicted using the model, and mean squared error is calculated and retained. To determine covariate importance, the values of each predictor are permuted, the model is re-estimated, and the response values of the out-of-bag samples are predicted using the new model [37]. The MSE from this new model is compared to the original, yielding an estimate of percent change in error (%IncMSE) as a result of the shuffling of predictor values [35], [37].

### 3.6.2 Significance Testing via Kruskal-Wallis

To evaluate whether performance varied between different groups of analysts for each background attribute, the Kruskal-Wallis (KW) test was used. This non-parametric alternative to the traditional t-test or ANOVA analysis does not require that responses be normally distributed [38], which was not necessarily expected for the performance measures used in this study.
Kruskal-Wallis procedure [38]:

1. Rank all N performance scores from smallest to largest.
2. For ties, assign each observation the mean rank of the group of tied attributes.
3. Sum the ranks for each group ($i = 1, 2, \ldots, k$) of the attribute of interest ($R_i$).
4. Compute the Kruskal-Wallis test statistic ($H$).

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{N_i} - 3(N+1)$$

5. When ties exist, correct the test statistic ($H$) by dividing it by $C$ (below, where $m$ represents the total number of ties and $t_i$ represents the number of observations included in the specific tie).

$$C = 1 - \frac{\sum_{i=1}^{m} (t_i^3 - t_i)}{N^3 - N}$$

The Kruskal-Wallis test statistic is approximately chi-square distributed (k-1 degrees of freedom). Therefore, the test statistic is compared to the chi-square distribution to obtain a $p$-value. In addition, the Benjamini-Hochberg *q*-statistic (BH(*q*)) was computed given the large number of background attributes under consideration [39]. The BH(*q*) is essentially a *p*-value adjusted based upon the false discovery rate, rather than the family-wise false positive rate [39], [40]. It is generally used to control for the detection of spurious effects with increased power when many individual tests are conducted, in which case a traditional Bonferroni-adjustment would be overly stringent and conservative [40].

### *3.6.3 Reporting Criteria: Associating Attributes with Performance*

Using the computed evaluation metrics (%IncMSE, *p*-value, and BH(*q*)), a threshold was determined for nominating an attribute association based upon the specific evaluation. For variable importance analysis, the threshold was adaptive depending on the distribution of the importance scores of all attributes. Conversely, for the Kruskal-Wallis *p*-value and associated *q*-statistic, an absolute threshold was set based upon tolerance for false positives and false detections, respectively. These criteria are outlined below.

Association thresholds for each evaluation:

1. For variable importance analysis, the distribution of importance ratings for each performance metric was considered in order to determine whether any attributes held more importance than “typical”. Therefore, the threshold for counting as an important variable is being classified as an extreme outlier with respect to percent increase in MSE. For the purposes of this study, an extreme outlier was defined as an attribute having an importance value greater than the third quartile (75th percentile) by more than three times the inter-quartile range (Q3 + 3IQR).
2. For the Kruskal-Wallis test, the *p*-value was considered for evaluation of the degree of association between attributes and performance. The association threshold for the KW *p*-value is 0.05. Theoretically, 5% of all tests may yield a significant result by chance alone at this significance level.
3. To control for spurious effects, the BH(*q*) for the Kruskal-Wallis test was also considered when evaluating the degree of association between attributes and performance. Any background attribute whose KW *q*-statistic falls below a significance level cutoff of 0.10 counted as meeting the association threshold for BH(*q*). Theoretically, 10% of all detected significant results will truly be null at this significance level (e.g., if there are 5 significant attributes, just 0.5 of those may be false rejections of a true null).

After quantifying the association of each background attribute with various performance metrics and evaluating each against association criteria for the three evaluations, it was then necessary to determine which (if any) attributes exhibited notable effects. In order to avoid reporting spurious associations, the following hierarchy of criteria for reporting associations was developed:

1. If an attribute met the criteria for all three association evaluations (variable importance analysis, Kruskal-Wallis *p*-value, and BH(*q*)) across at least three performance metrics, then the support for an association was deemed *notable* and reported. Six of the seven performance measures (Table 22, all except Decisiveness) are conceptually closely related, in that they are all measures of contradicting some type of control response (known cause, majority, or other analysts). Therefore, to avoid reporting spurious

associations and as an additional means of validation, we judged that an attribute must exhibit evidence of effects across at least three of these six measures to be reported.

2. If an attribute met the criteria for at least two association evaluations (variable importance analysis, Kruskal-Wallis *p*-value, and/or BH(*q*)) across three or more performance metrics, then the support for an association was deemed *limited* and reported with qualifications (e.g., *p*-value less than 0.05, but *q*-statistic not less than 0.10, so the result should be interpreted with caution).
3. Otherwise, there was not sufficient support to indicate a meaningful, stable association between the attribute and performance and no association was reported.

#### *3.6.4 Evaluating Associations between Analyst Background and Performance*

Table 24 details the results for variable importance analysis and significance testing of the 25 background attributes of interest, with respect to seven performance metrics. Based upon the reporting hierarchy outlined above, country of practice (US/Canada/Italy/Other) and extent of advanced workshop training exhibited *limited* associations with performance, which both fit into level 2 of the reporting hierarchy.

We did not detect support for any *notable* associations between participants' background and performance.

| | Contradict KC (ALL) | | | Contradict KC (MC) | | | Contradict Maj (ALL) | | | Contradict Maj (MC) | | | Contradict Rep (ALL) | | | Contradict Rep (MC) | | | Indeterminate | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VIA (4.15) | P (0.05) | Q (0.10) | VIA (4.14) | P (0.05) | Q (0.10) | VIA (4.60) | P (0.05) | Q (0.10) | VIA (5.60) | P (0.05) | Q (0.10) | VIA (3.39) | P (0.05) | Q (0.10) | VIA (1.60) | P (0.05) | Q (0.10) | VIA (1.30) | P (0.05) | Q (0.10) |
| Education | 0.07 | 0.90 | 0.97 | 0.07 | 0.62 | 0.95 | 0.13 | 0.84 | 0.87 | 0.06 | 0.94 | 0.97 | 0.18 | 0.85 | 0.97 | 0.09 | 0.80 | 0.94 | 0.17 | 0.92 | 0.96 |
| Experience | 0.10 | 0.20 | 0.55 | 0.06 | 0.58 | 0.95 | 0.14 | 0.44 | 0.73 | 0.05 | 0.54 | 0.81 | 0.27 | 0.29 | 0.72 | 0.45 | 0.09 | 0.68 | **2.80** | 0.15 | 0.83 |
| Cases/Year | 0.05 | 0.47 | 0.80 | 0.05 | 0.24 | 0.95 | 1.88 | 0.28 | 0.64 | 5.32 | 0.22 | 0.81 | 0.20 | 0.52 | 0.87 | 0.07 | 0.38 | 0.94 | 0.22 | 0.47 | 0.83 |
| Testified | 0.06 | 0.60 | 0.83 | 0.07 | 0.91 | 0.95 | 0.48 | 0.53 | 0.73 | 0.38 | 0.24 | 0.81 | 0.15 | 0.74 | 0.97 | 0.13 | 0.42 | 0.94 | 0.10 | 0.74 | 0.92 |
| Admissibility | 0.06 | 0.85 | 0.97 | 0.09 | 0.90 | 0.95 | 0.11 | 0.75 | 0.87 | 2.73 | 0.45 | 0.81 | 0.12 | 0.83 | 0.97 | 0.16 | 0.95 | 0.95 | 0.18 | 0.75 | 0.92 |
| %CS | 0.05 | 0.60 | 0.83 | 0.07 | 0.47 | 0.95 | 0.13 | 0.38 | 0.73 | 0.11 | 0.57 | 0.81 | 0.15 | 0.99 | 0.99 | 0.05 | 0.78 | 0.94 | 0.08 | 0.85 | 0.96 |
| # Disciplines | 0.07 | 0.28 | 0.64 | 0.25 | 0.06 | 0.80 | 0.20 | 0.68 | 0.84 | 0.16 | 0.06 | 0.61 | 0.13 | 0.75 | 0.97 | 0.16 | 0.28 | 0.94 | 0.07 | 0.56 | 0.87 |
| Training Program | 0.12 | 0.78 | 0.97 | 1.74 | 0.94 | 0.95 | 0.25 | 0.50 | 0.73 | 2.28 | 0.97 | 0.97 | 0.16 | 0.53 | 0.87 | 0.43 | 0.90 | 0.94 | 0.12 | 0.78 | 0.92 |
| Training Components | 0.11 | 0.97 | 0.97 | 0.17 | 0.62 | 0.95 | 0.15 | 0.87 | 0.87 | 0.04 | 0.87 | 0.94 | 0.13 | 0.91 | 0.98 | 0.14 | 0.80 | 0.94 | 0.16 | 0.89 | 0.96 |
| Basic Workshops | **7.02** | 0.08 | 0.33 | 0.47 | 0.35 | 0.95 | 1.25 | 0.23 | 0.64 | **6.12** | 0.09 | 0.61 | 1.91 | 0.32 | 0.72 | **2.37** | 0.11 | 0.68 | 0.80 | 0.27 | 0.83 |
| Advanced Workshops | **21.54** | **0.00** | 0.05 | **9.01** | **0.04** | 0.80 | 0.37 | 0.11 | 0.47 | 0.70 | **0.03** | 0.61 | 0.44 | 0.08 | 0.59 | **4.68** | **0.04** | 0.68 | 0.08 | 0.08 | 0.83 |
| Supervision | 0.58 | 0.16 | 0.51 | 0.24 | 0.43 | 0.95 | 2.49 | 0.08 | 0.39 | 0.42 | 0.39 | 0.81 | 0.96 | 0.12 | 0.59 | 0.06 | 0.43 | 0.94 | 0.42 | 0.46 | 0.83 |
| Certification | 0.05 | 0.45 | 0.80 | 0.05 | 0.58 | 0.95 | 0.28 | 0.48 | 0.73 | 0.06 | 0.38 | 0.81 | 0.21 | 0.74 | 0.97 | 0.05 | 0.43 | 0.94 | 0.18 | 0.66 | 0.92 |
| Employer | 2.05 | 0.68 | 0.89 | 2.07 | 0.77 | 0.95 | **7.52** | **0.02** | 0.17 | 1.44 | 0.47 | 0.81 | 1.02 | 0.13 | 0.59 | **3.71** | 0.82 | 0.94 | **1.33** | 0.39 | 0.83 |
| Country (US/not) | 0.75 | 0.91 | 0.97 | 0.92 | 0.51 | 0.95 | 0.32 | 0.78 | 0.87 | 0.89 | 0.71 | 0.84 | 0.24 | 0.82 | 0.97 | 0.28 | 0.90 | 0.94 | 0.17 | 0.97 | 0.97 |
| Country (Eng/non-native Eng) | **6.65** | **0.02** | 0.18 | 1.26 | 0.42 | 0.95 | 1.05 | 0.17 | 0.60 | 2.51 | 0.12 | 0.61 | 1.32 | 0.20 | 0.71 | 1.42 | 0.14 | 0.72 | 0.21 | 0.41 | 0.83 |
| Country (US/Canada/ Italy/Other) | **18.18** | **0.02** | 0.18 | **7.30** | 0.12 | 0.95 | **24.51** | **0.01** | 0.13 | 3.27 | 0.10 | 0.61 | **70.02** | **0.00** | **0.08** | **5.70** | 0.09 | 0.68 | **38.56** | **0.01** | 0.21 |
| Law Enforcement | 1.10 | 0.95 | 0.97 | 1.23 | 0.95 | 0.95 | 0.98 | 0.37 | 0.73 | 1.29 | 0.80 | 0.91 | 0.55 | 0.36 | 0.76 | 0.87 | 0.52 | 0.94 | 0.29 | 0.39 | 0.83 |
| Lab Work | 0.49 | 0.28 | 0.64 | 0.04 | 0.69 | 0.95 | 0.10 | 0.85 | 0.87 | 0.13 | 0.66 | 0.83 | 0.15 | 0.98 | 0.99 | 0.06 | 0.37 | 0.94 | 0.11 | 0.63 | 0.92 |
| Accreditation | 0.07 | 0.16 | 0.51 | 0.05 | 0.65 | 0.95 | 0.10 | 0.28 | 0.64 | 0.05 | 0.61 | 0.81 | 0.14 | 0.56 | 0.87 | 0.08 | 0.63 | 0.94 | 0.13 | 0.34 | 0.83 |
| # BPA Analysts | 0.13 | 0.32 | 0.67 | 0.20 | 0.83 | 0.95 | 1.67 | **0.03** | 0.22 | 0.06 | 0.31 | 0.81 | 1.31 | 0.08 | 0.59 | 0.14 | 0.47 | 0.94 | 0.42 | 0.33 | 0.83 |
| BPA Policies | 1.08 | 0.07 | 0.33 | 0.13 | 0.40 | 0.95 | **9.22** | **0.01** | 0.13 | 0.04 | 0.62 | 0.81 | 1.26 | 0.14 | 0.59 | 0.08 | 0.74 | 0.94 | 0.26 | 0.32 | 0.83 |
| Tech Review | 0.76 | 0.08 | 0.33 | 1.08 | 0.39 | 0.95 | 0.12 | 0.46 | 0.73 | 0.09 | 0.41 | 0.81 | 0.50 | 0.45 | 0.87 | 0.09 | 0.71 | 0.94 | 0.87 | 0.47 | 0.83 |
| % Cases Reviewed | 0.06 | 0.48 | 0.80 | 0.05 | 0.83 | 0.95 | 0.10 | 0.26 | 0.64 | 0.05 | 0.58 | 0.81 | 0.20 | 0.25 | 0.72 | 0.07 | 0.75 | 0.94 | 0.27 | 0.51 | 0.85 |
| Order of Materials | 0.04 | 0.58 | 0.83 | 0.15 | 0.43 | 0.95 | 0.12 | 0.63 | 0.83 | 0.12 | 0.60 | 0.81 | 0.34 | 0.31 | 0.72 | 0.05 | 0.56 | 0.94 | 0.67 | 0.15 | 0.83 |

*Table 24. Attribute versus performance. Results for variable importance analysis and significance testing (Kruskal-Wallis p-values and q-statistics), with the association threshold listed in parentheses. Cells highlighted yellow met the association criteria for a single evaluation, while those highlighted blue met the criteria for two evaluations. Cells highlighted green met the association criteria for all three evaluations (variable importance analysis, Kruskal-Wallis p-value, Benjamini-Hochberg q-statistic). (N=54 analysts)*

Figure 17 displays the distribution of performance as a function of country of practice for the four measures for which a limited association was detected. Based upon further investigation via a Bonferroni-adjusted Dunn post-hoc analysis (Table 25), Canadian participants exhibited lower contradiction rates and higher rates of indeterminacy as compared to analysts from other countries (excluding US and Canada), when considering the performance metrics based on all prompts. Note that higher rates of indeterminacy may in part explain lower contradiction rates: to be considered a contradiction in this study, a participant must have reported a determinate response (*Definitive*/*Yes* or *Excluded*/*No*), and therefore higher rates of indeterminacy would generally be associated with lower contradiction rates.

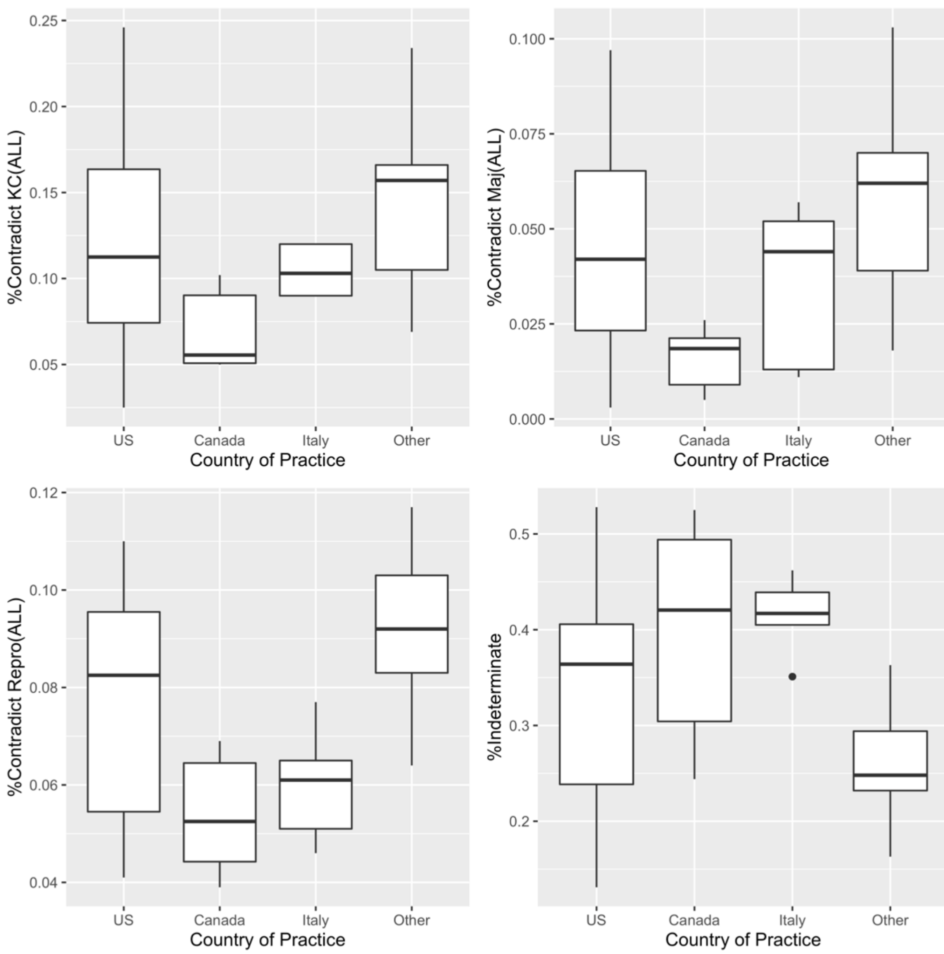


*Figure 17. Distribution of performance measures for associations with country of practice. We detected limited support for an association between participants' country of practice and performance as a function of % contradict known cause on all samples, % contradict majority on all samples, % contradict other analysts (Repro) on all samples, and % indeterminate. (N=54 analysts who completed at least 50 samples; omits 2 extreme outliers)*

| | *KC(All)* | *Maj(All)* | *Repro(All)* | *Indeterminate* |
|---|---|---|---|---|
| US vs Canada | 0.184 | 0.090 | 0.154 | 0.763 |
| US vs Italy | 1.000 | 1.000 | 0.774 | 0.433 |
| US vs Other | 0.571 | 0.730 | 0.352 | 0.343 |
| Canada vs Italy | 1.000 | 1.000 | 1.000 | 1.000 |
| Canada vs Other | 0.012 | 0.007 | 0.006 | 0.047 |
| Italy vs Other | 1.000 | 0.502 | 0.058 | 0.026 |

*Table 25. Bonferroni-adjusted Dunn post-hoc analysis of Kruskal-Wallis test for performance measures exhibiting associations with country of practice. Cells shaded gray indicate Bonferroni adjusted p-values that are significant at the α=0.05 level. (N=54 analysts who completed at least 50 samples; omits 2 extreme outliers)*

Figure 18 displays the distribution of performance as a function of extent of advanced workshop training for the three measures for which a limited association was detected. Most notably, analysts who attended 81+ hours of advanced workshop training generally contradicted known cause less than others when considering all prompts, according to a Bonferroni-adjusted Dunn post-hoc analysis of these distributions (as detailed in Table 26). This observation is of lower stability and utility for the other performance measures, with a slight difference between analysts who completed 1-40 hours of advanced workshop training and those who attended 81+ hours (significant for KC(MC) and just above the threshold for Repro(MC)).

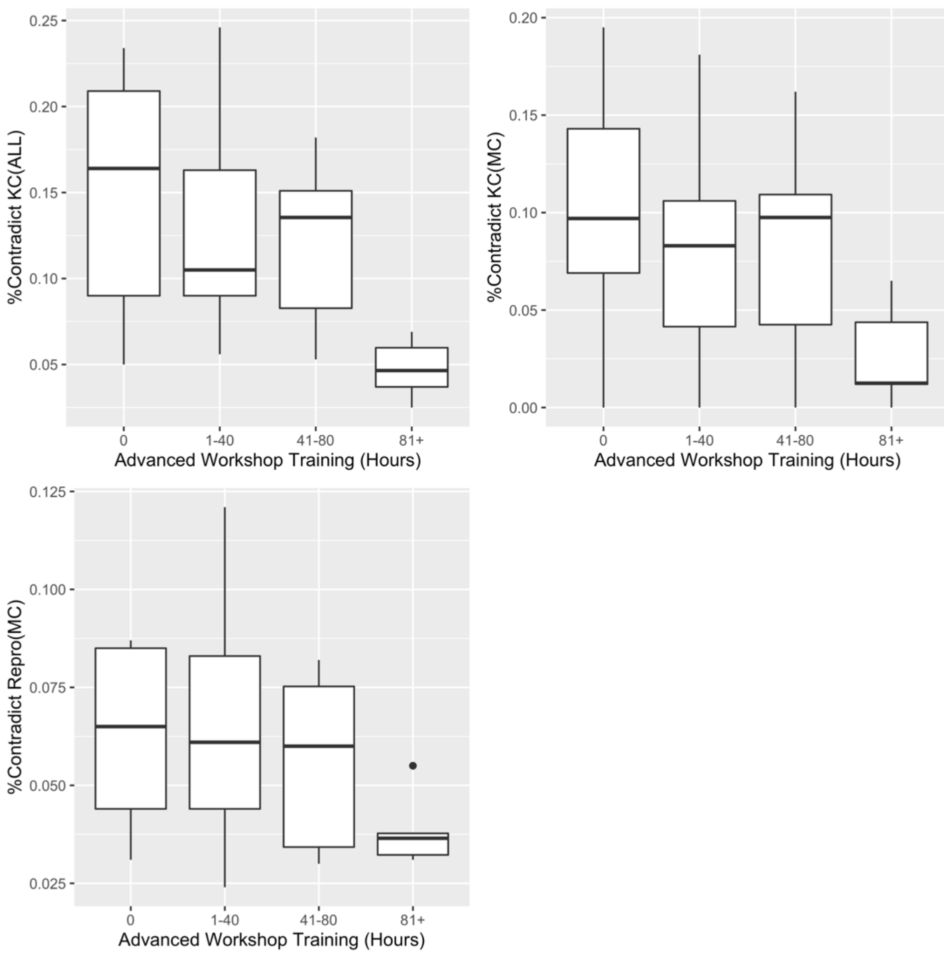


*Figure 18. Distribution of performance measures for associations with extent of advanced workshop training. We detected limited support for an association between participants' extent of advanced workshop training and performance as a function of % contradict known cause on all samples, % contradict known cause on most consequential samples, and % contradict other analysts (Repro) on most consequential samples. (N=54 analysts who completed at least 50 samples; omits 2 extreme outliers)*

| | *KC(All)* | *KC(MC)* | *Repro(MC)* |
|---|---|---|---|
| 0 vs 1-40 | 1.000 | 1.000 | 1.000 |
| 0 vs 41-80 | 1.000 | 1.000 | 1.000 |
| 0 vs 81+ | 0.002 | 0.047 | 0.135 |
| 1-40 vs 41-80 | 1.000 | 1.000 | 1.000 |
| 1-40 vs 81+ | 0.003 | 0.067 | 0.054 |
| 41-80 vs 81+ | 0.012 | 0.071 | 0.915 |

*Table 26. Bonferroni-adjusted Dunn post-hoc analysis of Kruskal-Wallis test for performance measures exhibiting associations with extent of advanced workshop training (N=54 analysts). Cells shaded gray indicate Bonferroni adjusted p-values that are significant at the a=0.05 level. (N=54 analysts who completed at least 50 samples; omits 2 extreme outliers)*

## 3.7 Accuracy and Reproducibility by Pattern Type

Table 27 shows the relationship between pattern classifications used in prompts and error rates. For example, splash has a 45.7% erroneous exclusion rate: 134 out of 293 responses were erroneous on prompts in which the classification was *splash* and the known cause was *true*; erroneous exclusions occurred on all eight such prompts. Although erroneous definitive rates and reproducibility contradiction rates are associated with pattern classifications, the associations are not as striking as those for erroneous exclusions.

Table 28 shows the effect on erroneous definitives of specific combinations of prompts and causes. (NOTE: "Cause" here refers to individual aspects of the known or assumed cause for each sample. This is the ONLY analysis in which we used "assumed cause", assessing the cause for casework samples.) Samples often include multiple causes, either multiple individual aspects of a single pattern (e.g., "*forward spatter*, *impact*, *spatter*") or descriptions of multiple distinct patterns present in a sample (e.g., "*spatter*, *transfer*"). To avoid the effects of individual samples, the table is limited to combinations of prompt and cause in which at least three prompts each received at least three erroneous definitives, and the erroneous definitive rate for the combination of prompt and cause was greater than erroneous definitive rate for the prompt alone.

For example (first row of Table 28), there were six classification prompts for which the cause was (or included) *splash*, and the (false) prompt was *drip pattern*; all six of those prompts each received at least three erroneous definitive responses. As shown, the erroneous definitive rate for *drip pattern* prompts alone was 19%, but the erroneous definitive rate for *impact* prompts where the known cause was or included *satellite stain* was 34% (delta of 15%).

| Class | Erroneous Exclusion | Erroneous Definitive | % Contradict | True Prompts | | False Prompts | | Responses on True Prompts | | Responses on False Prompts | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | All | Resulting in Excl errors | All | Resulting in Def errors | All | Excl errors | All | Def errors |
| Altered | 11.4% | 13.3% | 14.1% | 7 | 5 | 11 | 7 | 185 | 21 | 271 | 36 |
| Backspatter | 22.9% | 4.8% | 3.1% | 4 | 4 | 17 | 9 | 105 | 24 | 522 | 25 |
| Blood clot | 22.1% | 4.1% | 10.6% | 3 | 2 | 5 | 2 | 113 | 25 | 147 | 6 |
| Bubble ring | | | 13.9% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cast-off | 0.0% | 4.3% | 5.5% | 3 | 0 | 8 | 2 | 121 | 0 | 163 | 7 |
| Cessation | 20.3% | 5.7% | 4.4% | 2 | 1 | 6 | 5 | 59 | 12 | 246 | 14 |
| Drip pattern | 5.9% | 18.9% | 9.6% | 7 | 5 | 17 | 15 | 187 | 11 | 466 | 88 |
| Drip stain | 9.7% | 4.9% | 5.3% | 6 | 4 | 4 | 2 | 175 | 17 | 81 | 4 |
| Drip trail | 2.7% | 9.9% | 2.8% | 6 | 3 | 5 | 3 | 149 | 4 | 162 | 16 |
| Expiration | 23.8% | 3.8% | 6.5% | 4 | 4 | 24 | 10 | 147 | 35 | 728 | 28 |
| Flow | 0.0% | 17.9% | 9.0% | 2 | 0 | 3 | 1 | 66 | 0 | 84 | 15 |
| Forward spatter | 14.3% | 10.3% | 4.7% | 3 | 2 | 14 | 9 | 84 | 12 | 429 | 44 |
| Impact | 7.4% | 13.3% | 7.0% | 13 | 10 | 29 | 24 | 390 | 29 | 965 | 128 |
| Insect stain | | 2.4% | 2.9% | 0 | 0 | 6 | 4 | 0 | 0 | 170 | 4 |
| Perimeter stain | | | 11.5% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pool | 2.7% | 8.3% | 9.0% | 5 | 2 | 3 | 1 | 147 | 4 | 84 | 7 |
| Projected | 41.2% | 6.4% | 8.6% | 6 | 6 | 39 | 25 | 216 | 89 | 1191 | 76 |
| Satellite stain | 37.6% | | 10.1% | 6 | 5 | 0 | 0 | 165 | 62 | 0 | 0 |
| Saturation stain | 7.6% | 1.2% | 6.1% | 5 | 2 | 3 | 1 | 131 | 10 | 81 | 1 |
| Serum stain | | | 9.6% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Spatter | 7.3% | 0.0% | 10.5% | 34 | 25 | 1 | 0 | 1131 | 82 | 45 | 0 |
| Splash | 45.7% | 8.7% | 9.8% | 10 | 10 | 13 | 8 | 293 | 134 | 378 | 33 |
| Swipe | 28.8% | 6.7% | 8.7% | 8 | 8 | 1 | 1 | 222 | 64 | 45 | 3 |
| Transfer | 14.0% | 7.4% | 6.5% | 18 | 14 | 11 | 4 | 457 | 64 | 311 | 23 |
| Void | 0.0% | | 11.9% | 2 | 0 | 0 | 0 | 62 | 0 | 0 | 0 |
| Wipe | 13.6% | 9.1% | 10.3% | 3 | 1 | 15 | 11 | 88 | 12 | 372 | 34 |
| Mean | 15.2% | 8.5% | 8.0% | 157 | 113 | 235 | 144 | 4693 | 711 | 6941 | 592 |

*Table 27. Error rates and reproducibility contradiction rates for classification prompts by classification type. Rates greater than the mean that are based on more than two prompts are highlighted.*

| Prompt (False) | Cause | Erroneous Definitive Rate | | Prompts with 3+ Erroneous Definitives / All prompts | Erroneous Def responses / All responses |
|---|---|---|---|---|---|
| | | Prompt&Cause | Prompt | | |
| Drip pattern | Splash | 34% | 19% | 6/6 | 44/131 |
| Impact | Satellite | 30% | 13% | 4/4 | 34/113 |
| Impact | Swipe | 26% | 13% | 3/5 | 23/87 |
| Drip pattern | Spatter | 26% | 19% | 9/9 | 58/222 |
| Impact | Transfer | 25% | 13% | 4/6 | 27/107 |
| Altered | Spatter | 23% | 13% | 4/4 | 19/84 |
| Splash | Satellite | 22% | 9% | 3/3 | 24/111 |
| Impact | Expiration | 16% | 13% | 7/8 | 48/296 |
| Forward spatter | Impact | 14% | 10% | 5/8 | 37/259 |
| Forward spatter | Backspatter | 14% | 10% | 4/7 | 30/214 |
| Impact | Spatter | 14% | 13% | 18/28 | 128/948 |
| Wipe | Swipe | 13% | 9% | 4/8 | 25/188 |
| Splash | Drip pattern | 13% | 9% | 4/9 | 32/256 |
| Wipe | Transfer | 9% | 9% | 5/13 | 29/314 |
| Backspatter | Forward spatter | 9% | 5% | 3/6 | 19/214 |
| Expiration | Forward spatter | 8% | 4% | 3/5 | 14/172 |
| Projected | Spatter | 8% | 6% | 8/31 | 72/960 |
| Backspatter | Impact | 6% | 5% | 3/11 | 21/363 |
| Expiration | Impact | 5% | 4% | 4/16 | 27/500 |

*Table 28. Erroneous definitive rate with respect to combinations of prompts and causes. Combinations of classifications and causes that had a higher erroneous definitive rate than that for the prompt alone. (Limited to instances in which at least three distinct pairings of the specified classification prompt vs cause each resulted in three or more erroneous definitives.)*

## 3.8 Semantic Issues

Many of the disagreements between participants — and some of the errors — may be attributed to semantic differences rather than contradictory interpretations of samples. Such semantic issues include inadequate delineation between some pattern types (such as between *splash* and *drip patterns*, or whether *cessation* is a subcategory of *cast-off*), and ambiguity of some definitions (such as the minimum quantity necessary to classify a pattern as a *pool* or *saturation stain*, or the types of external factors that should be considered *alteration*).

As evidence that some apparent errors and contradictions may be due to semantic differences, observe that even for the simplest bloodstain patterns there was notable disagreement: when provided a classification prompt of *spatter* on samples consisting of a single drop of blood on a non-porous horizontal surface (a *drip stain*), out of 105 responses 42 were *definitive* and 46 were *excluded*. Even such an apparently simple example becomes more complex after further review, because such *drip stains* can contain small *satellite stains*. The 42 *definitive* responses may reflect participants who are seeing the *satellite stains* and calling them *spatter*, or they are referring to *drip stains* as *spatter*, or both. The 46 excluded responses presumably did not see *satellite stains* and presumably do not consider *drip stains* as *spatter*.

As Table 27 and Table 28 show, in the responses to classification prompts, errors were disproportionately associated with certain pattern types; this may in part be explainable by semantic issues. Participants erroneously excluded more than 35% of *splash*, *projected*, and *satellite* patterns (Table 27). Participants often incorrectly concluded that *splash* patterns were *drip patterns* (34% erroneous definitive rate), and often incorrectly concluded that *satellite stains* were *impact* patterns (30%). Issues with interpretation of *satellite stains* were reported in prior research [16]. These results indicate that there was not general agreement among participants on the delineation between *splash* and *drip patterns*; the high rates of exclusion on *projected* and *satellite* patterns may be attributable to a combination of semantic issues and differences in interpretation. Such semantic issues sometimes limited our ability to define known cause for a specific classification or question, even when video of the event was available. In the post-study survey, participants indicated that "several examples stretched the semantic interpretation of definitions," in particular regarding *projected* and *impact* patterns; this was bolstered by analysis of the short text responses.

In order to limit the effects of semantic issues, we calculated and report throughout results for both all prompts and the most consequential prompts. Although some semantic disagreements would presumably be unlikely to have significant consequences in actual casework, their prevalence obscures the extent of serious disagreements. This lack of agreement on the meaning and usage of BPA terminology and classifications illustrates the need for improved standards.

## 3.9 Analysis of Short Text Responses

In total, there were 1,760 short text responses collected as part of this study. As previously discussed, the short text responses were evaluated in two phases: classifications were extracted and associated with specific known cause(s) when possible and responses were assessed holistically for quality and thoroughness.

### *3.9.1 Short Text Responses vs Known Cause*

Each short text response was reviewed by the research team (which included two BPA experts) in order to extract classification responses and assign said classifications a conclusion (*definitive*, *included*, or *excluded*) and a classification type (physical characteristic observation, super-classification, or classification). Overall, the short text responses yielded 4,648 extracted classification decisions, of which 1,976 were on casework samples and 2,672 were on controlled samples.

The open-ended nature of the short text responses revealed that when unprompted, analysts were much more likely to report *definitive* classifications (68.9% of extracted decisions (3,201)) than *excluded* classifications (1.2%), and somewhat more so than *included* classifications (29.9%). In other words, the short text responses overwhelmingly described the mechanisms that the participants concluded happened or may have happened, and rarely mentioned mechanisms that they concluded did not happen. This parallels what would be expected in actual casework where analysts report the mechanisms that created a pattern but generally would not indicate how a pattern was <u>*not*</u> created.

Figure 19 details the distribution of responses as a function of the extracted classification; Figure 20 details the distribution of responses as a function of classification type. With respect to type, 14.3% of the 4,648 extracted classification decisions were physical characteristic observations, 28.0% were super-classifications, and 57.7% were classifications. Participants were more likely to report a *definitive* decision for physical characteristic observations and super-classifications than for classifications. In particular, participants were much less likely to be *definitive* for spatter-related classifications (49.1% of decisions were *definitive*) and somewhat less likely for non-spatter-related classifications (70.4%) as compared to both physical characteristic observations (79.0%) and super-classifications (81.2%). For the seven extracted classification terms that yielded a minority proportion of *definitive* decisions— wherein *definitive* decisions were reported less than 50% of the time for the given classification— five were spatter-related classifications (*projected, castoff, expiration, forward spatter, backspatter*) and two were non-spatter-related classifications (*splash, insect stain*), as shown in Figure 19.

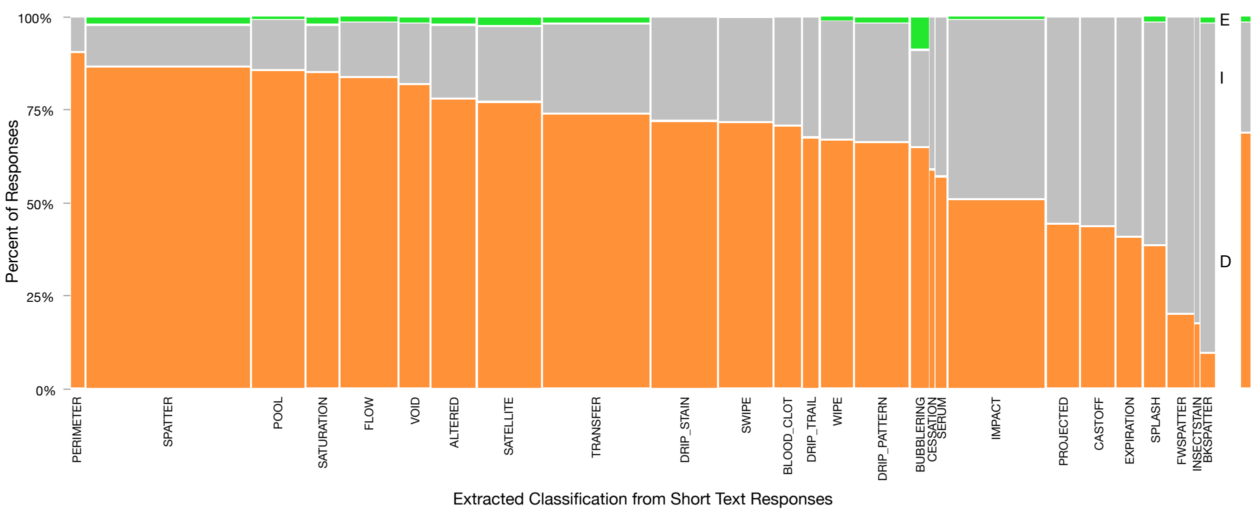

*Figure 19. Distribution of responses as a function of extracted classifications. (n=4,648 classification decisions from short text responses)*

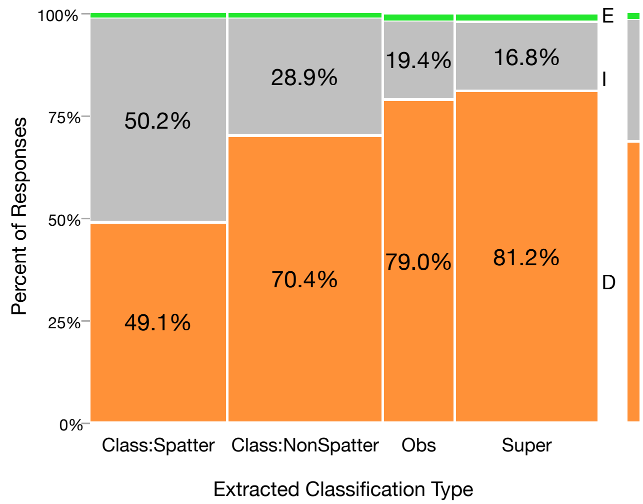


*Figure 20. Distribution of responses as a function of extracted classification type. (n=4,648 classification decisions from short text responses)*

Of all 1,760 short text responses, 1,052 could be assessed against known cause because they originated from a controlled sample and yielded at least one extracted classification associated with a known cause (*true* or *false*). Each of these 1,052 short text responses was then categorized into one of four categories: All Correct (all determinate classifications agreed with known cause); All Errors (all determinate classifications contradicted known cause); Mix (at least one determinate classification agreed with known cause and at least one determinate classification contradicted known cause); or All Indeterminate (all classifications reported were indeterminate).

Table 29 details the distribution of short text responses by known cause assessment categories. Most of the responses evaluated with respect to known cause were all correct (70%). In contrast, 5% were entirely erroneous, and an additional 11% were partially correct and partially erroneous (i.e., included both correct and incorrect statements). The remaining 14% of the short text responses that could be assessed against known cause were indeterminate.

| | *N Short Text Responses* | *% (known cause)* |
|---|---|---|
| All Correct | 739 | 70.2% |
| Mix | 118 | 11.2% |
| All Errors | 50 | 4.8% |
| All Indeterminate | 145 | 13.8% |
| Subtotal (known cause) | 1,052 | |
| Subtotal (unknown cause) | 708 | |
| Grand Total | 1,760 | |

*Table 29. Summary of short text response known cause assessments. The set of classifications extracted from each short text was compared against known cause, when possible, to obtain an overall estimate of accuracy for the conclusion.*

Table 30 details the distribution of short text responses by known cause assessment as a function of summary classification types. More specifically, each tweet could have yielded one or more classifications (and associated types). To assess short text response accuracy as a function of type, it is therefore necessary to combine all classification types into a summary type as follows:

- Obs only: short text response yields one or more physical characteristic observation(s) (no other types)
- Super only: short text response yields one or more super-classification(s) (no other types)
- Class only: short text response yields one or more classification(s) (no other types)

- Obs+Super: short text response yields one or more physical characteristic observation(s) and one or more super-classification(s) (no classifications)
- Obs+Class: short text response yields one or more physical characteristic observation(s) and one or more classification(s) (no super-classifications)
- Super+Class: short text response yields one or more super-classification(s) and one or more classification(s) (no physical characteristic observations)
- All: short text response yields one or more physical characteristic observation(s), one or more super-classification(s), and one or more classification(s)

With respect to summary types, entirely erroneous tweets were disproportionately comprised of classification responses—two-thirds of the entirely erroneous short text responses occurred when participants reported only classifications (31 were due to a single erroneous classification and 2 were due to two erroneous classifications). An additional 12 entirely erroneous responses occurred when participants reported at least one classification along with other types— 9 of these included incorrect classifications. Thus, 84% (42/50) of all entirely erroneous short text responses reported in this study can be attributed to erroneous classifications. Similarly, erroneous classification responses were largely responsible for short text responses that were assessed as Mix (i.e., partially correct and partially incorrect); nearly one-quarter occurred when participants reported only classifications, indicating that at least one classification decision was erroneous. An additional 85 Mix responses occurred when participants reported at least one classification along with other types— 68 of these included incorrect classifications. Thus, 80% (94/118) of all partially correct/partially incorrect short text responses reported in this study can be attributed to erroneous classifications.

In total, 182 of the 4,648 extracted decisions were erroneous and contributed to the 168 short text responses assessed as All Errors or Mix; 79% of these incorrect extracted decisions were classification type responses (12% were physical characteristic observations and 9% were super-classifications). These results suggest that errors are most likely to occur on classification responses, despite them being less likely to yield a determinate decision in the first place (i.e., a decision of *definitive* or *excluded* is required in this study in order to assess accuracy with respect to known cause). These results suggest that classification decisions, which are the most specific type and infer causal mechanisms, may benefit from additional standardization to improve their definitions and more specifically delineate the necessary observations and criteria required to report these types of responses in order to prevent reporting decisions that may be more specific than the evidence warrants.

| | *N Short Text Responses* | *Obs Only* | | *Super Only* | | *Class Only* | | *Obs+ Super* | | *Obs+ Class* | | *Super+ Class* | | *All* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All Correct | 739 | 22 | (3%) | 56 | (8%) | 218 | (30%) | 28 | (4%) | 52 | (7%) | 283 | (38%) | 80 | (11%) |
| Mix | 118 | 3 | (3%) | 1 | (1%) | 26 | (22%) | 3 | (3%) | 14 | (12%) | 47 | (40%) | 24 | (20%) |
| All Errors | 50 | 1 | (2%) | 2 | (4%) | 33 | (66%) | 2 | (4%) | 5 | (10%) | 5 | (10%) | 2 | (4%) |
| All Indeterminate | 145 | 6 | (4%) | 13 | (9%) | 78 | (54%) | 2 | (1%) | 7 | (5%) | 31 | (21%) | 8 | (6%) |
| Subtotal (known cause) | 1,052 | 32 | | 72 | | 355 | | 35 | | 78 | | 366 | | 114 | |

*Table 30. Summary of short text response known cause assessments by summary type.*

Of the 75 participants in this study, 43 (57%) did not have any short text responses that were entirely erroneous, 20 (27%) had one entirely erroneous text response, and the remaining 12 (16%) reported two to four responses that were entirely erroneous. Seven of the participants who completed the study (and thereby had 30 short text responses each) had no known cause errors extracted from their text responses. We did not find notable associations between participants' rates of errors on prompts and errors extracted from text responses: the participants with high error rates on prompts did not have high rates on extracted decisions, and vice versa.

### 3.9.2 Short Text Response Issue Categorization

As discussed in Section 2.7, the BPA experts on the team evaluated the quality and thoroughness of the short text responses; Table 31 details the results of this evaluation.

Four types of issues were counted as "Procedural Errors": *widely unsupported*, *error in reconstruction*, *inaccurate observations*, and *observations do not support pattern classification*. Approximately three-quarters of all 1,760 short text responses did not contain any issues, and an additional 13.2% were lacking observations or thoroughness. The remaining 11.6% were either clerical errors (0.3%) or errors in observations, reconstruction and/or unsupported conclusions (11.3%).

| | *Widely Unsupported* | *Error in Reconstruction* | *Inaccurate Observations* | *Observations X classification* | *N Text responses* | *%* | |
|---|---|---|---|---|---|---|---|
| None | | | | | 1,323 | 75.2% | |
| Not thorough | | | | | 232 | 13.2% | |
| Clerical error | | | | | 6 | 0.3% | |
| Procedural error | | | | X | 26 | 1.5% | 11.3% |
| | | | X | | 38 | 2.2% | |
| | | | X | X | 2 | 0.1% | |
| | | X | | | 33 | 1.9% | |
| | | X | | X | 1 | 0.1% | |
| | X | | | | 79 | 4.5% | |
| | X | | | X | 3 | 0.2% | |
| | X | | X | | 5 | 0.3% | |
| | X | X | | | 10 | 0.6% | |
| | X | X | X | | 2 | 0.1% | |
| | | | | | 1,760 | | |

*Table 31. Frequency of short text issue types.*

With respect to the 232 short text responses that were categorized as not thorough, 51 participants provided at least one response that fell into this category. Interestingly, the two anomalous participants in performance on classification and question prompts (one had the highest error rate and the other had the highest indeterminate rate) also had the most short text responses recorded as not thorough (19 and 22 out of 30 total).

Of the 199 short text responses that were considered procedural errors, 47 (19%) were committed by just four participants. All four participants currently conduct bloodstain pattern analysis as part of their employment but do not work in a laboratory setting or an accredited agency, have at least a Bachelor's degree, received less than six months of formal training (2 informal only, 2 formal less than 6 months), completed less than 6 months of supervised casework, and have never been told they failed a proficiency test; three of the four are from the US. A total of 10 participants meet these criteria – note the others did not show problematic performance. Five of the 45 participants who completed the study had no short text responses with procedural errors.

In addition to evaluating the responses for issues, we also determined whether short text responses rendered on controlled samples contained reconstruction statements and assessed any such statements for accuracy based upon the known deposition mechanisms. In total, 206 of the short text responses made reconstruction statements beyond simple classifications that could be assessed with respect to the known deposition mechanisms, 46 (28%) of which were inaccurate.

## 4 List of Products

The following products of this study are included as attachments to this report:

- Supplemental Data S1: Participant Instructions and FAQ (Frequently Asked Questions)
- Supplemental Data S2: Proofsheets (low-resolution summary images) and responses for all 192 samples
- Supplemental Data S3: Response Data (sample descriptions, classification prompts and questions with responses, short text responses, de-identified survey responses, summary results by participant)
- Supplemental Data S4: Example videos showing creation of controlled collection samples: gunshot forward spatter and backspatter (Samples 795,314,577,574), expirated (Sample 240), projected (Sample 661), satellite spatter over transfer (Samples 732,211,229,498)

The manuscript (Hicklin, et al. "Accuracy and Reproducibility of Conclusions by Forensic Bloodstain Pattern Analysts") has been submitted for journal publication [1].

The full-resolution images are archived on OSF [25]. This archive contains all images used in the study (532 images from the 192 samples). Each image is labelled "This image is from the NIJ/Noblis Bloodstain Pattern Analysis Black Box Study. It may be used without restrictions, but any usage must include the following citation: Bloodstain Pattern Analysis Black Box Study Dataset (V1, Dec 2020)."

- Noblis, Bloodstain Pattern Analysis Black Box Study Dataset (V1, Dec 2020) https:/doi.org/10.17605/osf.io/2ckhw

Videos showing controlled collection of samples are available to forensic researchers on request.

## 5 Dissemination of Research Findings

The research findings from this study are being disseminated in the following ways:

- This Final Technical Report
- Manuscript (Hicklin, et al. "Accuracy and Reproducibility of Conclusions by Forensic Bloodstain Pattern Analysts" (Submitted to journal))
- Briefing at the 2021 AAFS NIJ R&D Symposium (16 Feb 2021)
- Briefing at the 2021 IAI Educational Conference (planned)
- Briefing at the 2021 IABPA Conference (planned)

## 6 Conclusions

Our results show that conclusions by BPA analysts were often erroneous and often contradicted other analysts. The error rates corroborate the rates measured in previous studies [16], [17]. Such errors could have serious implications if they occurred in casework, as would conflicting conclusions among BPA analysts if those resulted in conflicting testimony in court. Many of the disagreements among BPA analysts — and some of the errors — may be attributable to semantic differences. The results show that there is often a lack of agreement on the meaning and usage of

BPA terminology and classifications, suggesting a need for improved standards. The lack of criteria for classification decisions means that we cannot expect high rates of reproducibility among analysts.

The results here are intended to provide estimates for use in decision making, improving procedures and training, and future research. These results should not be taken to be precise measures of operational error rates: the error rates reported here describe the proportion of erroneous results for this particular set of samples with these particular patterns; these rates cannot and should not be assumed to apply to all BPA analysts across all casework. The discipline of bloodstain pattern analysis is not solely defined by pattern classification, but rather it includes multiple other aspects that were not evaluated within this study. The study differed from operational casework in that analysts were asked to provide responses based solely on photographs, analysts were not provided case-relevant facts that may have aided in making conclusions, and the means of reporting conclusions were different from the manner in which BPA analysts typically reach conclusions. These results do not account for operational quality assurance measures, such as technical review or verification.

## 6.1 Limitations

The data collected as part of this study, and the results produced, must be considered within the constraints of the study design. The study design differed from operational casework in two key ways: the use of photographs for analysis, and the use of classification prompts, questions, and short text for responses.

In the design of the study, the distribution of digital images for analysis was a practical necessity but using photographs for analysis is not necessarily representative of operational casework. Rather than attending a crime scene and/or examining physical evidence as well as photographs, participants in this study were required to produce a response based *solely* on photographs. Based upon the background survey, none of the participants in this study conduct examinations using photographs alone (and three never use photographs at all), so this methodology was a departure from the manner in which they typically conduct casework. Furthermore, during operational casework, analysts can observe inter-relationships between patterns within the context of the entire scene, which can assist in pattern classification. In operational casework analysts are provided with additional case-relevant facts (e.g., observed victim injuries, coroner's reports, etc.), which can help to inform the conclusions reported by the BPA analyst; participants were not provided this contextual information as part of this study. Such contextual information may in some cases provide important information appropriate to the analysis of bloodstain patterns — but contextual information may also provide a source of bias. In this study, we explicitly chose to limit information to that provided in the images. The results of this study do not consider the effect of additional contextual information on the decision-making process; previous works have evaluated the impact of contextual information on performance on a smaller scale [18], [41], [42].

In addition, the mechanism of reporting for this study was different from the manner in which BPA analysts typically reach conclusions. BPA does not have a methodology or conclusion standard that could be adopted for this study, so we had to develop novel methods of collecting responses (classification prompts, questions, short text), as well as the three-level response structure for reporting responses to prompts. The specific impact of requiring analysts to examine and classify patterns in a manner different from that to which they are accustomed is unknown. However, the results from this study do provide estimates of analyst performance in recognizing various pattern types and in utilizing the ASB terminology standard to describe observations. To complement the prompted classifications/questions, short-text summary conclusions were also collected for a portion of each participant's assigned samples; this more closely replicates the manner in which analysts typically report their findings. The results of evaluating the short text responses showed that errors occurred in the short text responses at rates similar to the error rates from the prompts. This provides support for the use of classification prompts and questions in assessing the accuracy of participants.

## 6.2 Recommendations

In conducting this study and performing analyses, the authors developed the following recommendations, which we suggest may be considered by the BPA community in general, by standards bodies, and by laboratory management. These recommendations fall into the following broad categories:

- Methodology and general terminology recommendations (Section 6.2.1)
- Implications for casework (Section 6.2.2)
- Lessons learned (Section 6.2.3)
- Specific terminology recommendations (Section 6.2.4)

### *6.2.1 Methodology and general terminology recommendations*

The rates of errors and disagreements among analysis observed here raise potential concerns for BPA casework. These errors and disagreements can be explained at least in part by the lack of a standard BPA methodology, by limitations in the existing terminology standard, and by inconsistent use of the existing terminology standard by analysts.

In order to reduce errors and achieve greater agreement among analysts, we make the following recommendations:

- ***Stages of the BPA analysis methodology*** — A methodology standard should be developed for BPA that addresses these stages:
    - ***Observation*** — defines a standard process to describe observations and the observable features/attributes of bloodstains
    - ***Classification decision criteria*** — specifies the criteria required to make classification decisions, moving from observations to inferences regarding the cause(s) of a given bloodstain
    - ***Classification exclusions*** — explicitly considers classifications that can be excluded as causes
    - ***Articulation of levels of support for conclusions*** — provides a basis to articulate each specific classification conclusion for a given bloodstain, clearly expressing a level of support for that classification conclusion
- ***Review existing methodology*** — In developing the methodology standard, existing and proposed classification approaches, such as those developed by Bevel and Gardner [2], Wonder [43], James et. al. [3], and Esperança [44], should be reviewed for concepts that should be adopted for use in the new standard.
- ***Improved definitions*** — In improving the terminology standard, existing definitions from a variety of sources should be reviewed in detail to identify the consistencies in verbiage and detect any gaps to be addressed in future updates. The existing ASB definitions should be updated to be more explicit and precise, and should explicitly state any relationship(s)

between classifications. See Section 6.2.4 for specific comments on existing terminology. In particular, the terminology standard should state which patterns are subcategories of others (e.g., *forward spatter* is a subcategory of *impact*, which is a subcategory of *spatter*) as well as the distinctions or overlap between patterns (e.g., define whether *cessation* is distinct from (or a subcategory of) *cast-off*). See Appendix E for a comparison of BPA terminology and associated definitions from a variety of sources, including ASB [20], IABPA [45]–[47], FBI SWGSTAIN [48], and several BPA textbooks [2], [3], [43].

- ***Differentiate observations from causes*** — The terminology standard should explicitly delineate 1) which terms in the standard are observations that can be used to directly describe a bloodstain (e.g., *blood clot*, *bubble ring*, *edge characteristic*), 2) which terms are classifications of the causes of bloodstains that cannot be directly observed but can only be inferred (e.g., *altered stain*, *impact pattern*, *spatter stain*), and 3) which terms assist in reconstruction of the events leading to deposition of a bloodstain (e.g., *area of origin*, *area of convergence*). For example, an analyst does not observe that a bloodstain is *impact*, but observes specific droplet/pattern characteristics and can infer that the cause was *impact* and may be able to further determine the *area of origin* in order to aid reconstruction.
- ***Enhance observation terminology*** — The existing BPA terminology is disproportionately focused on classifications of the causes of bloodstains. The terminology standards should be enhanced to enable standardized, specific observations by developing additional terminology to describe the physical characteristics or attributes that should be observed and used to justify a given bloodstain pattern classification.
- ***Specify criteria for pattern classification decisions*** — The methodology/terminology standards should describe the specific observations and criteria that are required for making each pattern classification decision. Specific observable properties should be defined for each classification in order to explicitly describe the criteria that suffice for reaching a given classification conclusion. Guidance should be provided on what physical characteristics must be observed to report a given classification conclusion, those observations and/or limitations appropriate for reporting an indeterminate classification conclusion, and any absolute exculpatory criteria (e.g., when a specific characteristic is absent, does this indicate a certain exclusion, or could a given classification still be a possible cause?). This may be best achieved through verbal descriptions and accompanying high-quality exemplar imagery. A classification flowchart or hierarchy may be beneficial in visualizing relationships and decision points.
- ***Articulation of levels of support for conclusions*** — We believe that BPA requires standards for articulation and conclusions, which should detail a categorical conclusion scale and the associated verbiage for describing these conclusions — ideally allowing analysts to clearly ascribe a degree of support for a classification decision. For example, an analyst may decide that the characteristics of a given pattern show strong (but not definitive) support for an *impact* classification, but there is currently no standard means of articulating such a conclusion. Upon review of the short text responses collected in this study, it became apparent that there is a wide variety of verbiage used to describe the level of support for classifications (e.g., "likely", "possible", "consistent with", "seems" were all used as synonyms for a decision of *included*), which may lead to a misinterpretation of intended findings by other analysts, police, or in the legal system. Furthermore, we assert that this standard should provide additional requirements to ensure completeness of reporting. For example, several of the short text responses indicated a single pattern

classification with a decision of *included*—this conclusion suggests that there are other potential causes of the stain, but these are not explicitly stated. The reporting of a single classification, especially an indeterminate one, should be accompanied by additional commentary of any limitations and/or other possible mechanisms in order to prevent a single pattern classification being ascribed undue weight.

- ***Targeted research and validation*** **—** In order to develop standardized methodology and terminology, additional empirical research should be conducted in order to evaluate decision criteria and sufficiency thresholds for reaching specific pattern classifications and subsequently reporting appropriate conclusions. Changes to terminology and methodology should be evaluated by additional black-box-type validation studies in an effort to assess the impact of those changes on the accuracy and reliability of BPA analyst classification decisions. Further targeted white-box studies may provide additional insight into the specifics of errors in BPA conclusions that would inform methodology or terminology standards.

### *6.2.2 Implications for casework*

Although the error and reproducibility rates measured here should not be taken to be precise measures of operational error rates, their magnitude and the fact that they corroborate the rates measured in previous studies [16], [17] should raise concerns in the BPA community. Our primary recommendations to address error and reproducibility rates are to develop a standard methodology and enhance the current terminology standard (as detailed in Section 6.2.1). Additionally, we have the following observations regarding potential implications for casework:

- Participants' backgrounds generally did not exhibit support for an association with performance. There was no evidence to suggest that analysts with higher levels of education, more experience, or more frequent casework perform better: no differences in accuracy, consensus, and reproducibility were detected with respect to these attributes. This result should be considered in the development of standards defining minimum qualifications for BPA analysts.
- We did observe that participants with the most advanced workshop training (81+ hours) were generally more accurate than those with less such training. However, this association does not necessarily imply causality: while advanced workshop training would certainly be desirable, the differences in performance measured here may be a reflection of the individuals who currently seek such training rather than the result of the training itself.
- The accuracy rates measured here indicate that a single BPA analyst's conclusion cannot necessarily be assumed to be correct; and the reproducibility rates indicate that two BPA analysts cannot necessarily be assumed to reach the same conclusion. These results support the need for a rigorous verification process for bloodstain pattern analysis conclusions, to be incorporated into casework. Although we did observe some reproduced errors (two analysts agreed on an erroneous conclusion, meaning that a single verification would miss this error), majority conclusions were almost always correct—therefore, multiple independent verifications by different BPA analysts may be a reasonable path forward. We recommend that the BPA community develop standards for technical review and blind verification so that differing conclusions among BPA analysts can be detected and resolved in casework.
- On review of responses, the BPA experts on the study team felt that many conclusions expressed an excessive level of certainty given minimal observable data. This was true in

the short text responses as well as in the forced choices required by the prompts. We believe that developing a standard means of expressing levels of support for conclusions, and clearly defined criteria for making classification decisions would help mitigate this issue (as discussed in Section 6.2.1).

- As discussed under Limitations (Section 6.1), BPA casework is not limited to making classification decisions on photographs. Further studies may be conducted to assess the accuracy and reproducibility of other aspects of BPA, including reconstruction conclusions, the effect of incorporating other case information (e.g., case facts, pathology, forensic reports, other forensic experts' opinions), and conclusions involving multiple patterns.

### 6.2.3 Lessons learned

This study constituted the largest black box study of classification conclusions made by BPA analysts to date. As terminology and methodology standards are developed, studies such as this will need to be conducted to inform the development of these standards and evaluate their efficacy. Here we offer the following lessons learned to aid future researchers in designing their studies:

- A number of analyses in this study were limited by small sample sizes, as was the ability to generalize some results. To assure a larger number of participants, future studies may consider providing incentives for participants; such incentives could include a lump sum monetary reward, a monetary reward per comparison completed, a raffle drawing for prizes, and/or certification credits or another professional benefit.
- For the classification prompts, participants were presented with a set of pattern classifications to evaluate with respect to the given sample — to determine whether the classification was, could have been, or was not the cause of the stain. This directed collection mechanism is notably different from operational settings, in which analysts provide unprompted responses. To better approximate operational settings, an alternative approach would present participants with all pattern types, and have participants mark a response for each in separate checkboxes, as well as mark which pattern type(s) identifies the primary cause of the pattern in the sample. This process would more closely mirror the hypothesis-based decision-making process used by analysts in casework — in which several pattern classifications are simultaneously considered and appropriately included or excluded as potential causes.
- Although there was value in asking questions of participants to assess reproducibility of statements made in the short text responses, creating simple, targeted questions with responses that are easily evaluated proved challenging. Upon review, an alternative approach would be to develop questions within a working group comprised of legal experts in an effort to assess reliability of responses provided during testimony. BPA experts on the research team would essentially "testify" to evidence provided in each controlled sample, and the legal working group would develop a series of yes/no questions that they would ask in direct and cross-examination. These reconstruction-based questions would be based on known cause as well as possible alternative causes. Although this process would certainly be arduous, it may provide an alternative means of gleaning additional information from question data.
- Assessing the operational consequence for classification prompts and questions was very slow and somewhat subjective, limited to the BPA experts on the study team. In retrospect,

as part of the study we should have asked the participants to assess the operational consequence of each prompt. This would provide a voted measure reflecting all of the participants, and would have eliminated this review process.

- Estimating error rates with greater precision will require consistent use of terminology and a standardized method of stating conclusions. Until more detailed and explicit standards are developed and widely adopted in the BPA community, the error rates reported through reliability studies will necessarily be confounded with semantic issues and methodological and reporting inconsistencies. Until these challenges can be mitigated by standardization, it will be difficult to obtain a clear picture of BPA error rates.
- The design of this study is necessarily a departure from what is typically done during operational casework. In order to more closely mirror operational casework, future studies may consider having an in-person physical portion of their test — this may be implemented by setting up mock crime scenes, having physical evidence for examination alongside photographs, etc. which could be conducted at conferences or various testing sites. Note, however, that this process would presumably be a logistical challenge, would require additional time and effort by both researchers and participants, and may be difficult to accomplish at a sufficient scale for statistically useful results.
- Lastly, it should be acknowledged that pattern classification is just one portion of the entire BPA examination process, and a true estimate of overall operational error rates should consider the entire process (e.g., any relevant case information provided, relationships of patterns within the scene, etc.).

#### 6.2.4 Specific terminology recommendations

In addition to the general terminology recommendations outlined previously, we have also identified additional term-specific suggestions for updates and/or clarifications in an effort to maximize clarity and hopefully aid in consistency of usage. These recommendations are delineated below:

***Altered***

The current ASB definition for *altered* is: "A bloodstain with characteristics that indicate a physical change has occurred."

Potential updates/clarifications for this definition include:

- The standard should clarify what constitutes a "physical change". In particular, it should be clarified whether "physical changes" include changes associated with human action (e.g., *wipe* or *swipe*), non-human action (e.g., blowing of wind causing a *flow*), and/or passive aging processes (e.g., clotting or drying).

***Backspatter***

The current ASB definition for *backspatter* is: "A bloodstain pattern resulting from blood drops which can be produced when a projectile creates an entrance wound."

Potential updates/clarifications for this definition include:

- The phrase "which can be produced" can be interpreted as indeterminate (the pattern could be caused by a bullet entrance wound, rather than specifying that the pattern was definitively caused by a bullet entrance wound). This lack of clarity should be resolved by explicitly stating whether or not *backspatter* can be generated by a process not involving a bullet entrance wound. Perhaps, the fact that backspatter or forward spatter may not be produced in some circumstances when a projectile creates an entrance wound or exit wound, respectively, should be a training topic rather than a component of the definition.

- A more generalized approach for defining, describing and classifying projectile-related spatter should be developed. A single term for "impact spatter associated with gunshot" (a superset of *forward spatter* and *backspatter*) should be developed. Complicating factors such as avulsion, amputation, widely varying amounts of projectile kinetic energy, projectile characteristics, projectile number, the number of wounds, relative wound locations, complex anatomy and others should be considered.
- In the phrase "when a projectile creates an entrance wound", "when" implies an undefined time frame and will create challenges when attempting to discern backspatter from subsequent bleeding from the corresponding wound(s).
- "Creates an entrance wound" is rather narrowly defined. This phrase makes it ethically and practically impossible to create true backspatter (and true forward spatter) for training and research purposes without creating a wound. Consideration should be given to establishing validated methods for simulating certain bloodstain patterns including backspatter, forward spatter, and projected patterns even if these pattern types are re-defined more generally.
- The standard should explicitly state that *backspatter* is a subcategory of *impact*, which in turn is a subcategory of *spatter*.

***Blood clot***

The current ASB definition for *blood clot* is: "A gelatinous mass formed by a complex mechanism involving red blood cells, fibrinogen, platelets, and other clotting factors."

Potential updates/clarifications for this definition include:

- The standard should describe the physical characteristics and attributes of a *blood clot* in a stain, both for wet and dry stains.
- The standard should explicitly state whether *blood clot* is a subcategory of *altered*.

***Bubble ring***

The current ASB definition for *bubble ring* is: "An outline within a bloodstain resulting from air in the blood."

Potential updates/clarifications for this definition include:

- The standard should address whether surface effects or other mechanisms of deposition (other than *expiration*) can mimic the appearance of *bubble rings* and specify whether such stains also count as *bubble rings*.
- The standard should explicitly state that *bubble ring* is a physical characteristic, not a mechanism of deposition, and clearly delineate the pattern classifications that could be inferred by this observation.



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

***Cast-off***

The current ASB definition for *cast-off* is: "A bloodstain pattern resulting from blood drops released from an object due to its motion."

Potential updates/clarifications for this definition include:

- The standard should clarify whether *cessation* is a subcategory of *cast-off*, or whether the categories are disjoint.
- The standard should clarify whether a *drip trail* can be a subcategory of *cast-off* (e.g., when a *drip trail* is formed from movement of a bloody knife), or whether the categories are disjoint.
- The standard should explicitly state that *cast-off* is a subcategory of *spatter*.

***Cessation***

The current ASB definition for *cessation* is: "A bloodstain pattern resulting from blood drops released from an object due to its abrupt deceleration."

Potential updates/clarifications for this definition include:

- The standard should clarify whether *cessation* is a subcategory of *cast-off*, or if the categories are disjoint.
- The standard should explicitly state that *cessation* is a subcategory of *spatter*.

***Drip pattern***

The current ASB definition for *drip pattern* is: "A bloodstain pattern resulting from a liquid that dripped into another liquid, at least one of which was blood."

Potential updates/clarifications for this definition include:

- The standard should explicitly state whether "dripped" is meant to imply that only gravity is involved in the formation of a *drip pattern*.
- The standard should clarify the distinction between *drip pattern* and *splash*, or alternatively eliminate the distinction altogether.
- The standard should clarify whether *drip pattern* is a subcategory of *impact*, or if the categories are disjoint. In other words, is a liquid considered an "object striking liquid blood?"

***Drip stain***

The current ASB definition for *drip stain* is: "A bloodstain resulting from a falling drop that formed due to gravity."

Potential updates/clarifications for this definition include:

- The standard should clarify whether *drip stain* is a subcategory of *spatter*, or if the categories are disjoint. In other words, is "gravity" considered an "external force applied to liquid blood?"

***Drip trail***

The current ASB definition for *drip trail* is: "A bloodstain pattern resulting from the movement of a source of drip stains between two points."

Potential updates/clarifications for this definition include:

- The standard should specify what constitutes a "source". In particular, it should be clarified whether a "source" includes an actively bleeding human/animal, a non-actively bleeding human/animal, and/or an inanimate object.
- The standard should clarify the phrase "between two points". In particular, it should be clarified whether only two droplets constitute a *drip trail*.
- The standard should clarify whether a *drip trail* can be a subcategory of *cast-off* (e.g., when a *drip trail* is formed from movement of a bloody knife), or whether the categories are disjoint.

***Expiration***

The current ASB definition for *expiration* is: "A bloodstain pattern resulting from blood forced by airflow out of the nose, mouth, or a wound."

Potential updates/clarifications for this definition include:

- The standard should state explicitly that *expiration* is a subcategory of *spatter*.
- The standard should clearly differentiate *projected* and *expirated* blood—the latter including blood that has been expelled by, for example, vomiting or spitting.

***Flow***

The current ASB definition for *flow* is: "A bloodstain resulting from the movement of a volume of blood on a surface due to gravity or movement of the target."

Potential updates/clarifications for this definition include:

- No specific recommendations at this time.

***Forward spatter***

The current ASB definition for *forward spatter* is: "A bloodstain pattern resulting from blood drops which can be produced when a projectile creates an exit wound."

Potential updates/clarifications for this definition include:

- The phrase "which can be produced" can be interpreted as indeterminate (the pattern could be caused by a bullet entrance wound, rather than specifying that the pattern was definitively caused by a bullet entrance wound). This lack of clarity should be resolved by explicitly stating whether or not *forward spatter* can be generated by a process not involving a bullet entrance wound. Compare to note under *backspatter*.
- A more generalized approach for defining, describing and classifying projectile-related spatter should be developed (as discussed under *backspatter*).
- Consideration should be given to establishing validated methods for simulating forward spatter patterns.
- The standard should explicitly state that *forward spatter* is a subcategory of *impact*, which in turn is a subcategory of *spatter*.

***Impact***

The current ASB definition for *impact* is: “A bloodstain pattern resulting from an object striking liquid blood.”

Potential updates/clarifications for this definition include:

- The standard should clarify whether *impact* can also result from an object striking a blood-bearing material, such as a human body.
- The standard should clarify whether *drip pattern* is a subcategory of *impact*, or if the categories are disjoint. In other words, is a liquid considered an “object striking liquid blood?”
- The standard should explicitly state that *impact* is a subcategory of *spatter*.

***Insect stain***

The current ASB definition for *insect stain* is: “A bloodstain resulting from insect activity.”

Potential updates/clarifications for this definition include:

- The standard should clarify whether *insect stains* are subcategories of *transfer*, *alteration*, or both.
- The standard should clarify whether *insect stains* can ever be considered a subcategory of *spatter*, and if so, under what circumstances.

***Perimeter stain***

The current ASB definition for *perimeter stain* is: “An altered stain consisting of its edge characteristics, the central area having been partially or entirely removed.”

Potential updates/clarifications for this definition include:

- The standard should explicitly state that *perimeter stain* is a physical characteristic, not a mechanism of deposition, and clearly delineate the pattern classifications that could be inferred by this observation.

***Pool***

The current ASB definition for *pool* is: “A bloodstain resulting from an <u>accumulation</u> of liquid blood on a surface.”

Potential updates/clarifications for this definition include:

- The standard should clarify how much blood is needed to constitute an “accumulation”. For example: Do two drops of blood count as an accumulation? Does a single beaded drop (for instance, on a waxy surface) count?
- The standard should clarify whether a *pool* of blood continues to count as a *pool* after it dries.
- The standard should clearly differentiate *pool* from *saturation stain*.
- We propose the following rules: if the stain is fully absorbed, it is a *saturation stain*; if the stain is beaded or unabsorbed, it is a *pool*; if the stain is partially absorbed, it is both a *pool* and a *saturation stain*.
- The standard should explicitly state that *pool* is a physical characteristic, not a mechanism of deposition, and clearly delineate the pattern classifications that could be inferred by this observation.

***Projected***

The current ASB definition for *projected* is: "A bloodstain pattern resulting from the ejection of blood under hydraulic pressure, typically from a breach in the circulatory system."

Potential updates/clarifications for this definition include:

- Because all bloodstains ultimately result from a breach in the circulatory system, the example used to illustrate what is distinctive about the *projected* category should be more specific.
- Consideration should be given to establishing validated methods for simulating projected patterns.
- The standard should clearly differentiate *projected* and *expirated* blood—the latter including blood that has been expelled by, for example, vomiting or spitting.
- The standard should explicitly state that *projected* is a subcategory of *spatter*.

***Satellite stain***

The current ASB definition for *satellite stain* is: "A smaller bloodstain that originated during the formation of the parent stain as a result of blood impacting a surface."

Potential updates/clarifications for this definition include:

- The standard should clarify whether it is appropriate to report *satellite stains* when the parent stain is not present.
- The standard should clarify the circumstances under which a *satellite stain* may be considered an observation (e.g., when the parent stain is present) versus an inference of the mechanism of deposition (e.g., when the parent stain is not present).
- The standard should explicitly state that *satellite stain* is a subcategory of *spatter*.

***Saturation stain***

The current ASB definition for *saturation stain* is: "A bloodstain resulting from the accumulation of liquid blood in an absorbent material."

Potential updates/clarifications for this definition include:

- The standard should clarify how much blood is needed to constitute an "accumulation". For example: Do two drops of blood count as an accumulation? Does a single beaded drop (for instance, on a waxy surface) count?
- The standard should clearly differentiate *pool* from *saturation stain*.
- We propose the following rules: if the stain is fully absorbed, it is a *saturation stain*; if the stain is beaded or unabsorbed, it is a *pool*; if the stain is partially absorbed, it is both a *pool* and a *saturation stain*.
- The standard should explicitly state that *saturation stain* is a physical characteristic, not a mechanism of deposition, and clearly delineate the pattern classifications that could be inferred by this observation.

***Serum stain***

The current ASB definition for *serum stain* is: "The stain resulting from the liquid portion of blood (serum) that separates during coagulation."

Potential updates/clarifications for this definition include:

- No specific recommendations at this time.

***Spatter***

The current ASB definition for *spatter* is: "A bloodstain resulting from an airborne blood drop created when external force is applied to liquid blood."

Potential updates/clarifications for this definition include:

- No specific recommendations at this time.

***Splash***

The current ASB definition for *splash* is: "A bloodstain pattern created from a large volume of liquid blood falling onto a surface."

Potential updates/clarifications for this definition include:

- The standard should clarify how much blood is needed to constitute a "large volume".
- The standard should clarify what constitutes "falling". In particular, it should be clarified whether this implies that there is negligible lateral movement of the airborne liquid prior to deposition.
- We recommend that the standard consider removing the term "splash" altogether.

***Swipe***

The current ASB definition for *swipe* is: "A bloodstain resulting from the transfer of blood from a blood-bearing surface onto another surface, with characteristics that indicate relative motion between the two surfaces."

Potential updates/clarifications for this definition include:

- The standard should explicitly state that a *swipe* requires relative (side to side) motion; by contrast, pressing directly to the surface is not a *swipe*.
- The standard should clearly discuss the potential relationship between s*wipe* and *wipe*. For example, if a *wipe* transfers blood from the existing bloodstain onto a previously unstained area of a surface, the bloodstain is both a *wipe* and *swipe* (and thereby also both *altered* and *transfer*).
- The standard should explicitly state that *swipe* is a subcategory of *transfer stain*. Compare note under wipe.

***Transfer stain***

The current ASB definition for *transfer stain* is: “A bloodstain resulting from contact between a blood-bearing surface and another surface.”

Potential updates/clarifications for this definition include:

- Note that one can create *spatter* by roughly *swiping* (e.g. a drenched rag on a wall). The standard should clarify whether *spatter stains* that are generated in this way count as *transfer.*

***Void***

The current ASB definition for *void* is: “An absence of blood in an otherwise continuous bloodstain or bloodstain pattern.”

Potential updates/clarifications for this definition include:

- No specific recommendations at this time.

***Wipe***

The current ASB definition for *wipe* is: “An altered stain resulting from an object moving through a preexisting wet bloodstain.”

Potential updates/clarifications for this definition include:

- The standard should clearly discuss the potential relationship between s*wipe* and *wipe*. For example, if a *wipe* transfers blood from the existing bloodstain onto a previously unstained area of a surface, the bloodstain is both a *wipe* and *swipe* (and thereby also both *altered* and *transfer*).
- The standard should clarify what distinguishes *wipe* from *alteration* alone. In particular, it should be clarified whether *wipe* requires relative side-to-side motion, as *swipe* does. By contrast: if a person merely pushes and releases a bloody cloth, is the resulting stain a wipe stain?
- The standard should explicitly state that *wipe* is a subcategory of *altered*.

## Acknowledgements

We thank the BPA analysts who participated in this study, as well as Kyle Dalrymple for software development; Niki Osborne, Nikki Blackwell, and LeeAnn Singley for their insights during study design; and Andy Maloney for assistance in the comparison of terminology. The opinions, findings, and conclusions or recommendations expressed in this program are those of the authors and do not necessarily reflect those of the Department of Justice.

## References

[1] R. A. Hicklin *et al.*, "Accuracy and Reproducibility of Conclusions by Forensic Bloodstain Pattern Analysts," *(submitted)*, 2021.

[2] T. Bevel and R. M. Gardner, *Bloodstain Pattern Analysis with an Introduction to Crime Scene Reconstruction*. Boca Raton: CRC Press, 2008.

[3] S. H. James, P. E. Kish, and T. P. Sutton, *Principles of Bloodstain Pattern Analysis: Theory and Practice*. Boca Raton: CRC Press, 2005.

[4] H. Macdonell, "Bloodstain Pattern Interpretation," in *Wiley Encyclopedia of Forensic Science*, Chichester, UK: John Wiley & Sons, Ltd, 2009, pp. 359–395.

[5] H. L. MacDonell, *Bloodstain pattern interpretation*. Corning: Laboratory of Forensic Science, 1982.

[6] *David R. Camm v. State of Indiana (812 N.E.2d 1127, Court of Appeals of Indiana)*, vol. 22A01-0208. 2004.

[7] *David R. Camm v State of Indiana (908 N.E.2d 215, Supreme Court of Indiana)*, vol. 87S00-0612. 2009.

[8] *David R. Camm v. State of Indiana (957 N.E.2d 205, Court of Appeals of Indiana)*, vol. 87A01-1102. 2011.

[9] *State of Maine v. Knight (43 Me. 11, Supreme Judicial Court of Maine)*. 1857.

[10] National Research Council, *Strengthening Forensic Science in the United States: A Path Forward*. Washington, D.C.: The National Academies Press, 2009.

[11] 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*. Washington, D.C.: Executive Office of the President, 2016.

[12] President's Council of Advisors on Science and Technology, *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*. Washington, D.C.: Executive Office of the President, 2017.

[13] J. L. Mnookin, "Of Black Boxes, Instruments, and Experts: Testing the Validity of Forensic Science," *Episteme*, vol. 5, no. 3, pp. 343–358, Oct. 2008, doi: 10.3366/E1742360008000440.

[14] T. Laber, P. Kish, M. Taylor, G. Owens, N. Osborne, and J. Curran, "Reliability Assessment of Current Methods in Bloodstain Pattern Analysis (NIJ/NCJRS Report

247180),” 2014. [Online]. Available: https://www.ncjrs.gov/pdffiles1/nij/grants/247180.pdf.

[15] B. N. Meneses, B. J. Gestring, P. E. Kish, and B. J. Gestring, “A Preliminary Study: Evaluating Error Rate Associated with Bloodstain Pattern Analysis,” *61st Annual Meeting of the American Academy of Forensic Sciences*. 2009.

[16] M. C. Taylor, T. L. Laber, P. E. Kish, G. Owens, and N. K. P. Osborne, “The Reliability of Pattern Classification in Bloodstain Pattern Analysis, Part 1: Bloodstain Patterns on Rigid Non-absorbent Surfaces,” *J. Forensic Sci.*, vol. 61, no. 4, pp. 922–927, Jul. 2016, doi: 10.1111/1556-4029.13091.

[17] M. C. Taylor *et al.*, “The Reliability of Pattern Classification in Bloodstain Pattern Analysis - PART 2: Bloodstain Patterns on Fabric Surfaces ,” *J. Forensic Sci.*, vol. 61, no. 6, pp. 1461–1466, Nov. 2016, doi: 10.1111/1556-4029.13191.

[18] N. K. P. Osborne, M. C. Taylor, M. Healey, and R. Zajac, “Bloodstain pattern classification: Accuracy, effect of contextual information and the role of analyst characteristics,” *Sci. Justice*, vol. 56, no. 2, pp. 123–128, Mar. 2016, doi: 10.1016/j.scijus.2015.12.005.

[19] S. K. Y. Yuen, M. C. Taylor, G. Owens, and D. A. Elliot, “The Reliability of Swipe/Wipe Classification and Directionality Determination Methods in Bloodstain Pattern Analysis,” *J. Forensic Sci.*, vol. 62, no. 4, pp. 1037–1042, Jul. 2017, doi: 10.1111/1556-4029.13298.

[20] AAFS Standards Board, “ ASB Technical Report 033: Terms and Definitions in Bloodstain Pattern Analysis.” 2017.

[21] International Association for Identification (IAI), “Bloodstain Pattern Analyst Certification,” 2020. https://www.theiai.org/bloodstain_pattern.php (accessed Dec. 10, 2020).

[22] International Association of Bloodstain Pattern Analysts, “Published Standards Related to BPA,” *J. Bloodstain Pattern Anal.*, vol. 35, no. 3, p. 30, 2020, [Online]. Available: https://iabpa.org/docs/JBPA_Vol_35_No_3_September_2020_v1.pdf.

[23] Forensic Testing Program, “Bloodstain Pattern Analysis Test No. 18-5601/2/5 Summary Report,” Sterling, Virginia, 2018. Accessed: Dec. 10, 2020. [Online]. Available: https://cts-forensics.com/reports/38601_Web.pdf.

[24] Collaborative Testing Services Forensic Testing Program, “Footwear Imprint Evidence Test No. 20-5331/5 Summary Report,” 2020.

[25] Noblis, “Bloodstain Pattern Analysis Black Box Study Dataset (V1, Dec 2020),” 2020. https://osf.io/2ckhw/ (accessed Dec. 10, 2020).

[26] P. M. Comiskey, A. L. Yarin, and D. Attinger, “Theoretical and experimental investigation of forward spatter of blood from a gunshot,” *Phys. Rev. Fluids*, vol. 3, no. 063901, 2018.

[27] OSAC Human Factors Committee, “Draft Guidance on Testing the Performance of Forensic Examiners,” 2018. [Online]. Available: https://www.nist.gov/system/files/documents/2018/05/21/draft_hfc_guidance_document-may_8.pdf.

[28] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Accuracy and reliability of forensic latent fingerprint decisions," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 19, 2011, doi: 10.1073/pnas.1018707108.

[29] H. Tobi, P. B. van den Berg, and L. T. de Jong-van den Berg, "Small proportions: what to report for confidence intervals?," *Pharmacoepidemiol. Drug Saf.*, vol. 14, no. 4, Apr. 2005, doi: 10.1002/pds.1081.

[30] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Repeatability and reproducibility of decisions by latent fingerprint examiners," *PLoS One*, vol. 7, no. 3, 2012, doi: 10.1371/journal.pone.0032800.

[31] K. L. Gwet, *Handbook of Inter-Rater Reliability*, 4th ed. Gaithersburg, MD: Advanced Analytics, LLC., 2014.

[32] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, Mar. 1977, doi: 10.2307/2529310.

[33] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[34] J. Hair, W. Black, B. Babin, R. Anderson, and R. Tatham, *Multivariate Data Analysis*, 7th ed. Upper Saddle River: Prentice Hall, 2014.

[35] U. Grömping, "Variable Importance Assessment in Regression: Linear Regression versus Random Forest," *Am. Stat.*, vol. 63, no. 4, pp. 308–319, Nov. 2009, doi: 10.1198/tast.2009.08199.

[36] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010, doi: 10.1016/j.patrec.2010.03.014.

[37] A. Liaw and M. Wiener, "Classification and Regression by RandomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[38] S. Schmidt, "Encyclopedia of Research Design," Thousand Oaks: SAGE Publications, Inc., 2010.

[39] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Ser. B*, vol. 57, pp. 289–300, 1995.

[40] W. S. Noble, "How does multiple testing correction work?," *Nat. Biotechnol.*, vol. 27, no. 12, pp. 1135–1137, Dec. 2009, doi: 10.1038/nbt1209-1135.

[41] N. K. P. Osborne, M. C. Taylor, and R. Zajac, "Exploring the role of contextual information in bloodstain pattern analysis: A qualitative approach," *Forensic Sci. Int.*, vol. 260, Mar. 2016, doi: 10.1016/j.forsciint.2015.12.039.

[42] N. K. P. Osborne and M. C. Taylor, "Contextual information management: An example of independent-checking in the review of laboratory-based bloodstain pattern analysis," *Sci. Justice*, vol. 58, no. 3, May 2018, doi: 10.1016/j.scijus.2018.01.001.

[43] A. Wonder, *Bloodstain Pattern Evidence: Objective Approaches and Case Applications*. Academic Press, 2007.

[44] P. Esperança, "Bloodstain Pattern Identification Key," 2014. [Online]. Available: http://l-a-c.expert/wp-content/uploads/2019/01/Bloodstain_Pattern_Identification_Key_Philippe_Esperanca.pdf.

[45] International Association of Bloodstain Pattern Analysts (IABPA), "Suggested IABPA Terminology List," *Int. Assoc. Bloodstain Pattern Anal. Newsl.*, vol. 12, no. 4, pp. 15–17, 1996, [Online]. Available: https://static1.squarespace.com/static/543841fce4b0299b22e1956a/t/54be8e3ee4b0121e42a6c0d7/1421774398257/IABPA+Terminology+1996.pdf.

[46] International Association of Bloodstain Pattern Analysts (IABPA), "Suggested IABPA Terminology List," 2004, [Online]. Available: https://static1.squarespace.com/static/543841fce4b0299b22e1956a/t/54be8822e4b06fad9ba9d473/1421772834653/BPATerminology.pdf.

[47] International Association of Bloodstain Pattern Analysts (IABPA), "Recommended Training Standards/Guidelines for a Basic BPA Course," 2021. https://www.iabpa.org/iabpa_training_course_standard.php.

[48] Scientific Working Group on Bloodstain Pattern Analysis (SWGSTAIN), "Scientific Working Group on Bloodstain Pattern Analysis: Recommended Terminology," *Forensic Sci. Commun.*, vol. 11, no. 2, 2009, [Online]. Available: https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/april2009/standards/2009_04_standards01.htm.

[49] HemoSpat, "Bloodstain Pattern Analysis Terminology," 2021. https://hemospat.com/bloodstain-pattern-analysis-terminology/.

# Black Box Evaluation of Bloodstain Pattern Analysis Conclusions
# Final Research Report
# Appendices

## *Appendix A Participant Instructions and Frequently Asked Questions (FAQs)*

Participants were provided detailed instructions. During the testing period, any communications with participants (other than resolving software access issues) were provided to all participants in a Frequently Asked Questions (FAQs) file. No changes to the instructions were made after the start of the study period. The instructions, FAQs, and the ASB recommended BPA terms and definitions were emailed to interested parties and were available via the study website. The participation instructions and FAQs are included with this Report as Supplemental Data S1.

## *Appendix B Participant Background Survey*

All participants completed the background survey. Percentages in the summary below are based on 75 participants. Percentages may not total 100% due to rounding.

Survey result details (anonymized) are included as a spreadsheet in Supplemental Data S3.

| | | |
|---|---|---|
| **1. Select each degree completed and the field(s) associated with each degree.** | | |
| **Associate's Degree (Arts or Science) Major** | | |
| Criminology/Criminal Justice | 3 | 4% |
| Computer Science | 1 | 1% |
| Forensic Science | 3 | 4% |
| Natural Sciences | 3 | 4% |
| Other majors | 2 | 3% |
| Other scientific or technical field | 1 | 1% |
| **Bachelor of Arts Major** | | |
| (No response) | 1 | 1% |
| Criminology/Criminal Justice | 2 | 3% |
| Forensic Science | 1 | 1% |
| Natural Sciences | 2 | 3% |
| Other majors | 2 | 3% |
| **Bachelor of Science Major** | | |
| Criminology/Criminal Justice | 11 | 15% |
| Engineering | 2 | 3% |
| Forensic Science | 10 | 13% |
| Natural Sciences | 24 | 32% |
| Other majors | 3 | 4% |
| Other scientific or technical field | 1 | 1% |
| **Master of Arts Major** | | |
| Criminology/Criminal Justice | 1 | 1% |
| Forensic Science | 3 | 4% |
| Other majors | 1 | 1% |
| **Master of Science Major** | | |
| Criminology/Criminal Justice | 2 | 3% |
| Engineering | 1 | 1% |
| Forensic Science | 15 | 20% |
| Natural Sciences | 7 | 9% |
| Other majors | 2 | 3% |
| **PhD Major** | | |
| Engineering | 1 | 1% |
| Forensic Science | 1 | 1% |
| Natural Sciences | 3 | 4% |
| **Highest degree attained** | | |
| Associate's Degree (Arts or Science) | 6 | 8% |
| Bachelor of Arts | 4 | 5% |
| Bachelor of Science | 29 | 39% |
| Master of Arts | 4 | 5% |
| Master of Science | 24 | 32% |
| No degree | 3 | 4% |
| PhD | 5 | 7% |
| *For question 1, the total of majors by degree is greater than 75 participants and 100% due to multiple degrees.* | | |
| **2. Select the following statement which best describes your employment relevant to bloodstain pattern analysis.** | | |
| I currently conduct bloodstain pattern analysis as a part of my employment | 72 | 96% |
| I'm not currently conducting bloodstain pattern analysis as a part of my employment, but I have been within the last 2 years. | 3 | 4% |
| **3. After completing bloodstain pattern analysis training, provide the number of years of experience you have as a bloodstain pattern analyst.** | | |
| 0-1 years | 2 | 3% |
| 1-4 years | 18 | 24% |
| 5-10 years | 29 | 39% |
| 11-15 years | 11 | 15% |
| 16+ years | 15 | 20% |
| **4. Select the statement which best describes the frequency with which you perform bloodstain pattern analysis within your case work?** | | |
| 0 or 1 bloodstain pattern analysis cases per year | 8 | 11% |
| 2-4 bloodstain pattern analysis cases per year | 27 | 36% |
| 5-10 blood pattern analysis cases per year | 18 | 24% |
| 11-20 blood pattern analysis cases per year | 14 | 19% |
| 21-50 blood pattern analysis cases per year | 5 | 7% |
| Over 50 blood pattern analysis cases per year | 3 | 4% |

**5. Have you ever testified in court as an expert in bloodstain pattern analysis?**

| | | |
|---|---|---|
| Yes: More than 20 times | 6 | 8% |
| Yes: 10-20 times | 14 | 19% |
| Yes: less than 10 times | 42 | 56% |
| No, never | 13 | 17% |

**6. Have you ever testified as an expert in a bloodstain pattern analysis admissibility hearing?**

| | | |
|---|---|---|
| Yes: 5-10 times | 1 | 1% |
| Yes: less than 5 times | 18 | 24% |
| No, never | 56 | 75% |

**7. Please indicate the approximate percentage of your blood pattern analysis cases for the following (must add to 100%)**

A. Attend actual crime scenes to conduct bloodstain pattern analysis
B. Examine items of physical evidence and scene photographs (without actually attending the scene)
C. Conduct examination of scene photographs (without actually attending the scene or examining items of physical evidence)
D. Examine items of physical evidence (without actually attending the scene or examining scene photographs)

*Figure 21shows the responses to question 7 graphically.*
*Notes:*

A. *Crime scene: mean=50.7%. 2 participants replied 100%, 2 participants replied 0%, 40 participants replied 50% or more.*
B. *Physical evidence & photos: mean=21.2%. No participants replied 100%, 6 participants replied 0%, 8 participants replied 50% or more.*
C. *Photos only: mean=14.2%. No participants replied 100%, 12 participants replied 0%, 5 participants replied 50% or more.*
D. *Physical evidence only: mean=13.9%. No participants replied 100%, 23 participants replied 0%, 7 participants replied 50% or more.*

- *Photos (B & C combined): mean=35.4%. No participants replied 100%, 3 participants replied 0%, 14 participants replied 10% or less; 28 participants replied 50% or more.*
- *Note that response A (attending crime scenes) does not preclude evaluating photographs.*

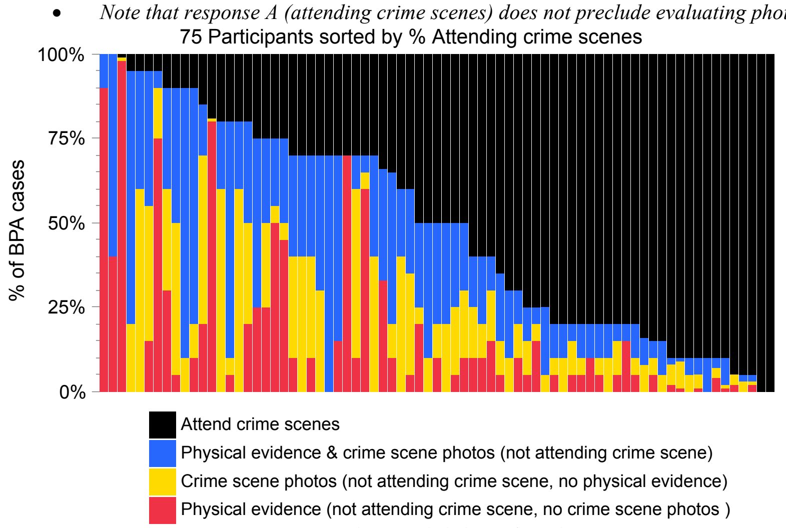


*Figure 21. Participants' BPA cases by availability of evidence. (Survey question #7)*

**8. What other forensic examinations are you currently qualified to perform or have been qualified to perform in the past? (Check all that apply)**

| | | |
|---|---|---|
| Breath alcohol | 5 | 7% |
| Chemistry (Seized drugs) | 7 | 9% |
| Crime scene investigation | 62 | 83% |
| Crime scene reconstruction | 41 | 55% |
| Digital/Multimedia | 4 | 5% |
| DNA | 17 | 23% |
| Fire and explosive investigations (on scene) | 7 | 9% |
| Fire debris and explosives (in lab) | 3 | 4% |
| Firearms/Toolmarks | 9 | 12% |
| Footwear/Tire impressions | 15 | 20% |
| Latent prints (friction ridge) | 24 | 32% |
| Medicolegal death investigation | 5 | 7% |
| Questioned documents | 2 | 3% |
| Serology (body fluid identification) | 23 | 31% |
| Shooting incident reconstruction | 20 | 27% |
| Toxicology | 3 | 4% |
| Trace evidence | 16 | 21% |

| | | |
|---|---|---|
| Traffic accident investigation and reconstruction (on scene) | 8 | 11% |
| Other | 4 | 5% |
| *For question 8, participants indicated 0-10 responses each (mean 3.6, median 4).* | | |
| *Frequent combinations:* | | |
| *Crime scene investigation and Crime scene reconstruction* | 40 | 53% |
| *Crime scene investigation and Latent prints* | 23 | 31% |
| *Crime scene investigation and Shooting incident reconstruction* | 18 | 24% |
| *Crime scene investigation and Serology (body fluid identification)* | 17 | 23% |
| *Crime scene investigation and Crime scene reconstruction and Latent prints* | 14 | 19% |
| *Crime scene investigation and DNA* | 12 | 16% |
| *Serology (body fluid identification) and DNA* | 11 | 15% |
| **9. Select the statement which best describes the most comprehensive blood pattern analysis training you have received.** | | |
| I completed a formal program of instruction / supervision for less than 6 months. | 10 | 13% |
| I completed a formal program of instruction / supervision for 6-12 months. | 9 | 12% |
| I completed a formal program of instruction / supervision for 1 year or more. | 16 | 21% |
| I received informal, on-the-job training. | 7 | 9% |
| I attended/completed courses and/or workshops. | 33 | 44% |
| **10. Select the statement(s) that describe the components of your blood pattern analysis training. (Select all that apply)** | | |
| Basic understanding of DNA reports | 55 | 73% |
| Basic understanding of medical reports and autopsy reports | 62 | 83% |
| Completed exercises or experiments as part of a formal training program | 62 | 83% |
| Conclusion writing | 67 | 89% |
| Creating bloodstain patterns | 73 | 97% |
| Enhancement of bloodstain patterns with chemicals and with various wavelengths of light | 58 | 77% |
| Examination of bloodstained clothing instruction | 65 | 87% |
| Limitations of bloodstain pattern analysis | 65 | 87% |
| Practical test of your knowledge, skills and abilities | 66 | 88% |
| Practiced testimony in a simulated court | 39 | 52% |
| Presumptive testing for blood | 71 | 95% |
| Presumptive testing for saliva | 34 | 45% |
| Processed crimes scenes for bloodstain patterns with a mentor | 55 | 73% |
| Proper documentation of bloodstain patterns at crime scenes | 72 | 96% |
| Read scientific articles on bloodstain pattern analysis | 70 | 93% |
| Report writing | 73 | 97% |
| Roadmapping | 41 | 55% |
| Terminology utilized in bloodstain pattern analysis | 74 | 99% |
| Wording conclusions, such as stating opinions about the mechanism(s) by which a pattern was created. | 68 | 91% |
| Worked with a mentor who guided the training | 46 | 61% |
| Written test of your knowledge and skills | 57 | 76% |
| *For question 10, participants indicated 8-21 responses each (mean 17.0, median 17).* | | |
| **11. For each of the following, estimate the total time spent over your career (for each type of training).** | | |
| **Creating bloodstain patterns** | | |
| None | 2 | 3% |
| 1-8 hours | 11 | 15% |
| 9-40 hours | 21 | 28% |
| 41-80 hours | 16 | 21% |
| 81-160 hours | 14 | 19% |
| Over 160 hours | 11 | 15% |
| **Lectures or continuing education specific to blood pattern analysis** | | |
| None | 2 | 3% |
| 1-8 hours | 6 | 8% |
| 9-40 hours | 10 | 13% |
| 41-80 hours | 17 | 23% |
| 81-160 hours | 23 | 31% |
| Over 160 hours | 17 | 23% |
| **Workshop: Basic blood pattern analysis** | | |
| None | 0 | 0% |
| 1-8 hours | 2 | 3% |
| 9-40 hours | 42 | 56% |
| 41-80 hours | 19 | 25% |
| 81-160 hours | 7 | 9% |
| Over 160 hours | 5 | 7% |
| **Workshop: Advanced blood pattern analysis** | | |
| None | 10 | 13% |
| 1-8 hours | 5 | 7% |

| | | |
|---|---|---|
| 9-40 hours | 35 | 47% |
| 41-80 hours | 16 | 21% |
| 81-160 hours | 6 | 8% |
| Over 160 hours | 3 | 4% |
| **Workshop: Math and physics as they relate to bloodstain pattern analysis** | | |
| None | 13 | 17% |
| 1-8 hours | 18 | 24% |
| 9-40 hours | 32 | 43% |
| 41-80 hours | 8 | 11% |
| 81-160 hours | 2 | 3% |
| Over 160 hours | 2 | 3% |
| **Workshop: Fluid dynamics of bloodstain pattern formation** | | |
| None | 23 | 31% |
| 1-8 hours | 20 | 27% |
| 9-40 hours | 25 | 33% |
| 41-80 hours | 5 | 7% |
| 81-160 hours | 1 | 1% |
| Over 160 hours | 1 | 1% |
| **Workshop: Examination of bloodstains on fabrics** | | |
| None | 10 | 13% |
| 1-8 hours | 22 | 29% |
| 9-40 hours | 23 | 31% |
| 41-80 hours | 14 | 19% |
| 81-160 hours | 5 | 7% |
| Over 160 hours | 1 | 1% |
| **12. How long did you do supervised bloodstain pattern analysis casework before performing bloodstain pattern analysis on your own?** | | |
| I did not do supervised bloodstain pattern analysis casework | 23 | 31% |
| One month or less | 8 | 11% |
| Two to six months | 14 | 19% |
| Six to twelve months | 17 | 23% |
| More than one year | 13 | 17% |
| **13. If you have been certified as a bloodstain pattern analyst, select the certifying body. Check all that apply.** | | |
| International Association for Identification (IAI) | 11 | 15% |
| Other | 11 | 15% |
| (No response) | 53 | 71% |
| *Three of the "Other" entries for question 13 indicated "Royal Canadian Mounted Police (RCMP)" (4%).* | | |
| **14. When did you last take a proficiency test in bloodstain pattern analysis?** | | |
| Within the past year | 45 | 60% |
| Within the past 2 years | 8 | 11% |
| Within the past 5 years | 4 | 5% |
| More than 5 years | 2 | 3% |
| Never | 16 | 21% |
| **15. Who prepared your most recent proficiency test? (Check all that apply)** | | |
| Employer (internal) | 13 | 17% |
| Collaborative Testing Services (CTS) | 39 | 52% |
| Forensic Assurance (FA) | 7 | 9% |
| Other proficiency test provider | 11 | 15% |
| (No response) | 17 | 23% |
| *Four of the "Other" entries for question 15 indicated "European Network of Forensic Science Institutes (ENFSI)" (5%). 10 participants made 2 selections and 1 participant made 3 selections.* | | |
| **16. Has it ever been brought to your attention that you failed a proficiency test in bloodstain pattern analysis?** | | |
| Yes | 3 | 4% |
| No | 72 | 96% |
| **17. Select the category that best describes the employer for whom you have done most of your bloodstain pattern analysis work:** | | |
| Local, city or county agency | 22 | 29% |
| Regional, state or provincial agency | 30 | 40% |
| National or federal agency | 16 | 21% |
| Independent consultant | 4 | 5% |
| Academic institution | 3 | 4% |
| **18. Select the country in which you performed most of your blood pattern analysis work:** | | |
| Brazil | 3 | 4% |
| Canada | 6 | 8% |

| | | |
|---|---|---|
| Finland | 3 | 4% |
| Italy | 6 | 8% |
| Sweden | 3 | 4% |
| United States | 43 | 57% |
| (Other) | 11 | 15% |
| *Eight countries (Argentina, Belgium, Denmark, Netherlands, Serbia, South Africa, United Arab Emirates, United Kingdom) had a total of 11 participants, grouped here as "Other." 23 participants (31%) were from countries where English is not an official language.* | | |
| **19. Are you in law enforcement?** | | |
| Yes: I am sworn law enforcement (i.e., uniformed) | 27 | 36% |
| Yes: I am a civilian working for a law enforcement agency | 36 | 48% |
| No | 12 | 16% |
| **20. Do you work in a laboratory setting?** | | |
| Yes | 50 | 67% |
| No | 25 | 33% |
| **21. If your current employer is accredited in a category of testing that includes bloodstain pattern analysis, what accreditation standard is held by your current employer? (Check all that apply. Leave blank if not accredited)** | | |
| ISO 17025 | 27 | 36% |
| ISO 17020 | 2 | 3% |
| Both ISO 17020 and 17025 | 3 | 4% |
| (No response) | 43 | 57% |
| **22. Other than yourself, how many blood pattern analysts work at your current employer?** | | |
| None | 20 | 27% |
| 1 other BPA analyst | 7 | 9% |
| 2-3 other BPA analysts | 25 | 33% |
| 4 or more other BPA analysts | 23 | 31% |
| **23. Please check all of the following that apply for your current employer's bloodstain pattern analysis policies and procedures: (check all that apply)** | | |
| A: My current employer does not have policies and procedures that apply to bloodstain pattern analysis<br>B: My current employer has written bloodstain pattern analysis policies and procedures<br>C: My current employer has written policies and procedures that provide guidance on how to write reports<br>D: My current employer has written policies and procedures that provide guidance about how to state opinions in reports about the mechanism(s) by which a pattern was created | | |
| (no response) | 3 | 4% |
| A (No BPA policies & procedures) | 24 | 32% |
| A,B (No BPA policies & procedures, written BPA policies & procedures; see note below) | 1 | 1% |
| A,C (No BPA policies & procedures, written guidance on writing reports; see note below) | 1 | 1% |
| B (Written BPA policies & procedures) | 16 | 21% |
| B,C (Written BPA policies & procedures, written guidance on writing reports) | 9 | 12% |
| B,C,D (Written BPA policies & procedures, written guidance on writing reports, written guidance on stating opinions) | 18 | 24% |
| C (Written guidance on writing reports) | 2 | 3% |
| D (Written guidance on stating opinions) | 1 | 1% |
| *For question 23, two of the responses (A,B and A,C) were self-contradictory; for analyses, we treat these as having written policies.* | | |
| **24. Are your bloodstain pattern analysis case conclusions / reports technically reviewed?** | | |
| Yes, they are reviewed by a trained bloodstain pattern analyst | 62 | 83% |
| Yes, they are reviewed, but not by a trained bloodstain pattern analyst | 6 | 8% |
| No | 7 | 9% |
| **25. In about what percentage of your blood pattern analysis cases are your conclusions / reports technically reviewed?** | | |
| 100% | 55 | 73% |
| 75-99% | 5 | 7% |
| 50-74% | 3 | 4% |
| 25-49% | 3 | 4% |
| Less than 25% | 2 | 3% |
| Never: my reports are not subjected to technical review | 7 | 9% |
| **26. If given the following case materials in what order would you review these materials when conducting bloodstain pattern analysis? (Number 1-4 to indicate order)** | | |
| Review Police Reports and Witness Statements<br>Review Autopsy and Medical Reports<br>Review Scene Photographs<br>Review DNA / Serology Reports | | |
| 1) Autopsy/Medical 2) Scene photos 3) DNA/Serology 4) Police/Witness | 2 | 3% |

| | | |
|---|---|---|
| 1) Autopsy/Medical 2) Scene photos 3) Police/Witness 4) DNA/Serology | 1 | 1% |
| 1) DNA/Serology 2) Autopsy/Medical 3) Scene photos 4) Police/Witness | 1 | 1% |
| 1) DNA/Serology 2) Scene photos 3) Autopsy/Medical 4) Police/Witness | 1 | 1% |
| 1) Police/Witness 2) Autopsy/Medical 3) DNA/Serology 4) Scene photos | 1 | 1% |
| 1) Police/Witness 2) Scene photos 3) Autopsy/Medical 4) DNA/Serology | 8 | 11% |
| 1) Police/Witness 2) Scene photos 3) DNA/Serology 4) Autopsy/Medical | 1 | 1% |
| 1) Scene photos 2) Autopsy/Medical 3) DNA/Serology 4) Police/Witness | 21 | 28% |
| 1) Scene photos 2) Autopsy/Medical 3) Police/Witness 4) DNA/Serology | 5 | 7% |
| 1) Scene photos 2) DNA/Serology 3) Autopsy/Medical 4) Police/Witness | 19 | 25% |
| 1) Scene photos 2) DNA/Serology 3) Police/Witness 4) Autopsy/Medical | 1 | 1% |
| 1) Scene photos 2) Police/Witness 3) Autopsy/Medical 4) DNA/Serology | 14 | 19% |
| Scene Photos before Police Reports and Witness Statements | 65 | 87% |
| Police Reports and Witness Statements before Scene Photos | 10 | 13% |

## Appendix C Post-study Survey Results

The results of the post-study survey are included verbatim below.

Survey result details (anonymized) are included as a spreadsheet in Supplemental Data S3.

| **Considering all 150 samples in this study, *how do the patterns in this study* compare in difficulty to your typical casework?** | | |
|---|---|---|
| Much easier than casework | 0 | 0% |
| Easier than casework | 2 | 4% |
| Similar to casework | 34 | 72% |
| Harder than casework | 9 | 19% |
| Much harder than casework | 2 | 4% |

**Do you have any comments about the overall test you'd like to provide the team?**

- Bloodstains and patterns excellent for case study
- Case information about the type of offense, the alleged facts, and maybe a few more photographs to give better context. In addition at times the website was slow and would time out and not submit your responses until you closed out and reopened the page.
- Due to my location outside the US, I have very little casework experience with gunshot spatter so my results may be skewed in that regard.
- Excellent study. I really enjoyed participating. Like casework there were some easy stains and some hard patterns. A lot of the scenario based questions seemed very far fetched.
- Good test but unlike casework we are only provided images to base our decisions on. I think we all agree that the methodology used in BPA is more than just looking at patterns, although some can be straight forward but looking at the total data collected.
- great experience, useful and well done. some easy pattern, and some very challenging. thanks for the opportunity.
- Great work. Thanks for putting this out there.
- I appreciate all the work going into this study. Some of the questions were badly worded: it was not possible to answer without further explanation. That may have been part of the study but it was frustrating.
- I felt this was a very good study and at the very least a great exposure to multiple kinds of bloodstains. I felt it really made me think more methodical and to stay more conservative with my conclusions when I don't have all the information.
- I know the "included" term was discussed on the video and at the IABPA conference, but there were some stains that were just "not classifiable" overall, it was a great exercise
- I loved the test! This made me study more and more about bloodstain pattern and consequently learn more. I think your team contributed much more with my improvement as bloodstain analyst than I did to the study.
- I really appreciate the study and is so important that more studies like these should be carried out.
- In some cases it was quite difficult to choose appropriate choice because saples was taken out of the "crime scene". And quality of some phothos was very poor!
- In some cases, without further information such as the coroner's report, it was difficult to correctly classify the bloodstains proposed.
- It was a great study! Best lucks for all of you! Thanks for let me participating

- It's not easy to analyze BPA on pictures because you can't see all the details. Also, with pictures, you're always thinking further... and the possibility of a pattern being under another one, so you can't always exclude...which means a lot of possible!
- Overall, I thought this was an excellent test. There were a few things which would could have improved it. For the pattern description parts, there was a character limit which did not allow me to be as detailed as I wanted with the descriptions.
- Quality of some of the pictures was poor. More overall pictures would've been good, sometimes you only saw a part of a pattern. Marking (A, B...) was poor. In casework we very seldom make conclusions from pictures only, go to crime scenes.
- Several examples stretched the semantic interpretation of definitions. I will be very interested to see the results of your study.
- Some of the pictures didn't load when I opened them in a new page. It didn't happen to many, but for whatever reason some pictures were blocked by the virus protection on the computer.
- Some of the questions I felt I couldn't answer due to subjective words, like "prolonged," where that can vary from person to person. It would have been nice to be able to add comments for each sample, to explain why I answered the way I did.
- Some photos were not taken with resolution high enough. Some patterns in photos were difficult as you can't see the "whole picture" as is the situation on crime scenes.
- Sometimes, I felt like I was answering "possible" or "included" way too many time. I felt like I had to cover all the bases, and I might not do that as much in casework because I have more of the back story (i.e. was this a gunshot wound or a beating)
- The actual case photos were of course similar to real casework.The tweets allowed for more analysis and was closer to casework, but more space is needed for the tweets,especially if there are multiple bloodstain patterns.Patterns on carpet was challenging.
- The amount of time it took to open samples and submit results was very frustrating.
- The difficulty in some of these samples is not having autopsy info or scene context in partial patterns. Also, I was not sure if "impact" was to be included as part of a gunshot type pattern as the "correct response."
- The test appear to be very interesting. Thank you.
- The testing was conveniently set up so it was easy to complete when time allowed.
- The time to take the test was grossly underestimated. The parts where we wrote descriptive comments took too much time and not enough space. Without having case info about injuries and seeing whole scenes, I was limited in opinions.
- There were areas (several in fact) where what i was being asked to comment on wasn't clear. A pattern would be labeled A and I would be asked to identify it, but there were multiple mechanisms applicable, so it was not clear. Some photographs not clear.
- Too many similar tests. Too long, too many patterns. "Possible" and "included" are too wide categories. Some solved examples could have specified better the meaning of these (and the other) categories.
- While I was doing the study I had problem with projected mechanism, becouse I think that You use this term only for Projected mechanism of arterial bleeding, but during my education we used that mechanism also for expirated blood, cast-off and cessation

# *Appendix D Glossary/Acronyms*

This section defines terms and acronyms as they are used in this report.

## Appendix D.1 ASB Terminology

The following definitions are taken verbatim from the OSAC/ASB terminology standard [20]. The classification prompts (2.3.1) were limited to these 26 terms.

| | |
|---|---|
| Altered | A bloodstain with characteristics that indicate a physical change has occurred. |
| Backspatter | A bloodstain pattern resulting from blood drops which can be produced when a projectile creates an entrance wound. |
| Blood clot | A gelatinous mass formed by a complex mechanism involving red blood cells, fibrinogen, platelets, and other clotting factors. |
| Bubble ring | An outline within a bloodstain resulting from air in the blood. |

| | |
|---|---|
| Cast-off | A bloodstain pattern resulting from blood drops released from an object due to its motion. |
| Cessation | A bloodstain pattern resulting from blood drops released from an object due to its abrupt deceleration. |
| Drip pattern | A bloodstain pattern resulting from a liquid that dripped into another liquid, at least one of which was blood. |
| Drip stain | A bloodstain resulting from a falling drop that formed due to gravity. |
| Drip trail | A bloodstain pattern resulting from the movement of a source of drip stains between two points. |
| Expiration | A bloodstain pattern resulting from blood forced by airflow out of the nose, mouth, or a wound. |
| Flow | A bloodstain resulting from the movement of a volume of blood on a surface due to gravity or movement of the target. |
| Forward spatter | A bloodstain pattern resulting from blood drops which can be produced when a projectile creates an exit wound. |
| Impact | A bloodstain pattern resulting from an object striking liquid blood. |
| Insect stain | A bloodstain resulting from insect activity. |
| Perimeter stain | An altered stain consisting of its edge characteristics, the central area having been partially or entirely removed. |
| Pool | A bloodstain resulting from an accumulation of liquid blood on a surface. |
| Projected | A bloodstain pattern resulting from the ejection of blood under hydraulic pressure, typically from a breach in the circulatory system. |
| Satellite stain | A smaller bloodstain that originated during the formation of the parent stain as a result of blood impacting a surface. |
| Saturation stain | A bloodstain resulting from the accumulation of liquid blood in an absorbent material. |
| Serum stain | The stain resulting from the liquid portion of blood (serum) that separates during coagulation. |
| Spatter | A bloodstain resulting from an airborne blood drop created when external force is applied to liquid blood. |
| Splash | A bloodstain pattern created from a large volume of liquid blood falling onto a surface. |
| Swipe | A bloodstain resulting from the transfer of blood from a blood-bearing surface onto another surface, with characteristics that indicate relative motion between the two surfaces. |
| Transfer | A bloodstain resulting from contact between a blood-bearing surface and another surface. |
| Void | An absence of blood in an otherwise continuous bloodstain or bloodstain pattern. |

| | |
|---|---|
| Wipe | An altered stain resulting from an object moving through a preexisting wet bloodstain. |

### Appendix D.2 Terminology as Used in this Report

| | |
|---|---|
| Accuracy | A metric describing a participant's tendency to agree with *known cause*—modeled for this study as % *contradict* known cause, wherein lower values indicate higher accuracy. |
| AUC | Area under the receiver operating characteristic (ROC) curve. As used here AUC ranges from 0 (representing complete *contradiction* of *known cause*, majority, other participants) to 1 (representing complete agreement with *known cause*, majority, other participants). Random performance would result in an AUC of 0.5. |
| Classification prompt | Described in *Section 2.3.1.* |
| Consensus | The extent to which a majority or supermajority of *responses* are achieved for a *prompt*, or the extent to which a specific *response* agrees with the majority. |
| Contradiction | A *response* diametrically opposed to *known cause*, majority, or another participant's *response*. (e.g., a *response* of *excluded* on a true prompt; a *response* of *excluded* on a prompt with a majority conclusion of *definitive*; *responses* of *excluded* and *definitive* rendered by two participants on the same *prompt*). *Indeterminate responses* are not considered *contradictions*. |
| Decisiveness | A metric describing a participant's tendency to report *determinate responses*—modeled for this study as % *indeterminate*, wherein lower values indicate higher decisiveness. |
| Definitive | (A *determinate response* to a *classification prompt*) A participant's *response* indicating that this classification/mechanism is correct and accurate for this bloodstain pattern, and that the analyst is certain that the given mechanism was involved in the creation of the pattern. |
| Determinate | A *response* of *definitive* or *excluded* to a *classification prompt*, or *yes* or *no* to a *question*. (Compare to *indeterminate*) |
| Disagreement | A *contradiction* between two participants' *responses* on the same *prompt*. |
| Effectiveness | A metric that considers the trade-off between *accuracy*/*consensus*/*reproducibility* and *decisiveness*, modeled for this study as *AUC*. Effectiveness is a measure of the ability of the participant to concur with *known cause*, the majority response, or with other participants. |
| Error | A *contradiction* between a participant's *response* and the *known cause* for the given *prompt*/sample. |
| Excluded | (A *determinate response* to a *classification prompt*) A participant's *response* indicating that the classification/mechanism cannot be responsible for creating this bloodstain pattern, and that the analyst is |

| | |
|---|---|
| | certain that the given mechanism could NOT have been involved in the creation of the pattern. |
| False | (An evaluation of *known cause*) A *classification prompt* or *question* that does not represent the *known cause(s)* of a given sample. |
| Included | (An *indeterminate response* to a *classification prompt*) This classification/mechanism is neither *definitive* nor *excluded*. |
| Indeterminate | A response of *included* to a *classification prompt*, or *possible* to a *question*. (Compare to *determinate*) |
| Known Cause | The known classification mechanism(s) that certainly produced a given sample. |
| Negative | A *response* of *excluded* to a *classification prompt*, or *no* to a *question*. |
| No | (A *determinate response* to a *question*) A participant's *response* indicating that he/she disagrees with the entire *question* as stated. |
| Positive | A *response* of *definitive* to a *classification prompt*, or *yes* to a *question*. |
| Prompt | A c*lassification prompt* or *question* that was presented to participants in the study. |
| Possible | (An *indeterminate response* to a question) A participant's *response* indicating either that he/she is not certain, or that the *question* is partly *true* and partly *false*. |
| Question | Described in *Section 2.3.2.* |
| Reproducibility | A metric describing a participant's tendency to agree with other participants—modeled for this study as % *contradict* other participants, wherein lower values indicate higher reproducibility. |
| Response | A participant's decision on a specific prompt. For *classification prompts*, responses include *definitive, included,* or *excluded*. For *questions*, responses include *yes, possible,* or *no*. |
| Short text summary conclusions | Described in *Section 2.3.3.* |
| True | (An evaluation of *known cause*) A *classification prompt* or *question* that represents the *known cause(s)* of a given sample. |
| Tweet | Nickname for short text summary conclusion. |
| Unknown | (An evaluation of known cause) A *classification prompt* or *question* that either originates from a casework sample or is considered to be debatable/semantic with respect to *known cause* for a controlled sample. |
| Yes | (A *determinate response* to a *question*) A participant's response indicating that he/she agrees with the entire question as stated. |

# Appendix E Comparison of BPA Terminology from Various Sources

This section compares current and archival BPA terms and associated definitions from a variety of different sources, including the ASB standard (used for this study) [20], a number of IABPA standards [45]–[47], the FBI SWGSTAIN standard [48], and several BPA books [2], [3], [43]. Herein, we provide a summary table detailing the terms included within each source (Appendix E.1), and a list of all definitions for each included term from these sources (Appendix E.2).

We also reviewed the terminology list provided within HemoSpat [49], which provides the terms/definitions from the FBI SWGSTAIN standard along with alternative terminology (from [2], [3], [43], [46]) and representative photographs exhibiting examples of the specific patterns.

## Appendix E.1 Summary of BPA Terms from Various Sources

| *Term* | *ASB* [20] | *IABPA (1996)* [45] | *IABPA Course* [47] | *IABPA (2004)* [46] | *FBI SWGSTAIN* [48] | *Bevel & Gardener* [2] | *James, Kish, & Sutton* [3] | *Wonder* [43] |
|---|---|---|---|---|---|---|---|---|
| *Absence Transfer*[*] | | | | | | | | ● |
| *Accompanying Drop** | ● | | | | ● | | | |
| *Altered Stain* | ● | | | | ● | | | "Physiologically Altered Bloodstain" |
| *Angle of Impact** | ● | ● | | ● | ● | ● | ● | "Incident Angle/ Inside Angle" |
| *Area/Point of Convergence** | ● | ● | | ● | ● | | ● | ● |
| *Area/Point of Origin** | ● | ● | | ● | ● | ● | ● | ● |
| *Arterial Damage Stains** | | | | | | | | ● |
| *Arterial Fountain** | | | | | | | | ● |
| *Arterial Gushing/Spurting Pattern** | | ● | ● | ● | | "(Arterial) Spurt/Gush" | "Arterial Pattern" | Utilizes both "Gushing" and "Spurting" |
| *Arterial Rain** | | | | | | | | ● |
| *Backspatter* | ● | ● | | ● | ● | | ● | ● |
| *Blockage Transfer** | | | | | | | | ● |
| *Blood Clot* | ● | | | | ● | ● | ● | "Clot (coagulation)" |
| *Bloodstain** | ● | ● | | ● | ● | | ● | ● |
| *Blood into Blood Patterns* | | | | | | ● | | ● |
| *Bloodstain Pattern** | ● | | | | ● | | | ● |
| *Blunt Force Impact** | | | | | | | | ● |
| *Bubble Ring* | ● | | ● | ● | ● | | ● | |
| *Capillary Action** | | | | | | ● | | |
| *Cast-off Pattern* | ● | ● | | ● | ● | ● | ● | "Castoff Spatter" |
| *Cessation Pattern* | ● | | | | "Cessation Cast-off Pattern" | | "Cessation Cast-off Pattern" | "Cessation Cast-offs" |
| *Contamination** | | | | | | | | ● |
| *Direction of Flight** | | ● | | ● | | | | "Direction of Travel" |
| *Directional/Directionality Angle** | ● | ● | | ● | ● | ● | | |
| *Directionality** | ● | ● | | ● | ● | ● | | |
| *Draw-back Effect** | | ● | | ● | | "Blowback Effect" | ● | |
| *Drip Castoff* | | | | | | | | ● |
| *Drip Pattern* | ● | ● | ● | ● | ● | | ● | |

[*] *Indicates that the classification term was not used as a prompt in this study.*

| *Term* | *ASB* [20] | *IABPA (1996)* [45] | *IABPA Course* [47] | *IABPA (2004)* [46] | *FBI SWGSTAIN* [48] | *Bevel & Gardener* [2] | *James, Kish, & Sutton* [3] | *Wonder* [43] |
|---|---|---|---|---|---|---|---|---|
| *Drip Stain* | ● | | | | ● | ● | | |
| *Drip Trail* | ● | | | | ● | ● | | |
| *Edge Characteristic** | ● | | | | ● | | "Scalloping" | |
| *Entrance Wound Spatter** | | | | | | | | ● |
| *Exit Wound Spatter** | | | | | | | | ● |
| *Expiration Pattern* | ● | "Expirated Blood" | ● | "Expirated Blood" | ● | "Expectorate Spatter/Blood" | "Expirated Blood" | |
| *Flight Path** | | ● | | ● | | | | |
| *Flow Pattern* | ● | ● | ● | ● | ● | ● | ● | |
| *Fly Spot Pattern** | | | ● | | | ● | | |
| *Forward Spatter* | ● | ● | | ● | ● | | ● | |
| *High Velocity Impact Spatter (HVIS)** | | ● | | ● | | | | ● |
| *Impact Pattern* | ● | ● | ● | ● | ● | ● | | |
| *Impact Site** | | ● | | ● | | ● | | |
| *Investigative Transfer** | | | | | | | | ● |
| *Insect Stain* | ● | | | | ● | | | |
| *Low Velocity Impact Spatter (LVIS)** | | ● | | ● | | | | ● |
| *Medium Velocity Impact Spatter (MVIS)** | | ● | | ● | | | | ● |
| *Mist Pattern/Misting** | | ● | | ● | ● | "Atomized Blood/ Misting" | ● | ● |
| *Moving Contact Bloodstains** | | | | | | | | ● |
| *Non-Spatter Stains** | | | | | | ● | | |
| *Parent Stain/Drop** | ● | ● | | ● | ● | ● | ● | ● |
| *Passive Stain/Drop (Bleeding)** | | ● | | ● | | | "Passive bloodstains" | |
| *Pattern Transfer** | | | | | | ● | | |
| *Perimeter Stain* | ● | ● | "Perimeter/ Skeletonized" | ● | ● | | | |
| *Pool* | ● | | ● | | ● | | | "Pool (Volume)" |
| *Projected Pattern* | ● | ● | | ● | ● | | ● | |
| *Primary Stain** | | | | | | ● | | |
| *Ricochet** | | ● | | ● | | ● | | |
| *Satellite Stain* | ● | "Satellite Spatter" | | "Satellite Spatter" | ● | ● | | ● |
| *Saturation Stain* | ● | | ● | | ● | ● | | |
| *Secondary Spatter** | | | | | | | | ● |
| *Serum Stain* | ● | | | | ● | | | ● |
| *Shadowing** | | | | | | | | ● |
| *Simple Direct Transfer Patterns** | | | | | | | | ● |
| *Skeletonized Stain/ Skeletonization** | | | | | | ● | ● | |
| *Smear** | | | | | | ● | | |
| *Spatter* | ● | ● | | ● | ● | ● | "Spatter and Spatter pattern" | ● |
| *Spine** | | ● | | ● | | ● | ● | ● |
| *Splash* | ● | | ● | | ● | | ● | ● |

* *Indicates that the classification term was not used as a prompt in this study.*

| *Term* | *ASB* [20] | *IABPA (1996)* [45] | *IABPA Course* [47] | *IABPA (2004)* [46] | *FBI SWGSTAIN* [48] | *Bevel & Gardener* [2] | *James, Kish, & Sutton* [3] | *Wonder* [43] |
|---|---|---|---|---|---|---|---|---|
| *Splatter** | | | | | | | | ● |
| *Swipe Pattern* | ● | ● | ● | ● | ● | ● | ● | "Swipe Moving Transfer Pattern" |
| *Target** | ● | ● | | ● | ● | | ● | ● |
| *Template Transfers** | | | | | | | | ● |
| *Terminal Velocity** | | | | | | | ● | |
| *Transfer Stain/Pattern* | ● | | "Transfer/ Contact" | "Transfer/ Contact" | ● | "Contact Stain" | ● | ● |
| *Void* | ● | ● | ● | ● | ● | "Shadowing/ Ghosting/ Void" | ● | ● |
| *Wave Cast-off** | | ● | | ● | | | ● | |
| *Wipe* | ● | ● | ● | ● | ● | ● | ● | "Wipe Moving Transfer Pattern" |

## Appendix E.2 BPA Definitions from Various Sources

**Absence Transfer***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The lack of blood spatters seen in an area of a recording surface (target) due to the angle from the distributing event to the target being such that none of the spatters are directed at that area of the target. This suggested definition is to separate two different kinds of template transfers included within the term "void."

**Accompanying Drop***

- ASB:
  - A small blood drop produced as a by-product of drop formation. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A small blood drop produced as a by-product of drop formation. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Altered Stain**

- ASB:
  - A bloodstain with characteristics that indicate a physical change has occurred. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain with characteristics that indicate a physical change has occurred. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - *Physiologically Altered Bloodstain:* A bloodstain that is recorded after a physiological change occurred to the blood. This change may be drying, clotting, or mixing with another substance.

**Angle of Impact***

- ASB:
  - The acute angle (alpha), relative to the plane of a target, at which a blood drop strikes the target. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - The acute angle formed between the direction of a blood drop and the plane of the surface it strikes. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The acute angle formed between the direction of a blood drop and the plane of the surface it strikes. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - The acute angle (alpha), relative to the plane of a target, at which a blood drop strikes the target. *(Same as ASB definition)*
- Bevel & Gardener:
  - The acute angle as viewed from the side, created by the intercept of the target by the droplet's vector.
- James, Kish, & Sutton:
  - The acute or internal angle formed between the direction of a blood drop in flight and the plane of the surface it strikes.
- Wonder:
  - *Incident Angle:* The angle between the flight path of a blood drop and the target surface on which the blood spatter from the drop is recorded.
  - *Inside Angle (Incident Angle or Impact Angle):* Angle within a calculated right angle triangle that is drawn between a recorded spatter and a theoretical blood drop in flight from a blood source.

**Area/Point of Convergence***

- ASB:
  - The space in two dimensions to which the directionalities of spatter stains can be retraced to determine the location of the spatter producing event.

- IABPA (1996):
  - The common point (area), on a two dimensional surface, over which the directionality of several blood drops can be retraced. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The common point (area), on a two dimensional surface, over which the directionality of several blood drops can be retraced. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - The area containing the intersections generated by lines drawn through the long axes of individual stains that indicates in two dimensions the location of the blood source.

- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - The area to which stains within a bloodstain pattern can be recon- structed on a two-dimensional surface determined by tracing the long axis of well-defined bloodstains within the pattern to a common area.
- Wonder:
  - The locus of points on a plane surface defining an area where accurately placed direction of travel lines intersect. This area should be proven to have resulted from a single impact event.

**Area/Point of Origin***

- ASB:
  - The space in three dimensions to which the trajectories of spatter can be utilized to determine the location of the spatter producing event.
- IABPA (1996):
  - The common point (area) in three dimensional space to which the trajectories of several blood drops can be retraced. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The common point (area) in three dimensional space to which the trajectories of several blood drops can be retraced. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - The three-dimensional location from which spatter originated.
- Bevel & Gardener:
  - *Origin/Area of Origin:* The area in three-dimensional space from where a blood drop originates.
- James, Kish, & Sutton:
  - The three-dimensional area to which stains within a bloodstain pattern can be reconstructed in space using the common area of convergence and the angles of impact.
- Wonder:
  - *Origin:* The locus of points in space from which a group of blood drops originated at the time of a single injuring impact event.

**Arterial Damage Stains***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined

- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Unique and identifiable bloodstain patterns that result from blood drops distributed by the pulsing, pressurized release from a breached arterial blood vessel. This type of pattern may continue to be recorded as long as the heart continues to beat. If the heart stops, the pressure drops, and the column of blood is no longer forced out as a recordable spatter pattern.

**Arterial Fountain***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The recording of blood drops, usually on an adjacent vertical surface, from fallout of a column during arterial projection. Arterial pressure may push a column of blood upward, while the drops that separate from the column fall in reverse by gravity. The downward directions may be recorded on adjacent surfaces in recognizable parabola-shaped patterns.

**Arterial Gushing/Spurting Pattern***

- ASB:
  - Not defined
- IABPA (1996):
  - Bloodstain pattern(s) resulting from blood exiting the body under pressure from a breached artery. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Bloodstain pattern(s) resulting from blood exiting the body under pressure from a breached artery. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - *(Arterial) Gush/Spurt:* The escape of blood under pressure, typically from a breach in an artery or heart, showing pressure, pressure fluctuations, or both.
- James, Kish, & Sutton:
  - *Arterial Pattern:* A pattern resulting from blood exiting the body under pressure from a breached artery.

- Wonder:
  - *Arterial Gush:* A large bloodstain, more than a few overlapping drops, projected from a breached artery that remains in one position while the heart continues to beat.
  - *Arterial Spurt:* Blood exiting a breached arterial vessel as a column that separates into drops as the victim moves. The drops may be recorded in parallel arrangements of similar-sized and similar-shaped spatters.

**Arterial Rain***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The scattered spatters recorded on a horizontal surface that have resulted from the fallout of blood drops from an arterial fountain.

**Backspatter**

- ASB:
  - A bloodstain pattern resulting from blood drops that traveled in the opposite direction of the external force applied; associated with an entrance wound created by a projectile. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Blood directed back towards the source of energy or force that caused the spatter. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Blood directed back towards the source of energy or force that caused the spatter. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from blood drops that traveled in the opposite direction of the external force applied; associated with an entrance wound created by a projectile. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Blood droplets directed back toward the force or energy that caused the spatter, often associated with gunshot wounds of entrance.
- Wonder:
  - A bloodstain pattern that results from a dynamic act suggestive of Newton's Third Law of Motion.

**Blockage Transfer***

- ASB:
  - Not defined

- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The lack of blood spatters found in an area of a target due to the location of a blockage somewhere along the flight path between the blood source of distributed blood drops and the recording surface of those drops. The term is called "transfer," although no contact between surfaces occurred, because the distribution of spatters can outline an object, thus transferring the shape of the object to a recording surface. Template transfers terminology is sometimes applied to this pattern identification.

**Blood Clot**

- ASB:
  - A gelatinous mass formed by a complex mechanism involving red blood cells, fibrinogen, platelets, and other clotting factors. *(Same as FBI SWGSTAIN and James, Kish & Sutton definitions)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A gelatinous mass formed by a complex mechanism involving red blood cells, fibrinogen, platelets, and other clotting factors. *(Same as ASB and James, Kish & Sutton definitions)*
- Bevel & Gardener:
  - A gelatinous mass formed by the collection of blood cells in fibrin; this mass will usually exhibit separation of the liquid and solid materials.
- James, Kish, & Sutton:
  - A gelatinous mass formed by a complex mechanism involving red blood cells, fibrinogen, platelets, and other clotting factors. *(Same as ASB and FBI SWGTREAD definitions)*
- Wonder:
  - *Clot (Coagulation):* The biochemical reaction within a quantity of blood where liquid fibrinogen molecules catalyzed by calcium cause a precipitation which forms in a semisolid fibrin matrix.
  - *Retraction (Clot):* The stage in the coagulation process in which fibrin strands tighten the bundle of cells into a solid mass while extruding liquid serum. Retraction can contribute to a time line if identified and recorded soon after a crime scene is found.

**Bloodstain***

- ASB:
  - A deposit of blood on a surface. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Evidence that liquid blood has come into contact with a surface. *(Same as IABPA (2004) definition)*

- IABPA (2004):
  - Evidence that liquid blood has come into contact with a surface. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A deposit of blood on a surface. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A stain on a surface caused by blood.
- Wonder:
  - Substance found on a surface, identified and verified by specific tests, to be blood.

**Blood into Blood Patterns**

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A pattern created when drops are deposited into one another or into another liquid resulting in an accumulation of blood surrounded by randomly oriented secondary spatter.
- Wonder:
  - A pattern created when blood drips into a pool of blood. This may be recognized as a periphery of secondary smaller spatters seen around the circumference of the pool, a pattern created when the internal cohesion and non-Newtonian elasticity of blood drops cause them to break up and bounce off the pool, rather than immediately diffuse into it.

**Bloodstain Pattern***

- ASB:
  - A grouping or distribution of bloodstains that indicates through regular or repetitive form, order, or arrangement the manner in which the pattern was deposited. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A grouping or distribution of bloodstains that indicates through regular or repetitive form, order, or arrangement the manner in which the pattern was deposited. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined

- Wonder:
  - *Pattern:* An identifiable arrangement of several blood spatters recorded on a target from one dynamic event.

**Blunt Force Impact***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - An act involving an object of some size and weight sufficient to cause bodily injury creating a blood source without specific penetration by knife or missile (bullet).

**Bubble Ring**

- ASB:
  - An outline within a bloodstain resulting from air in the blood. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Rings in blood that result when blood containing air bubbles dries and retains the bubble's circular configuration as a dried outline. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Rings in blood that result when blood containing air bubbles dries and retains the bubble's circular configuration as a dried outline. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - An outline within a bloodstain resulting from air in the blood. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Vacuoles in bloodstains that form when blood containing air bubbles dries and retains the circular configuration of the original bubble.
- Wonder:
  - Not defined

**Capillary Action***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined

- Bevel & Gardener:
  - The force exhibited in the attraction of a liquid to surfaces with which it is in contact and its own surface tension. This attraction often results in stain characteristics for which no corresponding defect may exist.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Cast-off Pattern**

- ASB:
  - A bloodstain pattern resulting from blood drops released from an object due to its motion. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - A bloodstain pattern created when blood is released or thrown from a blood-bearing object in motion. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A bloodstain pattern created when blood is released or thrown from a blood-bearing object in motion. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from blood drops released from an object due to its motion. *(Same as ASB definition)*
- Bevel & Gardener:
  - Patterns created when blood is flung or projected from an object in motion or one that suddenly stops some motion.
- James, Kish, & Sutton:
  - A pattern usually linear in configuration when blood is released or flung from a blood-bearing object in motion.
- Wonder:
  - *Castoff Spatters:* Spatters resulting from blood drops distributed by centripetal force and/or gravity, when the adhesion and cohesion are broken between blood and the weapon, carrier, or blood source. Drops that are distributed have sufficient mass and momentum to overcome fluid elasticity and the adhesion between the blood substance and carrier.

**Cessation Pattern**

- ASB:
  - A bloodstain pattern resulting from blood drops released from an object due to its abrupt deceleration.
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - *Cessation Cast-off Pattern:* A bloodstain pattern resulting from blood drops released from an object due to its rapid deceleration.
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - *Cessation Cast-off Pattern:* A pattern resulting from the rapid deceleration of an object wet with blood.

- Wonder:
  - *Cessation Castoffs:* Blood drops are formed when the motion of a blood carrier immediately stops or reverses in travel. Adhesion is broken between the blood and the moving object so that drops are formed and distributed. This event leads to continued travel of the drops formed in the relative direction the carrier was moving before motion was interrupted.

**Contamination***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Changes to the bloodstain patterns at a crime scene by assumed uninvolved persons after the crime events are over.

**Direction of Flight***

- ASB:
  - Not defined
- IABPA (1996):
  - The trajectory of a blood drop which can be established by its angle or impact and directionality angle. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The trajectory of a blood drop which can be established by its angle or impact and directionality angle. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - *Direction of Travel:* A line drawn from the smoothest side to the opposite most irregular edge of a blood spatter, which determines the direction the blood drop was traveling when it contacted the target. This is not necessarily the longest line through the stain because of differential absorption and/or texture of the target surface material.

**Directionality***

- ASB:
  - The characteristic of a bloodstain that indicates the direction blood was moving at the time of deposition. *(Same as FBI SWGSTAIN definition)*

- IABPA (1996):
  - The directionality of a bloodstain or pattern which indicates the direction the blood was traveling when it impacted the target surface. Directionality of a blood drop's flight can usually be established from the geometric shape of its bloodstain. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The directionality of a bloodstain or pattern which indicates the direction the blood was traveling when it impacted the target surface. Directionality of a blood drop's flight can usually be established from the geometric shape of its bloodstain. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - The characteristic of a bloodstain that indicates the direction blood was moving at the time of deposition. *(Same as ASB definition)*
- Bevel & Gardener:
  - Relating to or indicating the vector a droplet follows in relation to a target.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Directional/Directionality Angle***

- ASB:
  - The angle (gamma) between the long axis of a spatter stain and a defined reference line on the target. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - The angle between the long axis of a bloodstain and a predetermined line on the plane of the target surface which represents 0 degrees. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The angle between the long axis of a bloodstain and a predetermined line on the plane of the target surface which represents 0 degrees. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - The angle (gamma) between the long axis of a spatter stain and a defined reference line on the target. *(Same as ASB definition)*
- Bevel & Gardener:
  - The angle, as viewed from the front of the target between the long axis of the stain and a standard reference point
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Draw-back Effect***

- ASB:
  - Not defined
- IABPA (1996):
  - Blood in the barrel of a firearm that has been drawn backward into the muzzle. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Blood in the barrel of a firearm that has been drawn backward into the muzzle. *(Same as IABPA (1996) definition)*

- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - *Blowback Effect:* The process in which blood is deposited inside the barrel of a weapon after discharge.
- James, Kish, & Sutton:
  - The presence of blood in the barrel of a firearm that has been drawn back into the muzzle.
- Wonder:
  - Not defined

**Drip Castoff**

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - A pattern resulting from blood drops that overcome adhesion to an object and fall by force of gravity to a target surface.

**Drip Pattern**

- ASB:
  - A bloodstain pattern resulting from a liquid that dripped into another liquid, at least one of which was blood. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - A bloodstain pattern which results from blood dripping into blood. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A bloodstain pattern which results from blood dripping into blood. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from a liquid that dripped into another liquid, at least one of which was blood. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A bloodstain pattern created by free-falling drops of blood striking already- existing blood on a surface commonly associated with satellite spatter. The parent stain on the surface is larger than what would be associated with a single free-falling drop and is usually associated with satellite spatter.
- Wonder:
  - Not defined

**Drip Stain**

- ASB:
  - A bloodstain resulting from a falling drop that formed due to gravity. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain resulting from a falling drop that formed due to gravity. *(Same as ASB definition)*
- Bevel & Gardener:
  - Spatter resulting from blood dripping from an individual or otherwise bloodied object.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Drip Trail**

- ASB:
  - A bloodstain pattern resulting from the movement of a source of drip stains between two points. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from the movement of a source of drip stains between two points. *(Same as ASB definition)*
- Bevel & Gardener:
  - A drip trail is a deposit of a series of drips in linear orientations.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - *Inline Beading:* The in-tandem arrangement of spatters recorded on a target signifying blood drops from a column or single blood source as it moves. These may occur from castoff, arterial, or hair fiber transfer distribution.

**Edge Characteristic***

- ASB:
  - A physical feature of the periphery of a bloodstain. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A physical feature of the periphery of a bloodstain. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined

- James, Kish, & Sutton:
  - *Scalloping:* A serrated edge characteristic of bloodstains.
- Wonder:
  - Not defined

**Entrance Wound Spatter***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Blood drops distributed according to Newton's third law. This event occurs relative to the medically identified entrance wound of a victim, which is as an equal and opposite reaction to the impact of a bullet or missile. Newton's law applied to solid material rather than breaking up of a liquid but later was used long after his death to explain liquid behavior in the study of fluid mechanics. In reality, the result of a bullet penetrating flesh distributes blood drops away from the entrance of the bullet in an array suggestive of a cone shape.

**Exit Wound Spatter***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Specifically associated with gunshot wounds (GSW), the pattern that may result from blood distributed after a bullet/missile exits a body.

**Expiration Pattern**

- ASB:
  - A bloodstain pattern resulting from blood forced by airflow out of the nose, mouth, or a wound. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - *Expirated Blood:* Blood that is blown out of the nose, mouth, or a wound as a result of air pressure and/or air flow which is the propelling force. *(Same as IABPA (2004) definition)*

- IABPA (2004):
  - *Expirated Blood:* Blood that is blown out of the nose, mouth, or a wound as a result of air pressure and/or air flow which is the propelling force. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from blood forced by airflow out of the nose, mouth, or a wound. *(Same as ASB definition)*
- Bevel & Gardener:
  - *Expectorate Spatter/Blood:* Spatter created when blood is forced from the mouth, nose, or respiratory system under pressure.
- James, Kish, & Sutton:
  - *Expirated Blood:* Blood that has been blown from the nose, the mouth, or a wound in the respiratory system as the result of air flow or pressure.
- Wonder:
  - Not defined

**Flight Path***

- ASB:
  - Not defined
- IABPA (1996):
  - The path of the blood drop, as it moves through space, from the impact site to the target. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The path of the blood drop, as it moves through space, from the impact site to the target. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Flow Pattern**

- ASB:
  - A bloodstain pattern resulting from the movement of a volume of blood on a surface due to gravity or movement of the target. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - A change in the shape and direction of a bloodstain due to the influence of gravity or movement of the object. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A change in the shape and direction of a bloodstain due to the influence of gravity or movement of the object. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from the movement of a volume of blood on a surface due to gravity or movement of the target. *(Same as ASB definition)*
- Bevel & Gardener:
  - The movement of liquid blood as a mass under the effect of gravity.

- James, Kish, & Sutton:
  - A volume of blood on a surface that moves in one or more directions as a result of the influence of gravity.
- Wonder:
  - Not defined

**Fly Spot Pattern***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Stains resulting from fly activity
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Forward Spatter**

- ASB:
  - A bloodstain pattern resulting from blood drops which can be produced when a projectile creates an exit wound.
- IABPA (1996):
  - Blood which travels in the same direction as the source of energy or force which caused the spatter. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Blood which travels in the same direction as the source of energy or force which caused the spatter. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from blood drops that traveled in the same direction as the impact force.
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Blood droplets directed away from the force or energy that caused the spatter, often associated with gunshot wounds of exit.
- Wonder:
  - Not defined

**High Velocity Impact Spatter (HVIS)***

- ASB:
  - Not defined
- IABPA (1996):
  - A bloodstain pattern caused by a high velocity impact/force to a blood source such as that produced by gunshot or high speed machinery. *(Same as IABPA (2004) definition)*

- IABPA (2004):
  - A bloodstain pattern caused by a high velocity impact/force to a blood source such as that produced by gunshot or high speed machinery. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - A term coined by Dr. Paul L. Kirk to explain the exclamation mark appearance of blood spatters. The designation was relative to how fast a blood drop was traveling when it made contact with the target. This affects the shape of the resultant stain.

**Impact Pattern**

- ASB:
  - A bloodstain pattern resulting from an object striking liquid blood. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Bloodstain pattern created when blood receives a blow or force resulting in the random dispersion of smaller drops of blood. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Bloodstain pattern created when blood receives a blow or force resulting in the random dispersion of smaller drops of blood. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from an object striking liquid blood. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Impact Site***

- ASB:
  - Not defined
- IABPA (1996):
  - That point where force encounters a source of blood.
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - The point where a given source encounters a blood source
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Investigative Transfer***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The transfer patterns, where blood is deposited on crime scene surfaces from foot, shoe, hand, or equipment used at a crime scene by investigative personnel. This transfer should be recognized for value in developing a time line and/or eliminating necessary rescue and investigative presence to a crime scene.

**Insect Stain**

- ASB:
  - A bloodstain resulting from insect activity. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain resulting from insect activity. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Low Velocity Impact Spatter (LVIS)***

- ASB:
  - Not defined
- IABPA (1996):
  - A bloodstain pattern that is caused by a low velocity impact/force to a blood source. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A bloodstain pattern that is caused by a low velocity impact/force to a blood source. *(Same as IABPA (2004) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined

- Wonder:
  - Defined by Dr. Paul L. Kirk as blood drops falling by gravitational pull alone, thus impacting a target at relatively low velocity. This affects the shape of the resultant spatter.

**Medium Velocity Impact Spatter (MVIS)***

- ASB:
  - Not defined
- IABPA (1996):
  - A bloodstain pattern caused by a medium velocity impact/force to a blood source. A beating typically causes this type of spatter. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A bloodstain pattern caused by a medium velocity impact/force to a blood source. A beating typically causes this type of spatter. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Based on Dr. Paul L. Kirk's classifications, medium velocity impact spatter is spatter resulting from blood drops traveling slower than distributed by events such as gunshot but faster than dripping by gravity alone. All of the dynamic classifications of events—castoff, blunt force impact, arterial spurts, and/or respiratory coughs—may distribute blood drops that can leave blood spatters of a size and shape included in those historically called MVIS.

**Mist Pattern/Misting***

- ASB:
  - Not defined
- IABPA (1996):
  - Blood which has been reduced to a fine spray, as a result of the energy or force applied to it. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Blood which has been reduced to a fine spray, as a result of the energy or force applied to it. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from blood reduced to a spray of micro-drops as a result of the force applied.
- Bevel & Gardener:
  - *Atomized Blood/Misting:* Bloodstain patterns characterized by a mist-like appearance, which are generally associated with an explosive force such as a gunshot
- James, Kish, & Sutton:
  - Blood that has been atomized to a fine spray by the application of force, usually associated with gunshot or explosion events.
- Wonder:
  - *Mist Spatter or Misting:* When blood is atomized to form drops so small they ultimately leave stains of less than 0.1 mm in diameter. Detection of mist-sized drops may require microscopic examination.

**Moving Contact Bloodstains***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Transfer of blood during the movement of an object, skin, or material across a record- ing surface while depositing or spreading blood substance.

**Non-Spatter Stains**

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Any stain or pattern other than those defined by the spatter group (e.g., patterns not composed of small circular or elliptical shaped stains).
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Parent Stain/Drop***

- ASB:
  - A bloodstain from which a satellite stain(s) originated. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - A drop of blood from which a wave, cast-off, or satellite spatter originates. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A drop of blood from which a wave, cast-off, or satellite spatter originates. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain from which a satellite stain originated. *(Same as ASB definition)*
- Bevel & Gardener:
  - The spatter stain from which satellite spatter originates.
- James, Kish, & Sutton:
  - A bloodstain from which satellite spatter or wave cast-off stains originate.

- Wonder:
  - The original blood drop that leaves a spatter stain after travel from the origin to the target. The term "parent" is used to differentiate the source of a spatter from secondary or wave castoffs that occur after the drop has impacted the target.

**Passive Drop/Stain (Bleeding)***

- ASB:
  - Not defined
- IABPA (1996):
  - Bloodstain drop(s) created or formed by the force of gravity acting alone. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Bloodstain drop(s) created or formed by the force of gravity acting alone. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Stains and patterns whose physical features indicate that they were created without any significant outside force other than gravity and friction.
- Wonder:
  - Not defined

**Pattern Transfer***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - A pattern created by the transfer of blood from one object to another in which a recognizable characteristic or image is present in the pattern.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Perimeter Stain**

- ASB:
  - An altered stain consisting of its edge characteristics, the central area having been partially or entirely removed.
- IABPA (1996):
  - A bloodstain that consists of only its outer periphery, the central area having been removed by wiping or flaking after liquid blood has partially or completely dried. *(Same as IABPA (2004) definition)*

- IABPA (2004):
  - A bloodstain that consists of only its outer periphery, the central area having been removed by wiping or flaking after liquid blood has partially or completely dried. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - An altered stain that consists of the peripheral characteristics of the original stain.
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Pool**

- ASB:
  - A bloodstain resulting from an accumulation of liquid blood on a surface. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain resulting from an accumulation of liquid blood on a surface. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - *Pool (Volume):* The term "volume" may be used for science purposes. A volume of blood is a quantity in excess of a few drops, which collects from bleeding injuries and lies undisturbed.
  - *Volume (Pool):* The accumulation of blood larger than a few drops.

**Projected Pattern**

- ASB:
  - A bloodstain resulting from the ejection of blood under hydraulic pressure, typically from a breach in the circulatory system.
- IABPA (1996):
  - A bloodstain pattern that is produced by blood released under pressure as opposed to an impact, such as arterial spurting. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A bloodstain pattern that is produced by blood released under pressure as opposed to an impact, such as arterial spurting. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain resulting from the ejection of a volume of blood under pressure.
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Spatter created as the result of a force other than impact.



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

- Wonder:
  - Not defined

**Primary Stain***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - The main stain found in any pattern.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Ricochet***

- ASB:
  - Not defined
- IABPA (1996):
  - The deflection of blood after impact with a target surface that results in staining of a second target surface. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The deflection of blood after impact with a target surface that results in staining of a second target surface. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Blood that impacts an object and then bounces or falls to another target.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Satellite Stain**

- ASB:
  - A smaller bloodstain that originated during the formation of the parent stain as a result of blood impacting a surface. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Small droplets of blood that are distributed around a drop or pool of blood as a result of the blood impacting the target surface. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - Small droplets of blood that are distributed around a drop or pool of blood as a result of the blood impacting the target surface. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A smaller bloodstain that originated during the formation of the parent stain as a result of blood impacting a surface. *(Same as ASB definition)*

- Bevel & Gardener:
  - Small stains created when droplets detach from a large drop as it impacts a target.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - *Satellite Spatter:* Usually, several small to tiny spatters formed from a primary event such as blood drip- ping into blood or arterial gushing on the target surface.

**Saturation Stain**

- ASB:
  - A bloodstain resulting from the accumulation of liquid blood in an absorbent material. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain resulting from the accumulation of liquid blood in an absorbent material. *(Same as ASB definition)*
- Bevel & Gardener:
  - An accumulation of liquid blood created by contact with a volume of blood that is absorbed into a permeable surface.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Not defined

**Secondary Spatter***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The single small drop thrown ahead of a larger "parent" drop by the wave motion when the drop contacts a textured recording surface. The leading edge of contact breaks like an ocean wave with a small amount of the original drop thrown ahead of the final parent spatter.

**Serum Stain**

- ASB:
  - The stain resulting from the liquid portion of blood (serum) that separates during coagulation. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - Not defined

- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - The stain resulting from the liquid portion of blood (serum) that separates during coagulation. *(Same as ASB definition)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A yellowish stain resulting from the separation of serum from the retraction of a blood clot.
- Wonder:
  - A stain that may be separated from a clotted portion of blood but is seen as the color- less, shiny (reflective) edge, or spots. This pattern is sometimes seen in photographs later rather than when viewing the actual crime scene. Serum stains are associated with clot retraction, i.e., sometime after bloodshed, and occur before the bloodstain dries.

**Shadowing**

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - The effect seen on the edge of a blockage pattern by some of, but not all of, the blood drops from the distributing event outlining the object. The drops have been partially blocked or they touched a rounded surface, following paths of least resistance to form a shadow outline to the full blockage pattern.

**Simple Direct Transfer Patterns***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined

- Wonder:
  - A pattern which results from two surfaces, one or both holding blood, coming into contact without movement, i.e., touch transfer only.

**Skeletonized Stain/ Skeletonization***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - A bloodstain that, although disturbed, still reflects its original shape and size.
- James, Kish, & Sutton:
  - A bloodstain that consists of a darkened outer peripheral rim with the central portion of the stain having been removed by wiping through the partially dry stain. A skeletonized bloodstain is also produced by the flaking of the central portion of a completely dried stain.
- Wonder:
  - Not defined

**Smear***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Any stain or pattern created by the transfer of blood from one object to another, through some form of contact involving lateral motion.
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - Smear (Smudge): A moving transfer pattern that cannot be limited to an identifiable direction of travel. Back-and-forth motions in cleaning may leave this type of stain.

**Spatter**

- ASB:
  - A bloodstain resulting from an airborne blood drop created when external force is applied to liquid blood.
- IABPA (1996):
  - That blood which has been dispersed as a result of force applied to a source of blood. Patterns produced area often characteristic of the nature of the forces which created them. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - That blood which has been dispersed as a result of force applied to a source of blood. Patterns produced area often characteristic of the nature of the forces which created them. *(Same as IABPA (1996) definition)*

- FBI SWGSTAIN:
  - A bloodstain resulting from a blood drop dispersed through the air due to an external force applied to a source of liquid blood.
- Bevel & Gardener:
  - Those stains resulting from blood that has been put in free flight and subsequently impacted a surface.
- James, Kish, & Sutton:
  - *Spatter:* Blood stains that exhibit directionality and variation in size and are associated with a source of blood being subjected to external force(s).
  - *Spatter Pattern:* A distribution of individual spatters on a surface that can be traced to a common area of origin.
- Wonder:
  - The individual bloodstain recorded when a distributed blood drop contacts a target surface.

**Spine***

- ASB:
  - Not defined
- IABPA (1996):
  - The pointed or elongated stains which radiate away from the central area of a bloodstain. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The pointed or elongated stains which radiate away from the central area of a bloodstain. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Linear characteristics evident in both single drop stains and volume stains.
- James, Kish, & Sutton:
  - An edge characteristic of bloodstain consisting of narrow, elongated projections from the central area of the stain.
- Wonder:
  - A pointed irregularity located at the circumference of a spatter

**Splash Pattern**

- ASB:
  - A bloodstain created from a large volume of liquid blood falling onto a surface.
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - A bloodstain pattern resulting from a volume of liquid blood that falls or spills onto a surface.
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A bloodstain created by a free-falling volume of blood in excess of 1.0 ml onto a surface from a distance of at least 4 in., OR: An altered bloodstain pattern characterized by a preexisting volume on a surface that has been subjected to additional force creating elongated narrow spines such as stepping into blood.

- Wonder:
  - Spatters distributed from an impact to a volume (pool) of blood.

**Splatter***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - A spatter resulting from a blood drop distributed from a splash.

**Swipe Pattern**

- ASB:
  - A bloodstain resulting from the transfer of blood from a blood-bearing surface onto another surface, with characteristics that indicate relative motion between the two surfaces. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - The transfer of blood from a moving source onto an unstained surface. Direction of travel may be determined by the feathered edge. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - The transfer of blood from a moving source onto an unstained surface. Direction of travel may be determined by the feathered edge. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain resulting from the transfer of blood from a blood-bearing surface onto another surface, with characteristics that indicate relative motion between the two surfaces. *(Same as ASB definition)*
- Bevel & Gardener:
  - Any stain or pattern created by the transfer of blood from a bloodied object to another by some form of lateral motion.
- James, Kish, & Sutton:
  - The transfer of blood associated with motion onto a nonbloody surface.
- Wonder:
  - *Swipe Moving Transfer Pattern:* A moving contact transfer where a bloodied material brushes against a target.

**Target***

- ASB:
  - A surface onto which blood has been deposited. *(Same as FBI SWGSTAIN, IABPA (1996 & 2004), and James, Kish, & Sutton definitions)*
- IABPA (1996):
  - A surface upon which blood has been deposited. *(Same as ASB, IABPA (2004), FBI SWGSTAIN, and James, Kish, & Sutton definitions)*

- IABPA (2004):
  - A surface upon which blood has been deposited. *(Same as ASB, IABPA (1996 & 2004), FBI SWGSTAIN and James, Kish, & Sutton definitions)*
- FBI SWGSTAIN:
  - A surface onto which blood has been deposited. *(Same as ASB, IABPA (1996 & 2004), and James, Kish, & Sutton definitions)*
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A surface onto which blood has been deposited. *(Same as ASB, IABPA (1996 & 2004), and FBI SWGSTAIN definitions)*
- Wonder:
  - The surface upon which blood is recorded during accidental, suicidal, and/or criminal injurious events.

**Template Transfers***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - Not defined
- Wonder:
  - A description of a blockage pattern where the obstruction acts like a template or stencil blocking all distributed blood drops. The true shape and size of the obstruction may be seen.

**Terminal Velocity***

- ASB:
  - Not defined
- IABPA (1996):
  - Not defined
- IABPA (2004):
  - Not defined
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - The maximum velocity that a free-falling drop of blood can accelerate in air, determined to be approximately 25.1 feet per second.
- Wonder:
  - Not defined

**Transfer Stain/Pattern**

- ASB:
  - A bloodstain resulting from contact between a blood-bearing surface and another surface. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - *Transfer/Contact Pattern:* A bloodstain pattern created when a wet, bloody surface comes in contact with a second surface. A recognizable image of all or a portion of the original surface may be observed in the pattern. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - *Transfer/Contact Pattern:* A bloodstain pattern created when a wet, bloody surface comes in contact with a second surface. A recognizable image of all or a portion of the original surface may be observed in the pattern. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - A bloodstain resulting from contact between a blood-bearing surface and another surface. *(Same as ASB definition)*
- Bevel & Gardener:
  - *Contact Stain*: Any stain or pattern created by the transfer of blood from one object to another through physical contact.
- James, Kish, & Sutton:
  - The deposition of blood onto a surface as the result of contact.
- Wonder:
  - *Transfer Pattern:* A bloodstain pattern on a target surface of an image, material, or object resulting from touch or blood drop distribution from a blood distributing event. This may either outline or transfer the image of an object present. Unlike transfer evidence in general forensic terminology, blood spatters may outline an object without direct contact between the object and the recording target.

**Void**

- ASB:
  - An absence of blood in an otherwise continuous bloodstain or bloodstain pattern. *(Same as IABPA (1996 & 2004), and FBI SWGSTAIN definitions)*
- IABPA (1996):
  - An absence of stains in an otherwise continuous bloodstain pattern. *(Same as ASB, IABPA (2004), and FBI SWGSTAIN definitions)*
- IABPA (2004):
  - An absence of stains in an otherwise continuous bloodstain pattern. *(Same as ASB, IABPA (1996), and FBI SWGSTAIN definitions)*
- FBI SWGSTAIN:
  - An absence of blood in an otherwise continuous bloodstain or bloodstain pattern. *(Same as ASB and IABPA (1996 & 2004) definitions)*
- Bevel & Gardener:
  - *Shadowing/Ghosting/Void:* An area within a generally continuous bloodstain pattern that lacks bloodstains.
- James, Kish, & Sutton:
  - The absence of blood in an otherwise continuous bloodstain pattern that suggests the presence of an intermediate target that may have been removed.



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

- Wonder:
  - A term that is recognized in court but historically has shifted meaning with time. In the 1940s, it was used to identify a blank space, which fits this definition. In recent semantics, and for scientific application, the term has moved toward an emptying or canceling as null and void (voiding a check) or voiding (emptying out) the bladder for a urinary specimen.

**Wave Cast-off***

- ASB:
  - Not defined
- IABPA (1996):
  - A small blood drop that originates from a parent drop of blood due to the wave-like action of the liquid in conjunction with striking a surface. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A small blood drop that originates from a parent drop of blood due to the wave-like action of the liquid in conjunction with striking a surface. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - Not defined
- Bevel & Gardener:
  - Not defined
- James, Kish, & Sutton:
  - A small stain that has originated from a parent stain as the result of the wavelike action of the original drop striking a surface at an angle of less than 90 degrees.
- Wonder:
  - Not defined

**Wipe**

- ASB:
  - An altered stain resulting from an object moving through a preexisting wet bloodstain. *(Same as FBI SWGSTAIN definition)*
- IABPA (1996):
  - A bloodstain pattern created when an object moves through an existing stain, removing and/or altering its appearance. *(Same as IABPA (2004) definition)*
- IABPA (2004):
  - A bloodstain pattern created when an object moves through an existing stain, removing and/or altering its appearance. *(Same as IABPA (1996) definition)*
- FBI SWGSTAIN:
  - An altered bloodstain pattern resulting from an object moving through a preexisting wet bloodstain. *(Same as ASB definition)*
- Bevel & Gardener:
  - Any stain or pattern created when an object moves through a preexisting bloodstain on another surface.

- James, Kish, & Sutton:
  - An alteration of a preexisting wet or partially dry bloodstain caused by movement through the existing stain.
- Wonder:
  - *Wipe Moving Transfer Pattern:* A moving contact transfer where one material brushes across a bloodstained target.