



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

Document Title: Evaluation of Peer Review and Verification Processes and Evaluation of an Enhanced Screening Technique for Toxicological Specimens with High Resolution Accurate Mass Spectrometer

Author(s): Kathryn P. Lee

Document Number: 254412

Date Received: December 2019

Award Number: 2016-DN-BX-K005

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

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.

Final Summary Overview

NIJ Award Number: 2016-DN-BX-K005

Project: (1) Evaluation of Peer Review and Verification Processes
(2) Evaluation of an Enhanced Screening Technique for Toxicological Specimens with High Resolution Accurate Mass Spectrometer

Overview Prepared By: Kathryn P. Lee
Deputy Director
Georgia Bureau of Investigation Division of Forensic Sciences
kathy.lee@gbi.ga.gov
404-270-8082

Project Period: January 1, 2017 – December 31, 2018

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

Project 1:

Evaluation of Peer Review and Verification Processes

Purpose:

Technical and administrative review by a second qualified scientist, also known as peer review, consume a significant amount of time and reduce the scientist's ability to perform additional analyses, research, training and other critical functions. Peer review of 100% of reports has become the standard in many forensic laboratories. The goal of these reviews is to eliminate Type 1 errors, in which the data or conclusion incorrectly associates two samples or incorrectly identifies a substance, and to minimize Type 2 errors, in which the analysis fails to associate samples or identify a substance.

This study tests the hypothesis that conducting 100% technical and administrative review is the most effective approach to minimizing Type 1 and Type 2 errors.

Project Subjects:

This project examined cases completed at the Georgia Bureau of Investigation (GBI) Division of Forensic Sciences in the forensic disciplines of Chemistry (drug identification reports), Toxicology, Firearms, and Latent Prints.

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

Project Design and Methods:

A listing of reports completed during 2016 for the disciplines of Firearms, Latent Prints, Toxicology, and Chemistry was compiled using SQL queries from the Laboratory Information Management System (LIMS). Two lists of reports were generated from each forensic discipline; one list from reports in which no issues were identified during the initial review process, and the other list from reports in which an issue was identified (peer rejection) and corrected prior to release of the report.

Discipline	No Issue on Original Review	Issue on Original Review	Total Reports
Chemistry	22,461	2,431	24,892
Toxicology	25,942	1,232	27,174
Latent Prints	1,390	271	1,661
Firearms	2,876	210	3,086

All reports with an issue identified during review were utilized in the data collection. From the selection of reports without rejection, an equivalent number of reports were randomly sampled for data collection.

The reports from Chemistry and Toxicology in which no issue was identified during the initial review were sampled for data collection using the number of pages of data associated with the report as an initial proxy for the potential complexity of the analysis or the number of samples analyzed.

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

	Chemistry	Toxicology
Pages of Data Associated with Report	# Reports	# Reports
>200	46	0
100 - 200	135	30
76 - 99	110	73
51 - 75	280	176
41 - 50	196	113
31 - 40	288	158
21 - 30	425	148
16 - 20	372	59
11 - 15	460	100
<11	119	374

The number of evidence items related to the request for Latent Prints and Firearms analysis was used in those disciplines as the parameter for selecting reports where no rejection occurred.

Latent Prints	
#Evidence Items Related to Request for Analysis	# Reports
>20	9
11 – 20	15
6 – 10	55
<6	193

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

Firearms	
#Evidence Items Related to Request for Analysis	# Reports
>30	11
21-30	13
11-20	16
<11	170

Possible contributing factors to identification of an issue during the review process were examined. Statistical analysis of this data was used to identify any significant correlations between the various factors and the probability that the case documentation or report required correction prior to release. Examples of the factors to be considered were evidence type, quantity of evidence, and number of repeated or different analyses conducted.

A study to evaluate the effectiveness of the peer review process through re-review of reports and data was designed. The datasets from each discipline were provided to statisticians who conducted statistical tests to determine if any correlations existed between the various parameters and the probability that an issue would be identified during review of the case data and report.

Data Analysis:

Nearly every parameter was found to be correlated to potential issues requiring correction, but two parameters had very strong correlations across all statistical tests: the number of pages of data generated during the analysis and the number of items analyzed. The majority of issues identified during the review process were administrative in nature, e.g., missing initials on a notation or

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

missing information identifying the particular lot or batch of a reagent used in the testing process. The technical issues identified were very rarely linked to an incorrect conclusion. The most common technical issues were a need for additional testing to confirm a conclusion or additional documentation in the case file.

Although there were predictive correlations between the parameters studied for the Latent Prints and Firearms reports and the percentage of report rejections, it was decided that the laboratory would not pursue modification of the peer review practices in the Latent Prints and Firearms disciplines at the present time. This decision was partly due to the gravity of the types of cases which are being analyzed within those disciplines and partly due to the relatively low impact overall on operational efficiency that would be gained due to the low numbers of reports completed each month in those disciplines.

Review of the data from Postmortem Blood Alcohol analyses indicated very low rejection rates during review. The decision was made to release 100% of these types of reports without review following a seven day delay.

Because statistics showed that reports with at least 25 pages of data had a much higher probability of being rejected during review than those reports with less than 25 pages of data, the LIMS was modified to require the review of all drug identification reports with 25 pages or more of data and 10% of all remaining reports. The LIMS was set up to auto-release the remaining 90% of reports.

FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in Publicly Funded Laboratories

In order to determine if the operational changes were effective, further data sets were constructed following implementation of these peer review changes to answer the following questions:

- *Does reducing the amount of peer review an analyst conducts allow for a more effective peer review?*
- *Does limiting the amount of peer review being performed to only those cases where there is an increased likelihood of an error being found allow the analyst to complete the less complex cases with fewer errors – do they complete these cases with fewer errors if they know no one is coming behind them to review?*

Project Findings:

Because Postmortem Blood Alcohol reports showed an extremely low rejection rate, potentially due to the highly automated process with stringent operational checks that minimize the risk of reporting errors, the follow up study was limited to drug identification reports.

From the set of drug identification reports not originally reviewed prior to release by another analyst, a randomly selected set (n = 345) was distributed to technical staff for review. Additionally, a set of reports (n = 483) previously reviewed prior to being released from the laboratory was reviewed a second time, as in the initial re-review. Both of these sets were equivalent to

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

approximately 10% of cases released during a specified time frame. Statisticians compared these data sets to the baseline sets of data.

Does reducing the amount of peer review an analyst conducts allow for a more effective peer review?

Statistics did not indicate a significant change in errors identified in the re-review following the operational change that resulted in reduced peer review.

Does limiting the amount of peer review being performed to only those cases where there is an increased likelihood of an error being found allow the analyst to complete the less complex cases with fewer errors – do they complete these cases with fewer errors if they know no one is coming behind them to review?

Yes - Review of the statistics demonstrated a statistically significant ($P < 0.0001$) decrease in errors in reports with less than 25 pages of data following the operational change as compared to the percentage of reports with errors when 100% were being reviewed. There was not a statistically significant change in errors identified in reports with 25 pages or more of data.

Based on these findings, this laboratory continues to review all drug identification reports with 25 or more pages in the technical record and 10% of reports with less than 25 pages in the technical record. All reports completed by analysts in supervised casework are reviewed, and all reports not previously reviewed are reviewed prior to court testimony. This protocol will continue to be

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

monitored, and the auto-release of cases will be halted or adjusted as needed to maintain the highest quality of work product.

Implications for Criminal Justice Policy and Practice in the United States:

This study demonstrates the impact that reduced review has on the peer review process for one discipline of one laboratory. The benefits of a decrease in errors on cases released without a review and the time saved by reducing or eliminating peer review, combined with the absence of data indicating that the process is detrimental to the quality of the work product, may prompt other laboratories to conduct their own evaluations.

Summary:

Although the review and verification of forensic casework data and reports are critical elements in ensuring the quality of the final work product, the established requirements for the level of review and verification are not based on scientific studies demonstrating the effectiveness of the review process.

The scope and quantity of review and verification that the laboratory establishes should be based on relevant studies that demonstrate the effectiveness of the process.

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

This laboratory has implemented procedural changes for drug identification reports, focusing the review on reports that contain 25 pages or more of data. The protocol continues to be evaluated. Should the effectiveness of this process begin to diminish as scientists become accustomed to this level of review, 100% review may again be warranted.

Results:

This study shows that limiting the amount of peer review based on a review of statistical data could lead to less time spent on peer review by the scientist as well as a more effective peer review.

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

Project 2:

**Evaluation of an Enhanced Screening Technique for Toxicological Specimens with High
Resolution Accurate Mass Spectrometer (HRAMS)**

Purpose:

The Georgia Bureau of Investigation (GBI) Toxicology discipline currently performs drug screenings by a combination of enzyme immunoassay (EIA) and liquid chromatography tandem mass spectrometry (LC-MS/MS). Enzyme immunoassay screens for five classes of drugs: opioids, cocaine and cocaine metabolites, cannabinoids, barbiturates, and benzodiazepines. Liquid chromatography tandem mass spectrometry analyzes for approximately 180 drugs of interest. The information gathered from this screening process is used to guide confirmation testing of drugs of interest, which include prescription, over the counter, illegal, and designer drugs.

The GBI has purchased a Liquid Chromatography Q Exactive Focus™ High Resolution Accurate Mass Spectrometer (HRAMS). This instrument is expected to allow for increased screening capabilities, thus streamlining the front end process from a two instrument analysis to a single instrument screen. Additionally, the processing of data generated by the HRAMS has the potential to be automated. The manual processing of an average run of samples currently performed on the LC-MS/MS takes approximately 10 hours.

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

This evaluation will determine the capability of high resolution accurate mass spectrometry to streamline and automate the screening process performed at the GBI by comparing results of the HRAMS to those obtained by EIA and LC-MS/MS screens.

Project Design and Methods:

A method was created for sample preparation, data acquisition, and data processing of whole blood specimens using the HRAMS and a validation plan was created. This plan addressed such factors as matrix effects (ion enhancement/suppression), percent recovery, and interferences.

The detection limits for compounds were evaluated and compared with LC-MS/MS and EIA levels to ensure that the high resolution accurate mass spectrometer is able to detect forensically significant concentrations of drugs of interest.

Concordance studies were performed to obtain data that was used to evaluate the effectiveness and efficiency of the HRAMS as compared to the current two-instrument screening process.

Data Analysis:

In accordance with the validation plan, matrix effects (ion suppression/enhancement), percent recovery, limits of identification (LOI) and interferences were evaluated to assess method suitability. Matrix effects were calculated by determining the difference between aliquots that were un-extracted and those of a matrix that was extracted and fortified post-extraction. The percent recovery was calculated by comparisons of un-extracted aliquots to those of aliquots for fortified

FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in Publicly Funded Laboratories

samples that went through the entire extraction. The LOIs were determined by the concentration at which a high resolution accurate mass peak was detected and a mass spectrum was generated that produced a match to an internally created library. Interferences were evaluated by injecting all known isomers and determining the resolution between the different analytes.

Project Findings:

LOIs for all analytes were found to be acceptable for the purposes of this procedure. Most analytes generated detectable results at approximately 2 µg/L. These results are similar if not improved from those of the LC-MS/MS and EIA analysis, where generally analytes have detection limits in the 5-10 µg/L concentration.

The evaluations determining matrix effects (ion suppression/enhancement) demonstrated that for most samples ion suppression or enhancement is negligible. The results for some analytes are enhanced by the HRAMS method, yielding up to 2-5 times the expected responses.

Percent recovery for the analytes were typically in the 60-70% range, with some being greater than 100% when ion enhancement is involved. This recovery is deemed acceptable, as all analytes returned an LOI that was acceptable for the purposes of this procedure.

The interference evaluation of isomers demonstrated that most analytes were sufficiently resolved and had a unique enough mass spectrum that identifying them even in samples where both analytes are present is possible. While typically a resolution greater than 1.5 demonstrates full resolution,

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

it was determined that a resolution of approximately 0.4 or greater was sufficient for the instrument to correctly identify the presence of two distinct peaks and identify both correctly. Since this method is not quantitative, but rather a qualitative screening method, the 0.4 resolution is appropriate. The three sets of isomers that were not fully resolved are amitriptyline/maprotiline, crotonyl fentanyl/cyclopropyl fentanyl, and fluoroisobutyryl fentanyl (FIBF)/parafluoroisobutyryl fentanyl. If both isomer forms of the drug were to be present in the same case, the instrument would most likely be unable to determine that both were present.

As of the completion of this overview, a concordance study is underway comparing the findings of the currently used LC-MS/MS and EIA with those of the HRAMS. At this time 156 case samples have been analyzed. The only differences up to this point are attributed to different sensitivities of the two screening techniques. Both screening techniques are very sensitive for some analytes, leading to some slight differences in what is detected in either technique.

The average processing time for concordance samples run on an LC-MS/MS was shown to be approximately 10 hours. This process requires a scientist to manually review all of the transitions detected, print out library matches, and label the detected compounds with their respective retention times. The EIA data processing and instrument preparation time would add approximately one hour to the process.

The average time to process this data on the HRAMS using Trace Finder Forensic software was approximately 30 minutes.

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

Implications for Criminal Justice Policy and Practice in the United States:

HRAMS can provide an alternative for laboratories that use a multi-step process for front end drug screening. Reducing the technology used in drug screening lowers the cost of maintaining multiple instruments. With advanced data processing techniques, the time saved by laboratory personnel will ultimately result in a cost savings and timely reporting of results to customers.

Summary:

An analysis into the ability of the Georgia Bureau of Investigation to replace the current two-instrument method of screening using LC-MS/MS and EIA with an automated one instrument method utilizing HRAMS was considered. The method was evaluated to determine if it was fit for purpose testing matrix effects, percent recovery, limits of identification, and interferences, and by performing a concordance study. The results show that the HRAMS is capable of detecting all the same analytes and at equivalent concentrations as that of the currently utilized LC-MS/MS and EIA. The time saving ability to process the data in an automated way is the greatest benefit of moving front end screening for LC-MS/MS and EIA to HRAMS. On average the HRAMS can save approximately 9.5 hours of processing time.

Results:

The HRAMS analysis has shown to be capable of performing the duties of the combined analysis utilizing LC-MS/MS and EIA. Based on the results of this study, the HRAMS analysis has the

**FY16 Research and Evaluation for the Testing and Interpretation of Physical Evidence in
Publicly Funded Laboratories**

potential to meet the needs of the GBI and yield a significant increase in productivity and improved reporting timeliness.