Document Title:	**Smokeless Powder Reference Collection and SWGFEX Smokeless Powders Database Expansion**
Author(s):	**Michael E. Sigman, Ph.D.; Mary Williams, M.S.; Dana-Marie K. Dennis, Ph.D.**
Document Number:	**250933**
Date Received:	**July 2017**
Award Number:	**2013-R2-CX-K008**

Smokeless Powder Reference Collection and SWGFEX Smokeless Powders Database Expansion

Award No. 2013-R2-CX-K008

Author/ Principal Investigator	Michael E. Sigman, Ph.D. Department of Chemistry and National Center for Forensic Science University of Central Florida PO Box 162367 Orlando, FL 32816-2367 407-823-6469
Co-Authors	Mary Williams, M.S. Dana-Marie K. Dennis, Ph.D. National Center for Forensic Science University of Central Florida PO Box 162367 Orlando, FL 32816-236

Abstract

In 1998, the National Research Council issued a report "Black and Smokeless Powders: Technologies for Finding Bombs and the Bomb Makers". [1] The NRC report recommended that a comprehensive national powder database be developed, containing information about the physical characteristics and chemical composition of commercially available black and smokeless powders. An agency-independent effort to develop a smokeless powders database did not emerge until 2009, when the National Center for Forensic Science (NCFS) in collaboration with the Scientific Working Group for Fire and Explosions (SWGFEX) began work on an internet-accessible database of analytical information on smokeless powders (http://www.ilrc.ucf.edu/powders/). The database opened in early 2011 with 100 entries of legacy powders, provided by Mr. Ronald Kelly (FBI-retired). Under this research grant: (1) records for an additional 600 powders were entered into the database, (2) 100 new powders were analyzed and their records entered into the database, (3) reference collections of the 100 smokeless powders were provided, free of charge, to 50 forensic laboratories that conduct smokeless powder exams and, (4) the data corresponding to the legacy and new records were utilized in research to establish evidentiary and investigative value associated with matching records returned from a search of the Smokeless Powders Database.

Contents

Executive Summary

The Problem

In 1998, the National Research Council (NRC) issued a report "Black and Smokeless Powders: Technologies for Finding Bombs and the Bomb Makers" [1]. The Committee on Smokeless and Black Powder, which prepared the report, was appointed in response to a mandate in the Antiterrorism and Effective Death Penalty Act of 1996 to address the key issues of: (1) feasibility of adding tracers to smokeless and black powder to aid in detection and, (2) feasibility of adding tracers for sourcing black and smokeless powders. Key findings in that report included the fact that 90% of the deaths and 80% of the injuries caused by pipe bombs that contain black and smokeless powders occurred away from security screening points. This finding places a heavy emphasis on the forensic investigation to help identify the explosive and bomb maker.

The NRC report also pointed out that the materials recovered at bombing scenes (e.g. unburned powder, etc.) can aid in the identification of the product type and manufacturer, provided the appropriate resources (databases, spectral libraries, etc.) are available to forensic scientists. The report also pointed to the fact that existing databases on black and smokeless powders were used extensively, at that time, in bombing investigations, yet they were incomplete and it was difficult to match powders from a scene to database entries. Large federal organizations, such as the Federal Bureau of Investigation, the Bureau of Alcohol, Tobacco, Firearms and Explosives and the U.S. Postal Service had the resources to develop databases; however, smaller state and local laboratories, which also need access to these resources, often did not have the time and manpower to meet this need.

In addition to the need for a comprehensive database on smokeless powders, members of the SWGFEX Explosives Database Committee expressed the need for a set of reference materials for laboratory comparisons and training.

A more research oriented aspect of the problem was the need to assess the evidentiary or investigative value of a match between a questioned sample and a sample in the database. Implementation of probabilistic methods is heavily reliant on defining a relevant population. Defining a relevant population is aided by a survey of casework samples and the further expansion of the smokeless powders database to allow construction of a relevant population data source on which to base likelihood ratio methods. Estimation of the relevant population in this and related forensic problems remains a research problem. The results from this research are based on the population of the Smokeless Powders Database. Nonetheless, these results are important in demonstrating the methodologies under this constraint. Evaluation of the methodology and publication under peer-review was an important part of addressing this industry need.

Purpose of the Research

This research directly addresses the need for an extensive and searchable database of smokeless powders that was identified in the 1998 NRC report. The NRC's recommendation read, in part: "A comprehensive national powder database containing information about the physical characteristics and chemical composition of commercially available black and smokeless powders should be developed and maintained."[1] While various agencies, or individuals within agencies, created databases for their own use, a nationwide agency-independent effort did not emerge until 2009, when the National Center for Forensic Science (NCFS) in collaboration with the Technical and Scientific Working Groups for Fire and Explosions (TWGFEX/SWGFEX) began work on an internet-accessible database of analytical information on smokeless reloading powders (http://www.ilrc.ucf.edu/powders/). One purpose of this work was to expand the NCFS Smokeless Powders Database to contain over 800 records covering both legacy and newly purchased smokeless powders. As a result of meeting this goal, the Smokeless Powders Database would serve as a resource for local, state and federal law enforcement. The information would be freely available on the internet and therefore the database would also be a criminal justice tool for defense experts and the court.

Another purpose of this research was to address the need for reference materials by providing a set of 100 smokeless powder samples to each of 50 ASCLD/LAB accredited laboratories. Forty two (42) of the 50 laboratories accepted the samples. Data from the NCFS analysis of each reference material was entered into the Smokeless Powders Database and the references were provided along with the sample reference number for the corresponding database record.

Finally, a purpose of this research was to utilize the Smokeless Powders Database to provide a statistical assessment of the value of a match between a database record and the physical and chemical properties of a questioned sample, as the result of a database query. The need for this aspect of the research is clear, based on reports of matches to an individual product as the result of a search on other databases. It is easy to assume that the short list returned by a database query contains one or more powders made by the same manufacturer. However, over the years, several mergers and acquisitions of one smokeless powder manufacturer by another have occurred. The process of relabeling a powder prior to distribution further complicates the issue. While the impact of these industry practices are not directly addressed, the study presented here is based on a knowledge of the ground-truth manufacturer and provides an upper limit for the evidentiary or investigative value of a query-match.

Research Design

The Smokeless Powders Database opened in early 2011 with 100 entries. An additional 538 data sets from the FBI were available for uploading. In addition, at the time of the original proposal, NCFS had undertaken the task of maintaining the database, and performing the chemical analysis and data entry for new samples of smokeless powders following a standard operating procedure (SOP), collaboratively developed by SWGFEX and NCFS. Eighty-seven powders had been analyzed by NCFS and the associated data entered into the database. Under this proposal, FBI data from the remaining 538 legacy powders were uploaded into the database. An additional 100 powders, purchased during the performance period, were analyzed and the data was entered into the Smokeless Powders Database.

After each of the 100 new samples was analyzed, the remaining material was dispersed into 50 sample vials (5 dram each). The sample vials from all 100 samples were assembled into 42 sets of reference materials for distribution to accredited laboratories. The total number of vials prepared for distribution was 4200.

Following the entry of both new and legacy data records into the Smokeless Powders Database, cross validation was used to test the accuracy and evidentiary value of matches between a test sample and records returned by query. Two approaches were taken, one that relied entirely on discrete data (i.e., matching kernel shape, color, presence/absence of a perforation, measurement overlap, and matching subsets of 13 chemical analytes). The second method utilized Fisher transform of the correlation between the total ion spectra (average mass spectra across the chromatographic profiles) as a continuous score for comparison of powders, conditioned on the requirement of same shape, color, state of perforation and size. The continuous score based method was enhanced by further conditioning comparisons on membership of the two powders in the same clusters defined by the major ions in the total ion spectra.

A second study was conducted for the purpose of examining the accuracy of predicting the manufacturer of a smokeless powder based on the chemical and physical properties of the material. A Bayesian network was utilized in this work to calculate the posterior probability of each of 6 different manufacturers represented in the powder set analyzed by NCFS. The class conditional probabilities were calculated based on frequencies in the dataset and the sample was assigned to the manufacturer with the largest calculated posterior probability.

Findings and Conclusions

All of the records for the legacy powders were entered into the database and are available online. Similarly, all of the analyses were completed for the 100 new powders and the records were made available in the database for use by the forensic community. The reference material collections were sent to 42 laboratories.

Analysis of the evidentiary or investigative value of a query match between a questioned material and database entry was shown to be fairly low by either a discrete match of properties (likelihood ratio of 2.81), or a match based on a continuous score derived from the Fisher transform of the Pearson correlation coefficient between the total ion spectra from the questioned and query-returned samples (likelihood ratios below 5 for score values below 5). Both of these likelihood ratios provide weak evidence for the observation of matching criteria given the same manufacturer as opposed to different manufacturers.

Implications for Policy and Practice

The expansion of the Smokeless Powders Database has provided additional data on smokeless powders that assists the analyst in assessing casework on a daily basis. The reference collections provide a physical link to the database. These advances in data accessibility eliminate the need for individual databases in each laboratory and provides ASCLD/LAB accredited laboratories with smokeless powder references that they would not otherwise be able to attain. The results from probabilistic assessment of the evidentiary value of matching physical characteristics and chemical composition informs the community of the value of such comparisons for the purpose of investigative and probative applications. The generally low evidentiary value of these “matches” should influence practice and future policy regarding the introduction of smokeless powder evidence. The use of Bayesian networks to predict the manufacturer of a smokeless powder sample based on chemical and physical characteristics was shown to provide a nearly 80% correct prediction rate, which exceeds the 17% random probability based on a flat prior, or 52% probability of correctly assessing the manufacturer if the analyst always guesses Hodgdon Powder Company. These results provide an incentive for using these statistical methods in the investigative process.

I. Introduction

Statement of the problem

The National Research Council (NRC) issued a report "Black and Smokeless Powders: Technologies for Finding Bombs and the Bomb Makers" in 1998 [1]. Key findings in that report included the fact that 90% of the deaths and 80% of the injuries caused by pipe bombs that contain black and smokeless powders occurred away from security screening points. This finding places a heavy emphasis on the forensic investigation to help identify the explosive and bomb maker. The report also pointed out that the materials recovered at bombing scenes (e.g. unburned powder, etc.) can aid in the identification of the product type and manufacturer, provided the appropriate resources (databases, spectral libraries, etc.) are available to forensic scientists. The report also pointed to the fact that existing databases on black and smokeless powders were used extensively, at that time, in bombing investigations, yet they were incomplete and it was difficult to match powders from a scene to database entries. Large federal organizations, such as the Federal Bureau of Investigation, the Bureau of Alcohol, Tobacco, Firearms and Explosives and the U.S. Postal Service had the resources to develop databases; however, smaller state and local laboratories, which also need access to these resources, often did not have the time and manpower to meet this need.

In 1998, the NRC stopped short of recommending the implementation of a tagging system for black and smokeless powders due to the lack of information regarding safety issues associated with the addition of tags to these products. However, the NRC's recommendation read, in part: "A comprehensive national powder database containing information about the physical characteristics and chemical composition of commercially available black and smokeless powders should be developed and maintained."[1] While various agencies, or individuals within agencies, created databases for their own use, a nationwide agency-independent effort did not emerge until 2009, when the National Center for Forensic Science (NCFS) in collaboration with the Technical and Scientific Working Groups for Fire and Explosions (TWGFEX/SWGFEX) began work on an internet-accessible database of analytical information on smokeless powders (http://www.ilrc.ucf.edu/powders/). The database opened in early 2011 with 100 entries based on data provided by Mr. Ronald Kelly (FBI-retired). An additional 538 datasets were available for uploading into the system from the FBI. In addition, at the time of the original proposal, NCFS had undertaken the task of maintaining the database, performing the chemical analysis and data entry for new samples of smokeless powders following a standard operating procedure (SOP) collaboratively developed by SWGFEX and NCFS. The SOP is given in Appendix A. Eighty-seven new powders had been analyzed by NCFS and the associated data entered into the database. Entry of the remainder of the legacy data needed to be entered into the database.

NCFS also maintains the Ignitable Liquids Reference Collection and Database (ILRC) in collaboration with the SWGFEX Committee on ignitable liquids. Samples of each ILRC database entry are retained by NCFS and can be obtained by fire debris analysts upon request, which effectively makes the ILRC a “reference collection”. However, due to limits placed on the storage of smokeless powders by ATF and the Florida Fire Marshal’s Office, NCFS could not retain samples of all of the powders in the database to serve as a reference collection for forensic laboratories. Rather than dispose of newly purchased and analyzed samples under this proposal, we proposed to provide a “reference collection” of analyzed powders to ASCLD/LAB-accredited laboratories that perform smokeless powders analysis. Reference collections carry no requirement for traceability; however, the value of a reference collection is enhanced when the samples are cataloged in a larger database and when forensic laboratories can locate database samples with properties similar to the properties of the reference collection. Individualization of smokeless powder samples is not currently possible, and in the event that individualization should ever become a possibility, the continually changing pool of available powders would prevent accurate estimates of the probability of randomly encountering a specific powder. Knowledge of the amount produced and sold for a specific powder does not directly address this issue since use (detonation) consumes the powders and therefore the distribution of powders in the pool of available powders is continually in flux. Nonetheless, certain combinations of chemical and physical properties may prove to be relatively rare in the manufacture of smokeless powders in particular. Determination of the evidentiary or investigative value of matching chemical and physical properties of smokeless powders can be accomplished using the Smokeless Powders Database. In achieving this goal, it will be necessary to complete the database entry of records for 600 legacy samples and records from the analysis of 100 newly purchased samples. Regardless of the success in establishing a statistical basis for evaluating the value of casework samples as a function of chemical and physical properties, the proposed project provided the forensic community with a fully online and highly valuable smokeless powders database and a smokeless powders reference collection.

Literature review

In 2011 the Smokeless Powders Database, http://www.ilrc.ucf.edu/powders, became available online as a collaborative effort between the National Center for Forensic Science (NCFS) and the Scientific Working Group on Fire and Explosions (SWGFEX) Explosives Database Committee. The database contains information on the manufacturer, physical characteristics (size, shape and color) and chemical characteristics (major components). The database contains gas chromatography-mass spectrometry (GC-MS) datasets that can be downloaded by users. Information on the Fourier transform infrared (FTIR) spectra of new data entries are also included in the database. This information represents the data most commonly used in the laboratories by members of the SWGFEX Explosives Database Committee.

Other data types have been shown to be useful in the analysis of smokeless powders, although the analytical methodologies may not be common among all of the SWGFEX laboratories represented on the Explosives Database Committee. Capillary zone electrophoresis (CZE) has been used to measure inorganic ion profiles and characterize burned and unburned smokeless powders[2]. CZE has also been used to simultaneously detect both cations and anions from smokeless powders [3]. Micellar electrokinetic capillary chromatography (MEKC) has been successfully applied to the separation of organic components of smokeless powders and compared with separations performed by high performance liquid chromatography (HPLC) [4]. HPLC and MEKC gave comparable precision of elution and migration times; however, the highest reproducibility of peak area was exhibited by MEKC. The sensitivity and identification power of MEKC was improved by interfacing the instrument to a diode array UV detector. In another study, ultrasonic solvent extraction was used to recover nitroglycerin (NG), diphenylamine (DPA), and ethyl centralite (EC) for capillary electrophoresis analysis in less than one hour [5].

There have been several reports on the use of HPLC coupled with MS detection for the analysis of smokeless powders. In addition to the comparison of HPLC with MEKC [5], HPLC with electrospray ionization mass spectrometry detection (HPLC-ESI-MS) has been characterized for the analysis of selected smokeless powder stabilizers to optimize chromatographic analysis of these compounds [6].

Vibrational spectroscopies can be useful in the analysis of smokeless powders; however, Raman analysis can lead to ignition of smokeless and black powders [7].

The presence of smokeless powders have been shown to be identifiable by ion mobility with the aid of a planar solid phase microextraction air sampling device [8, 9]. The field detection of smokeless powders can be facilitated by standards developed by the National Institute of Standards and Technology [10].

There have been several reported studies that focused on reasonable and reliable analysis of smokeless powders by identification and quantitation of the propellant and stabilizers. In one study, quantitative measurements of nitroglycerin (NG), ethyl centralite (EC), diphenylamine (DPA) and N-nitrosodiphenylamine (NnDPA) were reported by five participating laboratories [11]. In this international measurement comparison exercise, NG was found to be the most unstable analyte. Participants self-reported a 1 – 5% measurement uncertainty, whereas a 5 – 50% uncertainty was reported among four concordant participants. In a follow-up study, the propellant to stabilizer ratio (P/S) was tested as a method of establishing a positive association between organic gunshot residue and 38 caliber ammunition powders. It was determined that the average and dispersity of the P/S values provided information for the association or exclusion of powder samples. Comparison of measurements from residue particles with those of residue samples collected from test firing of the same weapon and ammunition was found to be a useful technique to account for changes in composition that occurred during the firing of the weapon [12].

It is important to note that all of the comparative studies reported here were done on relatively small sample sets and the results of those studies are <u>not</u> representative of what can be expected when searching for a "match" or set of lead powders contained in a database of several hundreds or thousands of records. The NCFS/SWGFEX database currently uses a logical AND search routine that utilizes identification of smokeless powder shape, average measurement or maximum and minimum measurement, organic propellants and stabilizers present. Combining several properties (i.e., shape and phenylamine, etc.) limits the number of returned records; however, the independence of the frequency of property occurrence within the database and the value of the combined frequencies have not previously been established.

Smokeless powders are well-characterized materials comprised of the energetic component nitrocellulose (single base powder), along with nitroglycerin (double base powder) or nitroglycerin and nitroquanidine (triple base powder) [13]. Smokeless powders also contain a number of other organic and inorganic components that modify the performance characteristics of the materials [13-15]. It is the single and double base powders that are commercially available in the U.S. and these materials are often used in pipe bombs.

As previously described, the Smokeless Powders Database[16], contains data from analytical methods commonly used in the forensic analysis of these materials, including stereomicroscopy,[17] Fourier-transform infrared (FT-IR) spectroscopy,[18, 19] and gas chromatography-mass spectrometry (GC-MS).[20] The chemical and physical properties of intact powders, utilized in a database search, have been reported to facilitate the identification of an individual product or a short list of potential products.[17, 21] The certainty with which product identifications can be made has not been determined statistically and previous assertions are not grounded in peer-reviewed research. One aspect of the research conducted under this grant was to investigate the evidentiary and/or investigative value of smokeless powder identifications based on a search of the 726 samples in the Smokeless Powders Database.[22]

There have been several peer-reviewed reports on the differentiation of limited sets of smokeless powders. In 1989, an analysis of 12 smokeless powders from three different manufacturers was attempted using pyrolysis gas chromatography[23]. Each powder represented different lots of a single product. The distributions of the correlations between pyrograms resulting from between-manufacturer and within-manufacturer comparisons exhibited overlap, resulting in limited discrimination between manufacturers. A 1992 study of the correlation between the chromatographic profiles of pre-firing and post-firing powders from the same manufacturer and product suggested a possible route to linking shooting victims when the gun and/or projectile were not available[24]. The author of the study suggested, but did not demonstrate, the use of the chromatographic data to determine the manufacturer of a smokeless powder. A 2009 peer-reviewed study of a set of seven smokeless powders by nESI-MS (nanoelectrospray ionization mass spectrometry) and stereomicroscopy was unable to visually discriminate some powders[6]; however, combining the visual inspection, nESI-MS data and extraction efficiency resulted in the ability to discriminate all but two of the powders. In 2013, a set of smokeless powders from five manufacturers were correctly classified by manufacturer based on ESI-MS coupled with chemometric methods[25]. The chemometric methods employed in the study involved first using principal components analysis (PCA) for dimension reduction, followed by k-nearest neighbors classification. In this process, data from a new sample is projected into the PCA space and the probability of class membership is based on the number of (k) nearest neighbors from the known samples. These studies suggest the possibility of discriminating between smokeless powders, or identification of the product and/or manufacturer, based on studies of relatively small data sets. These studies do not provide a probabilistic estimate of the evidentiary or investigative value of the results from a smokeless powder database search. The evidentiary/investigative value question was addressed in this research and the results were published under peer-review [26]. In this report, we provide a synopsis of the methods used in that research and the results obtained. For full details, the reader is referred to our published work.

Hypothesis or Rationale for the Research

The rationale for this research was to meet the needs of the forensic community, as identified by the National Research Council, by providing a publically accessible database of smokeless powders for use in casework and research. An additional need of the community was met by providing reference samples that were indexed in the database. The hypotheses of the research were that the database would assist in casework by providing data to aid in the identification of samples, and establish the statistical significance of comparisons of physical and chemical characteristics among powders. Based on the literature review and statement of the problem, the goals of this research were:

(1) Complete entry of data from 600 legacy samples into the database.

(2) Purchase 100 new smokeless powders, which were analyzed by NCFS and the data entered into the database. The 100 new powders were distributed as a reference collection to ASCLD/LAB accredited laboratories that perform smokeless powders analysis.

(3) Utilize the properties of the new 100 samples and the legacy samples to establish the value of property-based relationships in estimating the uniqueness of a smokeless powder.

II. Methods

Chemical Analysis

The Smokeless Powders Database was developed in 2009 – 2010 by NCFS and T\SWGFEX Explosives committee. Part of the development included creating analytical methods for the measurement of the kernels, gas chromatography – mass spectrometry (GC-MS) analysis, and Fourier Transform – infrared (FTIR) analysis for samples analyzed by the NCFS. The original SOP can be found in Appendix A. Since 2010, NCFS has modified the SOP because both the GC-MS and FTIR spectrometer were replaced with newer instruments in 2013 and 2014, respectively. The internal reflectance element in the new Bruker FTIR is germanium, the wavenumber range is 4000 – 600 cm^{-1}, and the number of scans was increased to 64. During the transition to the new GC-MS, the chromatographic column was replaced; however, the stationary phase and dimensions were the same as the original column. The new SOP can be found in Appendix B.

The powder images for all of smokeless powders in the database were reviewed by the Explosives committee and consensus agreement of the shape for each powder was determined. In the course of assigning powder shape, additional shapes were added to the "shape" search field in the database.

The data used in the statistical aspects of this research were taken from the Smokeless Powders Database. Samples analyzed at NCFS followed the methods and protocols described in Appendices A and B. Legacy samples had been analyzed by source-specific protocols listed in the database.

Likelihood Ratio Approach to Smokeless Powder Comparisons

General Methodology

The Smokeless Powders Database contains nominal discrete data relating to several chemical and physical characteristics of smokeless powders. Nominal discrete data is limited to specified values (i.e., there are a limited number of kernel shape descriptors; disk, ball, flattened ball, etc.). If two powders do not share the same nominal values for a set of properties, the powders differ (a “non-match”). If two powders share the same nominal values for a set of properties, they are not discriminated on the basis of these properties and we designate the powders as a “match”. In this research, we used comparisons of nominal properties shape (s), color (c), presence/absence of a perforation (p), overlapping measurement range (m), and the presence/absence of identical subsets of 13 chemical components or analytes, designated A. Other properties recorded in the Smokeless Powders Database, such as distinguishing features and marker colors, are so infrequent or subjective that they were not used to condition record comparisons in this work. In a second comparison method, rather than examining the correspondence of subsets of 13 chemical components, the chemical comparison was based on a similarity score measure. The Fisher transform of the Pearson correlation of the total ion spectra, designated as z, was used as the similarity score. Pairwise comparisons were made between 726 records contained in the Smokeless Powders Database. The evidentiary/investigative value of a comparison was assessed as a likelihood ratio value. The likelihood ratio discussed here addresses propositions at the source level [27, 28], where the source is defined under the prosecution’s proposition, H_P, as the same product (P), with additional conditioning requirements of same kernel shape, color, etc., as described above. Under the defense’s proposition, H_D, the respective statement is that the two powders are not the same product ($\bar{P}$), again with the additional conditioning requirements. Designating two powders as the same product implicitly requires that they correspond to the same manufacturer, but does not speak to the same or different production lot. The likelihood ratio is designated in equations 1 and 2.

$$LR = \frac{Pr(A|P,s,c,p,m)}{Pr(A|\bar{P},s,c,p,m)} \tag{1}$$

$$LR = \frac{f(z|P,s,c,p,m)}{f(z|\bar{P},s,c,p,m)} \tag{2}$$

The numerator of equation 1, is the probability of observing the same subset of the 13 specified chemicals, A, given the proposition of the same product, P, and matching s, c, p and m nominal discrete properties. The denominator of equation 1 is the probability of observing the same subset of 13 specified chemicals given the proposition of different product, $\bar{P}$, and matching s, c, p and m nominal discrete properties. Equation 2 expresses a similar ratio of probability density estimates, *f*, of the distributions of *z*, given the propositions and matching nominal discrete values as described above. The significance of equation 2 is shown graphically in Figure 1. In this figure, the curve on the left represents the probability density estimate of *z* values that come from different products that have the same shape, color, perforation and measurements, i.e., the dashed curve represents the function $f(z|\bar{P},s,c,p,m)$. This function of *z* reflects the probability of observing different values of *z* for comparison of different products. The solid curve on the right represents the probability density estimate of *z* values that come from same product comparisons, i.e., the solid curve represents the function $(z|P,s,c,p,m)$. This function of *z* reflects the probability of observing various values of *z* for comparison of samples of the same product. The vertical red line represents a value, *z*', that was determined for two powders that had matching s, c, p and m nominal discrete properties; however we do not know if these two powders are the same product. The likelihood ratio is determined by the ratio of the height where the vertical line intersects the solid curve, divided by the height where the vertical line intersects the dashed curve. In this example, the likelihood ratio will be greater than one. The meaning of this ratio is that “it is more likely to observe a score of *z'* if the two powders are the same product, than if they are different products”. Similarly, if *z'* were smaller (i.e., the vertical red line were further to the left), the height where the vertical red line intersects the dashed curve may be larger than the height where it intersects the solid curve. In that case the likelihood ratio would be less than one and the meaning would be that “it is more likely to observe a score of *z'* if the two powders are different products”. These two results do not mean that it is either more or less likely that the powders are the same product or different products! To answer the question of the relative probabilities that the powders are the same product or different products, you must know the prior probability of two powders being the same product or different products. You need equation 3, the odds form of Bayes’ equation.

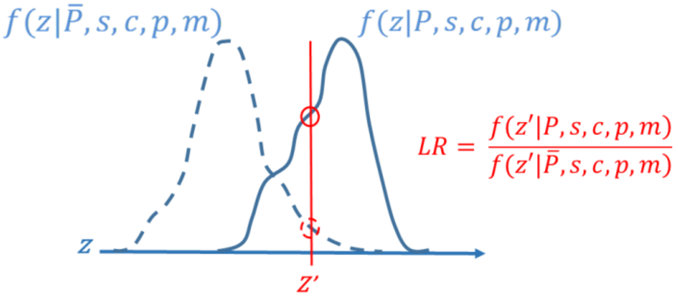

Fig. 1. Shows the probability density estimates of z values for different product (dashed curve) and same product (solid curve) comparisons

$$\frac{Pr(P|z,s,c,m,p)}{Pr(\bar{P}|z,s,c,m,p)} = \frac{f(z|P,s,c,p,m)}{f(z|\bar{P},s,c,p,m)} \times \frac{Pr(P|s,c,p,m)}{Pr(\bar{P}|s,c,p,m)} \tag{3}$$

In equation 3, the first term on the left is the posterior odds, which is the odds of two samples being the same product versus different products given the value of z and the same s, c, p and m nominal discrete values. The second term is the likelihood ratio, discussed above. The third term is the prior odds, which is the odds of two powders being the same product versus different products given that they have the same s, c, p and m nominal discrete values. Note the prior odds does not take into account the value of z. For example, if it were 100 times more likely for same product powders (rather than different products) to share matching s, c, p and m, the prior odds would be 100. If you measure z and determine your likelihood ratio to be 10, then the posterior odds are 1,000 (10 x 100). The odds that the two powders were the same product would be 1,000. On the other hand, if you measure a value of z corresponding to a likelihood ratio of 1×10^{-4} , the odds that the two powders were the same product would be 0.01 ($100 \times 1 \times 10^{-4}$).

The prior and posterior odds are generally accepted to be within the purview of the court, and should not be addressed by the forensic scientist. The likelihood ratio is generally accepted to be within the purview of the forensic scientist and this is the only term in equation 3 that should be addressed by the forensic scientist.

Limiting smokeless powder comparisons

The values for the nominal discrete variables for each record in the Smokeless Powders Database were tabulated in a comma-separated values file, along with the manufacturer and product designations. The values were read into R – statistical software and unique pairwise comparisons were performed [29]. The number of "match" and "non-match" comparisons were counted and tabulated for same-product and different product comparisons. The comparisons are given in tabular form in the results section.

Discrete chemical and physical data

The presence of 13 organic compounds (2,4-dinitrotoluene, 2-nitrodiphenylamine, 2-nitrosodiphenylamine, 4-nitrodiphenylamine, 4-nitrosodiphenylamine, dibutyl phthalate, diethyl phthalate, dioctyl phthalate, diphenylamine, ethyl centralite, methyl centralite, nitroglycerin, dipentyl phthalate), commonly observed in smokeless powders, were determined for each of 726 smokeless powders based on the database records. When two powders share the same subset of components (i.e., both samples contain only diphenylamine, ethyl centralite and nitroglycerin), they are designated as a "match", so long as they have the same values for the nominal discrete characteristics. If two samples have the same values for the nominal discrete characteristics, but contain different subsets of the 13 organic compounds, they are designated a "non-match". The comparison is limited to the 13 organic compounds that are tabulated in the database. The output of each comparison is discrete ("match" or "non-match") and the number of each discrete outcome is counted for each of the two ground-truth cases, same-product and different-product. The values are tabulated in a 2x2 confusion matrix, as shown in Table 1. The values in a 2x2 confusion matrix can be interpreted to determine a true positive rate (TPR = TP/ΣP) and false positive rate (FPR = FP/ΣN).[30] The slope of a line passing from the intercept (0,0) in ROC space through the single point (FPR, TPR) is equivalent to the likelihood ratio[31, 32], and is a measure of the evidentiary/investigative value of a returned "match" from a database search.

Table 1. Shows the number of discrete outcomes for each of two ground-truth states, same-product and different product

		Comparison of Organics		
		"match"	"non-match"	Total
Ground-Truth	Same Product	TP	FN	ΣP
	Different Product	FP	TN	ΣN
	Total	Σp	Σn	

Continuous chemical composition data

The total ion spectra from 43 – 400 m/z were calculated for data from GC-MS analysis of 726 smokeless powders (87 analyzed at NCFS and 639 analyzed by FBI) [33]. The TIS is equivalent to the average MS across the chromatographic profile with normalization to the base peak, and has been useful for the analysis and modeling of ignitable liquids and fire debris [33-38]. Undecane has been added before GC-MS analysis of the smokeless powders in the database which originated from NCFS or FBI. While the concentration of undecane is known in the NCFS samples, it is not known for the FBI samples. To remove influence of undecane on the score calculations performed in this research, intensities of ions corresponding to undecane (m/z 43, 55, 56, 57, 69, 70, 71, 84, 85, 98, 99, 113 and 156) were removed from the TIS. The remaining ions were used for calculating the score for pairwise comparisons between same and different products, conditioned on the requirement that the compared samples match in the shape, color, perforation and measurement nominal discrete values.

Score-based Likelihood Ratio Calculations

The Fisher transform, z, of the Pearson correlation coefficients, r, was calculated by equation 4 as a similarity score of each pairwise comparison of TIS. The distributions of same-product and different-product scores were modeled using a kernel density estimate based on a Gaussian kernel. All calculations were performed in the R Statistical Software using code written in house [16].

$$z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right) \tag{4}$$

Agglomerative Hierarchical Cluster Analysis

The kernel density estimates of the same-product and different-product scores, calculated as described in the previous section, were found to be bimodal. In an effort to identify the source of the bimodal nature of the distributions, we investigated the possibility of natural groupings of the smokeless powders, beyond the single base and double base designations. Following previous work in the determination of natural groupings in chemical data [39, 40], agglomerative hierarchical clustering, AHC, was used to identify additional groupings of the smokeless powders. AHC calculations were also performed in R [16], using the **amap** [41] package for distance calculations and the **cba** [42] package for optimal leaf ordering in the dendrograms that can be produced from AHC. Optimal leaf ordering was utilized to *"maximize the sum of the similarities of adjacent elements"* [43] in the dendrogram so that highly similar samples were arranged in the center of the overall cluster. The final ordering based on distances between samples is given in a permutation vector. The resulting data can be viewed by plotting the m/z intensities for an ordering along the sample axis determined by AHC [44-52]. The AHC calculations performed in this research are explained in greater detail elsewhere [26].

Cross Validation

The calculated set of z values for same product and different product comparisons were divided into model and test sets by withdrawing 10% of the z values for same product comparisons and 10% of the different product comparison z values to comprise the cross validation test set. The remaining values constituted the model (training) data set. The cross validation draw of test data was performed using a random uniform distribution. The data in the model set was used to construct kernel density functions for the same product and different product z values. Likelihood ratios were calculated for the cross validation test data using the modeled kernel density estimates. Results from the cross validation calculations were compared against the known ground truth (same product or different product) for each sample.

Probabilistic Predictions of Manufacturers using Bayesian Networks

General Methodology

Bayesian networks are probabilistic graphical structures comprised of nodes and arrows. Within the network structure, the nodes which are depicted by circles, represent random variables or events. The causal relationships between the nodes are denoted by arrows to form a directed acyclic graph (DAG); where a node located at the arrow's tail is referred to as a parent node, a node at the arrow's head is referred to as a child node, and a node without an arrow pointing into it is known as a root node. Figure 2 illustrates a simple Bayesian network structure comprising four nodes: **A**, **B**, **C**, and **D**. Nodes **A** and **C** are parent nodes; **A** is the parent of **B** and **C**, and **C** is the parent of **D**. Node **A** is also a root node, since there is no arrow directed into the node. Within the network each node, representing either a discrete or continuous random variable, contains an exhaustive list of mutually exclusive states which can be either categorical, discrete, or continuous [53, 54]. States represent possible outcomes for the random variable, and each state has a probability value ascribed to it. A table associated with each node encodes the probability distribution across all states, or combination of states, within the node. Accordingly, a root node table encodes unconditional probabilities for the states within that node, and the table associated with a child node encodes conditional probabilities for all states within the child node, where each state is conditioned on the states of the parent node(s) [55].

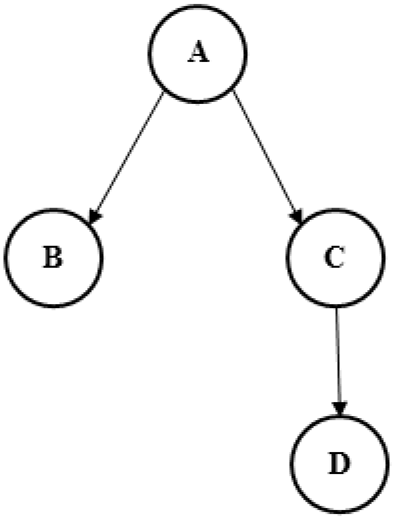

Fig. 2. Shows a simple Bayesian Network structure comprising four nodes, A – D. Nodes A and C are parent nodes, and nodes B, C and D are child nodes.

Referring to Figure 2, the table associated with node **A** encodes unconditional probabilities, while the tables associated with nodes **B**, **C**, and **D** encode conditional probabilities. The probabilities for each state in nodes **B** and **C** are conditioned on each state in node **A**, and the probabilities for each state in node **D** are conditioned on each state in node

C. The probabilities across all states of a child node, conditioned on each state in its parent node, sum to one. A requirement which must be satisfied within the structure of a Bayesian network is that of conditional independence, where two child nodes must be conditionally independent given their parent(s). In Figure 2, nodes **B** and **C** must be conditionally independent given node **A**. Consequently, a Bayesian network describes the probabilistic and independence relationships between a set of random variables by conditioning child nodes upon their parent nodes. Additionally, the joint probability distribution for a set of random variables, $X_1, \ldots., X_n$, is decomposed to the product of their probabilities conditioned on their parents [53, 54]. The expression describing this relationship, known as the *chain rule*, is given in Equation 5.

$$\mathbf{Pr}(\mathbf{X_1}, \ldots., \mathbf{X_n}) = \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Pr}\left(\mathbf{X_i} \middle| \mathbf{parents}(\mathbf{X_i})\right) \qquad (5)$$

Bayesian networks are instantiated to update probabilities of all variable states based on the observation of new evidence. Within the network, a random variable is instantiated by setting one state in the variable's node to certainty; that is, the probability of the state is set equal to 1.

Discrete chemical and physical data

GC-MS data for 726 smokeless powder samples from the NCFS database have been utilized in a previous study [26]. In this study, the smokeless powders GC-MS data were extracted to common data format (CDF) files which were used to generate total ion spectra (TIS) for the samples [26, 56]. Each powder was assigned to a class, or cluster, C1 – C6, resulting from agglomerative hierarchical cluster analysis (AHC). The AHC method is described above and the results leading to the generation of six clusters (based on the most intense ions of *m/z* 46, 120, 134, 149, 165 and 169 in the TIS) are described in the results section of this report. A discrete representation of each sample was generated by first assigning the cluster designation C1 – C6, and then breaking the relative intensities of the observed AHC ions into one-third increments of the base peak intensity. Consequently, if an ion was not observed it was assigned as zero (z); an ion having an intensity on the interval (0, 0.33] was assigned as weak (wk); an ion with a relative intensity on the interval (0.33, 0.67] was assigned as medium (med); and an ion of relative intensity (0.67, 1] was assigned as strong (strg). For example, the TIS of SRN 121 (assigned to C1) had relative intensities of 1, 0.127, 0.005, 0.037, 0.007, and 0.199 for ions *m/z* 46, 120, 134, 149, 165 and 169, respectively. The discrete representation of SRN 121 was C1, strg, wk, wk, wk, wk, wk.

Stereomicroscopic examination of intact smokeless powders enables the determination of the powder's physical attributes such as kernel shape and dimensions. For each Smokeless Powder Database record, kernel dimensions are reported as minimum and maximum measurements of the diameter, and length or thickness, in millimeters. Average and standard deviations of the measurements are also reported for smokeless powders which were analyzed by NCFS. In this work, the average measurements (AM) were converted to discrete data by assigning alphanumeric variables to define specific ranges of measurement. To define the complete range of measurement, the minimum value was set to 0 for kernel shapes for which length/thickness measurements were not recorded, such as ball powders; the maximum value was set to the maximum length observed in the data for cylindrical powders. The specific ranges of measurement were broken into 0.4 mm increments of the maximum length observed. For example, kernel dimensions within the range 0 mm ≤ AM < 0.4 mm were assigned "d1" and "l1" for average diameter and average length, respectively. Table 2 lists the ranges of measurement, and their corresponding alphanumeric assignments. For example, SRN 121 had a diameter (1.6 mm) within the range designated d5 and a length (0 mm – disk shape) within the range designated l1.

Table 2. Shows alphanumeric assignments for average diameter and average length of the smokeless powder kernels

Kernel Dimension (mm)	**Average Diameter**	**Average Length**
0 ≤ AM < 0.4	d1	l1
0.4 ≤ AM < 0.8	d2	l2
0.8 ≤ AM < 1.2	d3	l3
1.2 ≤ AM < 1.6	d4	l4
1.6 ≤ AM < 2.0	d5	l5
2.0 ≤ AM ≤ 2.4	d6	l6

Bayes' Theorem

The computational capacity of Bayesian networks relies on the concept of Bayes' theorem. Within the network, dependence relationships exist between child and parent nodes since the state(s) of the child node depend on the state(s) of the parent node(s). These dependencies, which are expressed as conditional probabilities, are updated as new information is observed in the network. New information is propagated through the network to update the probabilities in other nodes. Bayes' theorem, shown in Equation 6, provides the mathematical framework for calculating these probabilities [55, 57, 58].

$$Pr(\Theta|X) = \frac{Pr(X|\Theta)Pr(\Theta)}{Pr(X)}$$

(6)

Bayes' theorem, which is derived from the third law of probability for dependent events, relates unconditional and conditional probabilities. In Equation 6, the conditioning bar, | , is read "given", the symbol Θ represents the hypothesis, and X represents the data [57, 58]. Equation 6 describes the posterior probability that the hypothesis is true given the data, $Pr(\Theta|X)$, as the product of the likelihood, $Pr(X|\Theta)$, and the prior probability, $Pr(\Theta)$, divided by the marginal likelihood, $Pr(X)$ [58]. Note that the posterior probability and the likelihood are conditional probabilities; whereas, the prior probability and the marginal likelihood are unconditional probabilities. In this work, Θ is the smokeless powder manufacturer, and X represents the chemical and physical characteristics of the smokeless powder sample.

Bayesian Network

As previously discussed, the goal of smokeless powder analysis is to characterize smokeless powders through determination of their physical and chemical features for the purpose of identification. The records in the NCFS Smokeless Powders Database contain chemical and physical information about the powders which is typically used toward smokeless powder characterization. The chemical and physical characteristics described above were utilized in this work to predict smokeless powders manufacturers using Bayesian networks. To determine the network structure, the causal relationships between manufacturers and the chemical and physical characteristics of their smokeless powder products were considered and assigned as random variables. Additionally, the conditional independence relationships between the variables within the proposed network structure were determined using the asymptotic chi-squared test. Conditional independence calculations were performed in R software [29] using the ***bnlearn*** package [59] which was accessed from the Comprehensive R Archive Network (CRAN) repository.

The Bayesian network structure that satisfied the causal relationships and conditional independence requirements was generated using R code written in-house. The DAG was specified using the ***gRbase*** [60] package, and the probability tables were extracted and compiled within the network using the ***gRain*** [61] and ***gRim*** [62] packages which were accessed from the CRAN repository. Within the code, 100 iterations were performed for cross validation of the network. For each repetition, 10% of the data was withheld as the cross validation test set, and the remaining 90% was used to develop and compile the network model comprising the nodes and their associated probability tables. The probabilities for all states within the nodes were calculated based on frequencies within the model data. The network was instantiated by entering the chemical and physical characteristics of each smokeless powder within the test set. Manufacturer posterior probabilities were calculated for each test sample, and predictions were made based on the highest posterior probability. Percent correct rates were determined for each smokeless powder manufacturer by calculating the number of correct predictions; that is, where the ground truth and predicted manufacturer were the same.

III. Results

Historical Records

Ron Kelly from the Federal Bureau of Investigation provided 638 records of smokeless powders they had analyzed over the last 30 - 40 years. In 2010, one hundred of these records were chosen to be the first records in the database (SRN 1 – 100). The remaining 538 records were uploaded to the Smokeless Powders Database during this project.

New Reference Materials

One hundred new smokeless powders were purchased and analyzed following the SOP in Appendix B. The data and product information for these samples were uploaded to the database. These smokeless powders were used to create a reference collection which was distributed to ASCLD/LAB accredited laboratories performing explosives analysis.

Evidentiary Value of Smokeless Powder Comparisons

Discrete chemical composition

Searches of the Smokeless Powders Database requires the user to select discrete values of the various properties that are indexed in the database. The search returns all of the records that match the specified criteria and have additional properties not specified in the search. For example, a search for records having the same shape, color, perforation, measurement, and containing nitroglycerin and diphenylamine, will return all records matching these criteria, as well as, those that match the criteria but may also contain 2,4-

dinitrotoluene, which was not specified in the search. The research conducted under this grant considered only exact matches to the search criteria. This approach corresponds to specifying the entire subset of the 13 organic compounds that are present in a sample. Returned records that contain non-matching subsets of organic compounds are therefore designated as a non-match. When the possible 263,175 pairwise combinations were restricted to comparisons between records having the same physical properties (s, c, p and m), a total of 4,367 comparisons were identified. The resulting confusion matrix of "matching" and "non-matching" comparisons are given in Table 3.

Table 3. Shows comparisons of "matching" (identical subset of organic compounds) and "non-matching" (subset of organic compounds did not match) records for same-product and different-product pairwise combinations.

		Comparison of Organics		
		"match"	"non-match"	Total
Ground-Truth	Same Product	52	116	168
	Different Product	462	3737	4199
	Total	514	3853	4367

Following the formulations given in an earlier section for the calculation of true positive rate, TPR, and false positive rate, FPR, a TPR of 0.31 and an FPR of 0.11 can be calculated. The TPR is a measure of the sensitivity of the database search to return records with characteristics matching the search criteria which correspond to the same product. The

FPR is equal to the quantity of one minus the specificity (1-specificity) of the search. These two criteria (TPR and FPR) define the receiver operating characteristics (ROC) space and the discrete results from Table 3 represent a single point in ROC space [30]. The slope of a line passing from the intercept through the single point in ROC space is equivalent to the likelihood ratio of 2.81 [31, 32]. The likelihood ratio from this single point represents the probability of observing a "match" in composition, given that the search sample and database record are the same product (with the same s, c, p, and m properties), divided by the probability of observing a "match" in composition given that the search sample and database record correspond to different products (also having the same s, c, p and m). The likelihood ratio offers only limited support for the same product proposition as opposed to the different product proposition support, when using Evett's verbal scale [28]. The accuracy of the method is 0.87 (87%) [30].

The database search evaluation presented here does not take into account the presence of compounds other than the 13 organic analytes considered. The likelihood ratio would change if the cut-off criteria for detection of the organic analytes were changed. In order for the likelihood ratio provided here to be useful, the presence or absence of the 13 organics in the sample being searched against the database would have to be determined with the same criteria used to establish the database. Positive identification of an analyte for samples in the database that were analyzed by NCFS was determined by a 3:1 signal-to-noise ratio in the total ion chromatograph, the presence of at least three major ions (for all analytes except nitroglycerin which only has two major ions in the EI-MS) , and a retention time within ± 0.1 min of the corresponding standard.

Continuous chemical composition

The score, *z*, was calculated for all pairwise comparisons of same product and different product pairs that share the same s, c, p and m. The kernel distributions were calculated for the *z* scores from same product and different product distributions and examined graphically. The different sample distribution was bimodal and the same sample distribution exhibited negative kurtosis. The two distributions were significantly overlapped. The area under the ROC curve (AUC) was 0.814 and represents the accuracy of the method. The accuracy of the continuous score method was slightly less than the for the discrete comparison method (0.87), as described above. The likelihood ratios calculated from Equation 1, have values of 0.10, 0.29, 1.32, 2.20 and 10.61 at *z* values of 0.3, 1, 2, 3, 4 and 5, respectively. These values provide only limited support for either H_D or H_P using Evett's verbal scale [63], with the exception of the likelihood ratio calculated at $z = 5$, which provides moderate support for H_P and corresponds to a correlation coefficient of r = 0.99991 between TIS. A large correlation between TIS will only result from virtually identical total ion chromatograms. Cross validation with a 10% hold-out repeated 20 times was used to further test the method. The discriminating power of the method was estimated from detection error tradeoff (DET) plots to be approximately 30% [64.]. Tippett and empirical cross entropy (ECE) plots revealed approximately 30% misleading evidence; however, the method appeared to be well calibrated. Tippett, DET and ECE plots will be discussed in greater detail in the text that follows. The following observations were made regarding the kernel density estimate overlap at low *z* values:

> "The contributions to the same product distribution at low z values result from comparisons between products that have undergone changes in formulation. For example, comparison of two samples of Hodgdon TRAP 100 (SRN 400 and 670), yields a z of 0.618 (r = 0.549). Both samples contain nitroglycerin, 2,4-dinitrotoluene, diphenylamine, ethyl centralite, dibutyl phthalate and 2-nitrodiphenylamine; however the SRN 670 sample also contains 4-

> nitrodiphenylamine as a major constituent. A similar example is a comparison of two samples of Hercules 2400 (SRN 219 and 680), which have a z score of 1.39 (r = 0.883). Both samples contain nitroglycerin, diphenylamine and 2-nitrodiphenylamine in similar ratios; however, SRN 219 also contains a significant amount of ethyl centralite." [26]

An overlap in kernel density estimates for same product and different product scores was also observed at higher score values (i.e., at higher similarities). The overlap at higher scores reflects the fact that some powders representing different products have highly similar physical and chemical characteristics. Some examples of these were provided in our publication:

> "An example is the comparison of DuPont 700x and Alliant Bullseye (SRN 405 and 800). This comparison has a z of 4.72 (r = 0.9998), and both samples contain nitroglycerin as well as ethyl centralite as major components. The total ion chromatograms of the two samples are highly similar. Another example is the comparison of Accurate 3100 and Accurate Magnum Rifle (SRN 609 and 618). This comparison has a z of 4.84 (r = 0.9999), both samples contain 2,4-dinitrotoluene, diphenylamine, 2-nitrodiphenylamine and dioctyl phthalate as major components, and their total ion chromatograms are highly similar. A third example is the comparison of Hodgdon 870 and Accurate MR-8700 (SRN 472 and 668). In this case z for the comparison is 3.58 (r = 0.998), and both samples contain nitroglycerin, diphenylamine, dibutyl phthalate and 2-nitrodiphenylamine; however, the Hodgdon product also contains a small amount of 2,4-dinitrotoluene." [26]

Implementing a method with a continuous similarity score, while conditioning comparisons on matching s, c, p and m discrete values, did not have a superior performance to the discrete comparisons of matching subsets of the 13 specified organic analytes. In an attempt to improve on the performance of the continuous model, classification of smokeless powders beyond single and double base was addressed.

Agglomerative hierarchical cluster (AHC) analysis

The results of AHC analysis of the smokeless powders is shown in Figure 3. The optimally ordered dendrogram is shown along the top of Figure 3, with boxes drawn to demonstrate the six clusters of samples. The intensities of the ions in the TIS are shown below the dendrogram. The ions are ordered by increasing *m/z*, starting from the bottom of the graph. From the graph, it can be seen that each of the six groups has a corresponding most intense ion that defines the group. This observation allows the analyst to classify smokeless powders into one of the six clusters based on observation of the most intense ion in the TIS. The chemical explanation of cluster membership is summarized in Table 4. While the grouping is not perfect, as observed in Figure 3, 98% of all smokeless powders were grouped by AHC into clusters corresponding to a common most intense ion in their TIS.

Table 4. Shows the six clusters observed from AHC analysis along with the dominant ions that define the clusters, and their associated analytes.

Cluster	Most intense ion (*m/z*)	Associated Analyte
C1	46	nitroglycerin
C2	149	phthalates
C3	120	ethyl centralite
C4	165	dinitrotoluenes
C5	169	diphenylamines
C6	134	methyl centralite

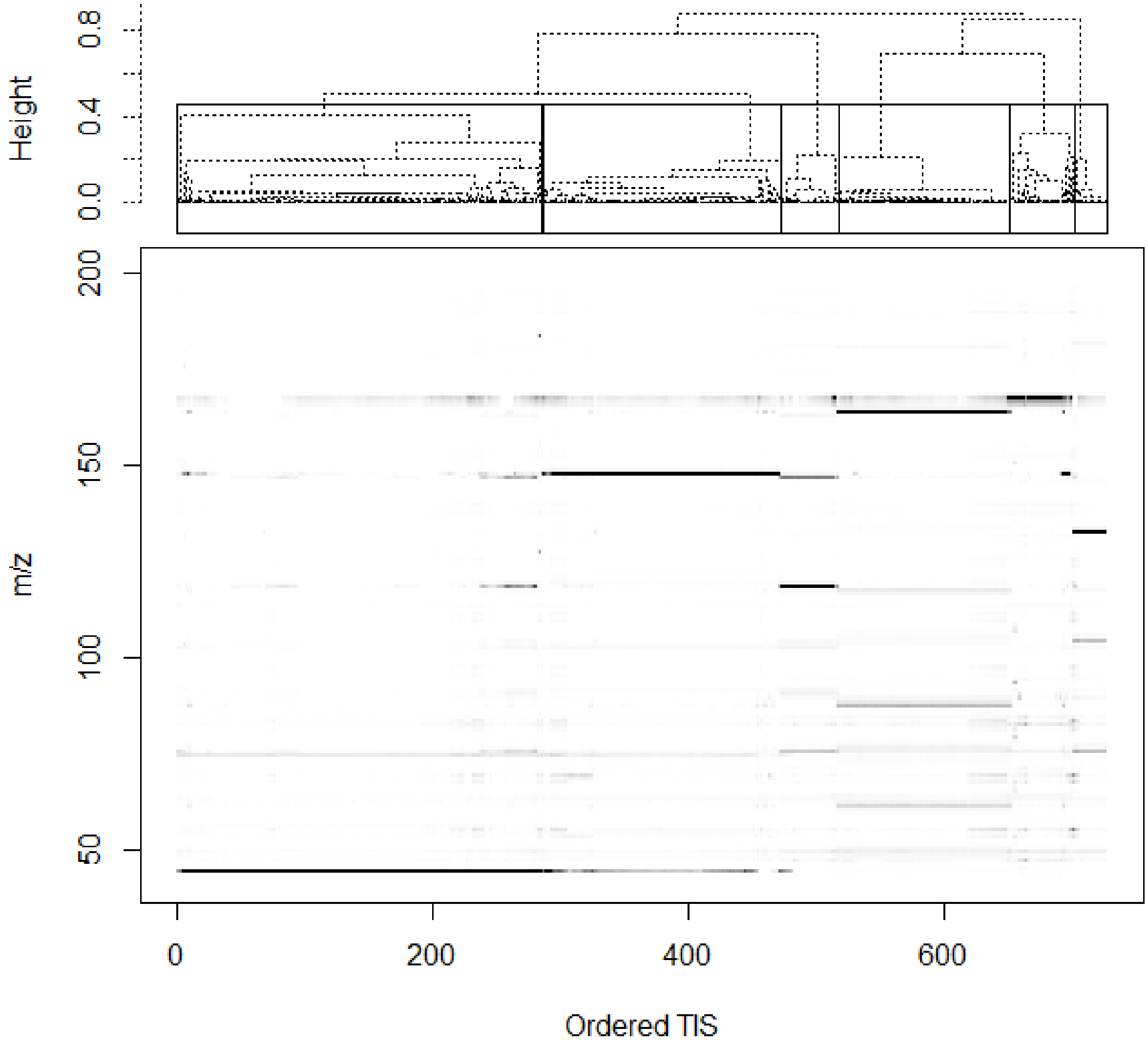

Fig. 3. Shows the optimally ordered dendrogram and heat map obtained from AHC analysis of 726 Smokeless Powders Database samples.

Score-based likelihood ratios (same color, shape, cluster)

The score-based likelihood ratio discrimination method discussed above was extended by limiting smokeless powder comparisons to those powders that had same discrete values for shape (s), color (c), perforation (p), measure (m) and cluster (cl). The likelihood ratio equation then becomes the following:

$$LR = \frac{f(z|P,s,c,p,m,cl)}{f(z|\bar{P},s,c,p,m,cl)} \tag{7}$$

Limiting the comparisons by including cluster membership reduced the number of comparisons to 2,154, which included 157 same product and 1,996 different product comparisons. The histograms of the distributions of the score, z, for same product (dark shading) and different product (light shading) groups are shown in Figure 4a. These distributions are closer to normal and are modeled by normal distributions, N (mean, standard deviation). The solid curve, $N(3.17, 0.93)$, shows the same product model and the dashed curve, $N(2.52, 0.95)$, shows the different product distribution. The graph on the right, Figure 4b, shows the ROC curve for the data and exhibits an AUC of 0.6832. The square symbol in Figure 4b represents the single point in ROC space for the discrete comparisons based on subsets of the thirteen organic components. The likelihood ratios corresponding to Equation 7, have values of 0.25, 0.54, 1.13, 2.33 and 4.65 at z values of 1, 2, 3, 4 and 5, respectively. These values provide only limited support for either H_D or H_P using Evett's verbal scale [63]. The range of values once again indicates that the evidentiary and investigative value of a database search match, while helpful, is nowhere near a unique powder identification. The importance of these results is that the significance attached to any individual item returned from a database search is not high; however, the investigative value of a list of powders matching the physical and chemical characteristics of a questioned item may prove useful.

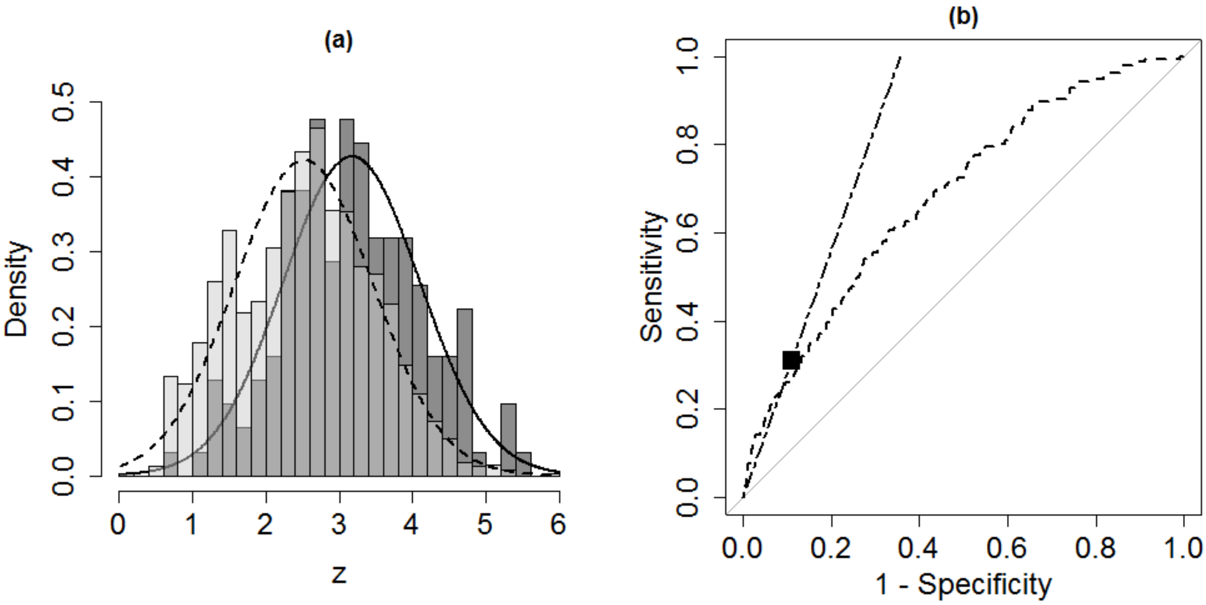

Fig. 4. (a) Shows the distribution for the Fisher transform of Pearson correlation coefficients for same product (dark bars) and different product (light bars) comparisons after imposing the same AHC cluster membership constraint. The solid line and dashed line show the normal density distributions for the same product and different product comparisons, respectively. (b) Shows the ROC curve (dashed) derived from the two normal distributions shown in (a). The square symbol represents the single point in ROC space for the discrete comparisons based on subsets of the thirteen organic components

Probability of a Smokeless Powder Manufacturer for Investigative Purposes

Conditional Independence Testing of the Smokeless Powder Random Variables

The smokeless powder physical features which were tested for conditional independence include kernel shape; color; luster; average diameter; average length; the presence/absence of a perforation; and the presence/absence of a bias cut. Additional physical characteristics which may be observed in smokeless powders, such as marker color; and distinguishing features such as teardrops, oblongs, etc., were not included in this study since they were either observed infrequently, or their identification was subjective. The smokeless powder chemical characteristics which were tested for conditional independence include cluster, as previously determined from AHC analysis, and the cluster associated dominant ions – m/z 46, 120, 134, 149, 165, and 169. The causal relationships between the manufacturer and the smokeless powder features described above were used to propose the structure of the Bayesian network, and to determine the combination of features to be tested for conditional independence. Consequently, the physical features and cluster were tested against each other given the manufacturer; while the AHC dominant ions were tested against each other given the cluster. Of those characteristics tested, perforation was found to be conditionally dependent on other smokeless powder physical features given the manufacturer; therefore, it was removed from the network structure.

Predicting Smokeless Powder Manufacturers using High Posterior Probabilities

The smokeless powder Bayesian network structure, comprising 14 nodes, which satisfied the causal relationships and conditional independence requirements is shown in Figure 5. Two parent nodes, **Man** and **Clstr**, representing smokeless powder manufacturers and AHC cluster identifiers, respectively, are present in the network's structure. The **Man** node is the only root node within the network structure; that is, this node is not a child of any other node. Conversely, the **Clstr** node is both a child of the **Man** node, and the parent of the dominant ion nodes: ***m/z* 46**, **120**, **134**, **149**, **165**, and **169**. Additional smokeless powder features represented in the network as children of the **Man** node are: average diameter, **Dia**; average length, **Len**; color, **Clr**; kernel shape, **Shpe**; luster, **Lstr**; and the absence/presence of a bias cut, **Bias**.

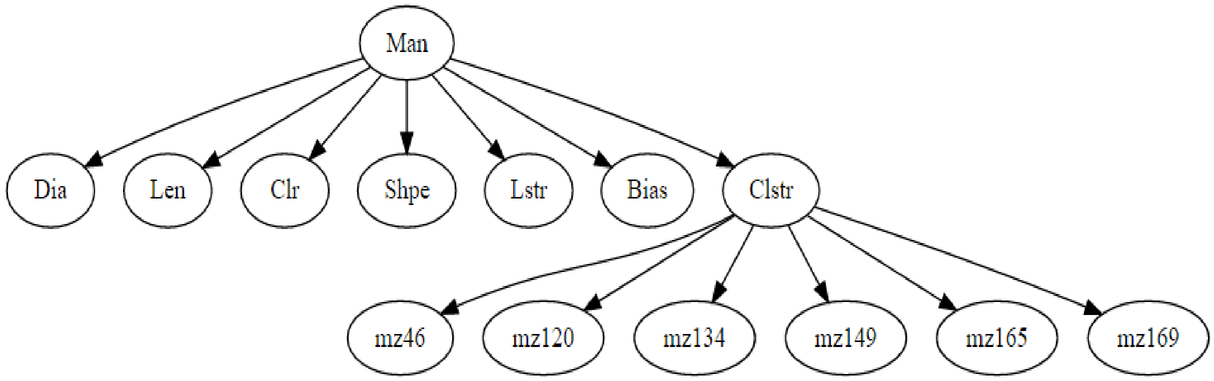

Fig. 5. Shows the Smokeless Powder Bayesian Network which satisfied the causal relationships and conditional independence requirements. The Man and Clstr nodes are parent nodes; the Clstr node is also a child of the Man node.

Within the network, probability tables associated with root parent and child nodes encode unconditional and conditional probabilities, respectively. For example, the table associated with the **Man** node encodes unconditional probabilities for six states representing the smokeless powder manufacturers: Alliant Powder Company, *APC*; Hodgdon Powder Company, *HPC*; Norma Precision AB, *NPAB*; Nobel Sport, *NS*; Vihta Vuori, *ViVu*; and Western Powders, *WP*; whereas, the table associated with the **Clstr** node encodes conditional probabilities for multiple states describing the unique AHC cluster identifiers, *C1* to *C6*, each of which are contingent on each smokeless powder manufacturer within the **Man** node. Similarly, each node corresponding to an AHC dominant ion, such as ***m/z* 46**, contains multiple states describing the strength of the ion as previously described, conditioned on each state in the **Clstr** node. The unconditional probabilities in the **Man** node represent the prior probabilities for each smokeless powder manufacturer; whereas, conditional probabilities in the child nodes represent the likelihoods of observing the states within those nodes, conditioned on the states within their parent nodes.

The probabilities ascribed to the states in each node were calculated based on frequencies in the model data, comprising 90% of the smokeless powder dataset. Smokeless powders manufactured by HPC comprise approximately 50% of the model data; while powders manufactured NPAB and ViVu occur with the lowest prior probabilities, 5% and 8%, respectively. Each smokeless powder sample in the test set, comprising 10% of the smokeless powders data, was used to instantiate the network. The network was instantiated by setting one state in each child node to certainty; that is, the probability was set equal to 1 when the state was observed. For example, using SRN 121, the probability of the state "C1" was set equal to 1 in the **Clstr** node; the probability of state "strg" was set equal to 1 in the ***m/z* 46** node; the probability of state "wk" was set equal to 1 in the ***m/z* 120** node; the probability of state "wk" was set equal to 1 in the ***m/z* 134** node, etc. These probabilities were propagated in the network, and the prior probabilities in the **Man** node were updated to posterior probabilities. Within the code, the smokeless powder manufacturer for each test sample was predicted based on the highest posterior probability. Correct predictions were determined when the ground truth and predicted manufacturer were the same. Incorrect predictions resulted when ground truth and predicted manufacturers were not the same. As an example, APC was the known manufacturer of SRN 121 and the network calculated a posterior probability of 0.99 for the same manufacturer; this resulted in a correct prediction. Conversely, the known manufacturer of SRN 151 was HPC; however, the highest posterior probability (0.97) was observed for WP which resulted in an incorrect prediction.

Percent correct rates were calculated for each smokeless powder manufacturer, and for the overall test set. The confusion matrix shown in Table 5 summarizes the percent correct rates for predicting manufacturers based on the physical and chemical characteristics of the smokeless powders. Smokeless powders manufactured by HPC and ViVu exhibited the highest percent correct rates at 82% for each manufacturer; low percent correct rates of 39% and 65% were observed for powders manufactured by Nobel Sport (NS) and Norma Precision AB (NPAB). The low correct rate for NS and NPAB powders is attributed to the high similarity between their physical and chemical profiles and those of Hodgdon, Alliant Powder Company and Vihta Vuori powders. Table 5 also summarizes incorrect manufacturer predictions by the network (off-diagonal values). Smokeless powders manufactured by APC which were incorrectly predicted to have been manufactured by HPC, NPAB, and WP. Similarly, smokeless powders manufactured by HPC which were incorrectly predicted were assigned as APC and WP products. NPAB powders which were incorrectly predicted were assigned as APC and HPC products, and incorrect predictions of NS powders were assigned as HPC and ViVu products. Incorrect predictions of smokeless powders manufactured by ViVu and WP were assigned exclusively to NS and HPC, respectively. An explanation toward incorrect predictions by the network was provided by comparing the physical and chemical profiles of smokeless powder samples in the test set, to profiles of samples in the model data. Specifically, the profiles of the test samples were compared to profiles of samples in the model data which corresponded to the manufacturer predicted by the network. A number of samples in the test data which were incorrectly predicted exhibited similar or identical profiles to samples in the model data. Table 6 lists some examples of test samples where the network predicted different manufacturers than the ground truth along with corresponding samples from the predicted manufacturers, which have identical physical and chemical profiles. For example the ground truth manufacturer for SRN 119 was APC; however, the network predicted HPC as the manufacturer of that sample. SRN 368, manufactured by HPC, exhibited an identical physical and chemical profile as SRN 119. Other examples are given in Table 6. Though there were a number of incorrect predictions for smokeless powder manufacturers, the overall percent correct rate was 76%, which is significantly higher than the approximate 17% random chance of correct assignment when considering six manufacturers.

Table 5. Confusion Matrix summarizing predictions of Smokeless Powder Manufacturers based on high posterior probabilities using the Bayesian network. The highest percent correct rates were obtained for smokeless powders manufactured by HPC and ViVu, and the lowest percent correct rate was obtained for smokeless powders manufactured by NS. The overall percent correct rate for predictions by the network was 76%.

Ground Truth	**Predictions based on the Highest Posterior Probability**							**Percent Correct**
	APC	**HPC**	**NPAB**	**NS**	**ViVu**	**WP**	**SUM**	
APC	173	36	9	0	0	12	230	75
HPC	66	665	0	0	0	76	807	82
NPAB	20	5	46	0	0	0	71	65
NS	0	32	0	50	47	0	129	39
ViVu	0	0	0	20	93	0	113	82
WP	0	53	0	0	0	163	216	75
SUM	259	791	55	70	140	251	1566	76

Table 6. Discrete profiles of some smokeless powder test samples incorrectly predicted by the network (first line) and model data having similar or identical physical and chemical profiles (second line).

SRN	**Ground Truth**	**Prediction**	**Shpe**	**Bias**	**Clr**	**Lstr**	**Dia**	**Len**	**Clstr**	***m/z* 46**	***m/z* 120**	***m/z* 134**	***m/z* 149**	***m/z* 165**	***m/z* 169**
119	**APC**	**HPC**	Disk	No	Gray	Dull	d5	l1	C1	strg	wk	wk	wk	wk	wk
368	**HPC**		Disk	No	Gray	Dull	d5	l1	C1	strg	wk	wk	wk	wk	wk
151	**HPC**	**WP**	Ball	No	Black	Dull	d2	l1	C2	strg	wk	wk	strg	wk	med
186	**WP**		Ball	No	Black	Dull	d2	l1	C2	strg	wk	wk	strg	wk	med
377	**ViVu**	**NS**	Cylinder	No	Brown	Dull	d2	l2	C5	wk	wk	wk	wk	wk	strg
395	**NS**		Cylinder	No	Brown	Dull	d2	l2	C5	wk	wk	wk	wk	wk	strg
179	**WP**	**HPC**	Flattened Ball	No	Gray	Dull	d2	l1	C2	strg	wk	wk	strg	wk	med
148	**HPC**		Flattened Ball	No	Gray	Dull	d2	l1	C2	strg	wk	wk	strg	wk	med

IV. Conclusions

Discussion of Findings

Database Expansion and Reference Collection

The database expansion included the addition of 538 historical records spanning 30 - 40 years and the addition of 100 new smokeless powders currently on the reloading market. It is important to continue populating the database with powders that are being sold on the market. For example, the French manufacturer Nobel Sport began exporting their Vectan smokeless powders to the United States market in 2015. Some of these powders have shapes and colors not seen elsewhere in the database. The database provides a means to search though large numbers of smokeless powders by their physical characteristics and chemical components.

The 100 new smokeless powders were used to create smokeless powder reference collections. These were distributed to ASCLD/Lab accredited laboratories to incorporate into their in-house collections. Smokeless powders are classified as flammable solids, which have regulatory limitations for the quantity stored in a facility. The smallest containers sold on the market weigh 12 – 16 ounces. This means a reference collection would be limited in the number of powders it contained, or the excess powder would need to be disposed in order to expand the collection. Disposal of excess amounts of smokeless powder is difficult since it is a hazardous material. It was advantageous for NCFS as well as the accredited laboratories to share the smokeless powder reference materials. The reference collections are also valuable for comparing analytical results between laboratories and the database.

Evidentiary/Investigative Value of Database Searches

The evidentiary and investigative value of a same product assertion based on a Smokeless Powders Database search has been examined. The probability of observing the evidence given the same product proposition, as opposed to the different product proposition will generally be less than 10. The probabilities that constitute the likelihood ratio are conditioned on the smokeless powders having the same kernel shape, color, presence/absence perforation and measurement. The evidence corresponds to an exact match in the subset of organic analytes present in the sample (discrete data) or the Fisher transformation of the Pearson correlation between the TIS of two samples. When the pairwise comparisons are further conditioned by restricting comparisons to samples belonging to the same AHC cluster, the range of evidentiary and investigative value is slightly smaller. The results demonstrate the limits on the strength of probabilistic assertions that can be made regarding smokeless powder comparisons using the properties examined in this work.

Probabilistic Assessment of Manufacturer

A Bayesian network for predicting smokeless powder manufacturers based on high posterior probabilities has been examined. The physical and chemical characteristics which are typically utilized in a database search were used to instantiate the network and to calculate posterior probabilities for each manufacturer using these smokeless powder features. Higher percent correct rates were observed when predicting six smokeless powder manufacturers than the approximate 17% rate which would be observed for random choice. The results demonstrate the utility of a Smokeless Powders Bayesian network as an investigative tool for making probabilistic predictions of the manufacturer based on the physical and chemical properties of a smokeless powder sample.

Implications for Policy and Practice

The expansion of the Smokeless Powders Database has provided additional data on smokeless powders that assists the analyst in assessing casework on a daily basis. The reference collections provide a physical link to the database. These advances in data accessibility eliminate the need for individual databases in each laboratory and provides ASCLD/LAB accredited laboratories with smokeless powder references that they would not otherwise be able to attain. The results from probabilistic assessment of the evidentiary value of matching physical characteristics and chemical composition informs the community of the value of such comparisons for the purpose of investigative and probative applications. The generally low evidentiary value of these "matches" should influence practice and future policy regarding the introduction of smokeless powder evidence. The use of Bayesian networks to predict the manufacturer of a smokeless powder sample based on chemical and physical characteristics was shown to provide a nearly 80% correct prediction rate, which exceeds the 17% random probability based on a flat prior, or 52% probability of correctly assessing the manufacturer if the analyst always guesses Hodgdon Powder Company. These results provide an incentive for using these statistical methods in the investigative process.

Implications for Further Research

The Smokeless Powders Database has been expanded, as detailed in the proposed work, and reference collections have been prepared and disseminated to forensic laboratories. The composition of the marketplace (i.e., manufacturers, types of powders, powder composition, etc.) continues to change and the database is only a snapshot in time of the marketplace composition. There is a need to continually update and expand the database, especially given the widespread availability of smokeless powders in the U.S. and their frequent use in bombings.

While it can be argued that individualization of smokeless powders is not possible, the need still exists for improved evidentiary and investigative value from comparisons of the physical characteristics and chemical composition of smokeless powders. The path forward may require the application of additional analytical methods to the analysis of smokeless powders, or improved data analysis. Implementation of probabilistic methods is heavily reliant on defining a relevant population. Defining a relevant population is aided by a survey of casework samples and the further expansion of the smokeless powders database to allow construction of a relevant population data source on which to base likelihood ratio methods.

Improved characterization of smokeless powders, especially the inorganic components, may offer better characterization of the manufacturer and improve the evidentiary and investigative value of smokeless powders.

V. References

[1] Council NR. Black and Smokeless Powders: Technologies for Finding Bombs and the Bomb Makers. Washington, DC: The National Academies Press; 1998.
[2] Hopper KG, McCord BR. A comparison of smokeless powders and mixtures by capillary zone electrophoresis. J Forensic Sci. 2005;50:307-15.
[3] Hopper KG, Leclair H, McCord BR. A novel method for analysis of explosives residue by simultaneous detection of anions and cations via capillary zone electrophoresis. Talanta. 2005;67:304-12.
[4] Cascio O, Trettene M, Bortolotti F, Milana G, Tagliaro F. Analysis of organic components of smokeless gunpowders: high-performance liquid chromatogaphy vs. micellar electrokinetic capillary chromatography. Electrophoresis. 2004;25:1543-7.
[5] Reardon MR, MacCrehan WA. Developing a quantitative extraction technique for determining the organic additives in smokeless handgun powder. J Forensic Sci. 2001;46:802-7.
[6] Scherperel G, Reid GE, Smith RW. Characterization of smokeless powders using nanoelectrospray ionization mass spectrometry (nESI-MS). Analytical & Bioanalytical Chemistry. 2009;394:2019-28.
[7] Harvey SD, Peters TJ, Wright BW. Safety considerations for sample analysis using a near-infrared (785 nm) Raman laser source. Appl Spectrosc. 2003;57:580-7.
[8] Guerra-Diaz P, Gura S, Almirall JR. Dynamic planar solid phase microextraction-ion mobility spectrometry for rapid field air sampling and analysis of illicit drugs and explosives. Anal Chem. 2010;82:2826-35.
[9] Joshi M, Delgado Y, Guerra P, Lai H, Almirall JR. Detection of odor signatures of smokeless powders using solid phase microextraction coupled to an ion mobility spectrometer. Forensic Sci Int. 2009;188:112-8.
[10] MacCrehan WA, Bedner M. Development of a smokeless powder reference material for propellant and explosives analysis. Forensic Science International.163:119-24.
[11] MacCrehan WA, Reardon MR, Duewer DL. A quantitative comparison of smokeless measurements. J Forensic Sci. 2002;47:1283-7.
[12] MacCrehan WA, Reardon MR, Duewer DL. Associating gunpowder and residues from commercial ammunition using compositional analysis. J Forensic Sci. 2002;47:260-6.
[13] de Perre C, Corbin I, Blas M, McCord BR. Separation and identification of smokeless gunpowder additives by capillary electrochromatography. Journal of Chromatography A. 2012;1267:259-65.
[14] Heramb RM, McCord BR. The Manufacture of Smokeless Powders and their Forensic Analysis: A Brief Review. Forensic Science Communications. 2002;4.
[15] Bender EC. Analysis of Low Explosives. In: Beveridge A, editor. Forensic Investigation of Explosions. Bristol, PA: Taylor & Francis Ltd.; 1998. p. 343-88.
[16] R Core Team. R: A Language and Environment for Statistical Computing. Version 3.0.3 ed. Vienna, Austria: R Foundation for Statistical Computing; 2011.
[17] Moorehead W. Characterization of Smokeless Powders. In: Blackledge RD, editor. Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis: John Wiley & Sons, Inc.; 2007. p. 241-68.

[18] López-López M, de la Ossa MÁF, Galindo JS, Ferrando JL, Vega A, Torre M, et al. New Protocol for the Isolation of Nitrocellulose from Gunpowders: Utility in their identification. Talanta. 2010;81:1742-9.
[19] Lopez-Lopez M, Ferrando JL, Garcia-Ruiz C. Comparative analysis of smokeless gunpowders by Fourier transform infrared and Raman spectroscopy. Analytica Chimica Acta. 2012:92-9.
[20] Joshi M, Rigsby K, Almirall JR. Analysis of the Headspace Composition of Smokeless Powders using GC–MS, GC-μECD and Ion Mobility Spectrometry. Forensic Science International. 2011;208:29-36.
[21] Wallace CMC. Smokeless Powder Characterization: An Investigative Tool in Pipe Bombings. In: Yinon J, editor. Advances in Analysis and Detection of Explosives Proceedings of the 4th International Symposium on Analysis and Detection of Explosives, September 7-10, 1992, Jerusalem, Israel. Dordrecht, Netherlands: Kluwer Academic Publishers; 1992. p. 29-39.
[22] National Center for Forensic Science (NCFS). The NCFS Smokeless Powders Database can be found at the Universal Resource Locator http://www.ilrc.ucf.edu/powders/. Last accessed May 31, 2016.
[23] Keto RO. Comparison of Smokeless Powders by Pyrolysis Capillary Gas Chromatography and Pattern Recognition. Journal of Forensic Sciences. 1989;34:74 - 82.
[24] Andrasko J. Characterization of Smokeless Powder Flakes from Fired Cartridge Cases and form Discharg Patterns on Clothing. Journal of Forensic Sciences. 1992;34:1030 - 47.
[25] Perez JJ, Flanigan PM, Brady JJ, Levis RJ. Classification of Smokeless Powders Using Laser Electrospray Mass Spectrometry and Offline Multivariate Statistical Analysis. Analytical Chemistry. 2013;85:296-302.
[26] Dennis DMK, Williams MR, Sigman ME. Assessing the evidentiary value of smokeless powder comparisons. Forensic Science International. 2016;259:179-87.
[27] Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA. A hierarchy of propositions: deciding which level to address in casework. Science & Justice. 1998;38:231-9.
[28] Evett Iw, Jackson G, Lambert JA, McCrossan S. The impact of the principles of evidence interpretation on the structure and content of statements. Science & Justice. 2000;40:233-9.
[29] R Development Core Team. R: A Language and Environment for Statistical Computing. Version 3.2.3 ed. Vienna, Austria: R Foundation for Statistical Computing; 2015.
[30] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27:861-74.
[31] Choi BCK. Slopes of a Receiver Operating Characteristic Curve and Likelihood Ratios for a Diagnostic Test. American Journal of Epidemiology. 1998;148:1127-32.
[32] Johnson NP. Advantages to transforming the receiver operating characteristic (ROC) curve into likelihood ratio co-ordinates. Statistics in Medicine. 2004;23:2257-66.
[33] Sigman ME, Williams MR, Castelbuono JA, Colca JG, Clark CD. Ignitable Liquid Classification and Identification using the Summed-Ion Mass Spectrum. Instrumentation Science & Technology. 2008;36:375-93.

[34] Williams MR, Sigman ME, Lewis J, Pitan KM. Combined target factor analysis and Bayesian soft-classification of interference-contaminated samples: Forensic Fire Debris Analysis. Forensic Science International. 2012;222:373-86.
[35] Waddell EE, Song ET, Rinke CN, Williams MR, Sigman ME. Progress Toward the Determination of Correct Classification Rates in Fire Debris Analysis. Journal of Forensic Sciences. 2013;58:887-96.
[36] Waddell EE, Frisch-Daiello JL, Williams MR, Sigman ME. Hierarchical Cluster Analysis of Ignitable Liquids Based on the Total Ion Spectrum. Journal of Forensic Sciences. 2014;59:1198-204.
[37] Waddell EE, Williams MR, Sigman ME. Progress Toward the Determination of Correct Classification Rates in Fire Debris Analysis II: Utilizing Soft Independent Modeling of Class Analogy (SIMCA). Journal of Forensic Sciences. 2014;59:927-35.
[38] Lopatka M, Sigman ME, Sjerps MJ, Williams MR, Vivo-Truyols G. Class-conditional feature modeling for ignitable liquid classification with substantial substrate contribution in fire debris analysis. Forensic Science International. 2015;252:177-86.
[39] Mat Desa WNS, Nic Daéid N, Ismail D, Savage K. Application of unsupervised chemometric analysis and self-organizing feature (SOFM) for the classification of lighter fuels. Analytical Chemistry. 2010;82:6395-400.
[40] Mat Desa WNS, Ismail D, Nic Daéid N. Classification and source determination of medium petroleum distillates by chemometric and artificial neural networks: a self organizing feature approach. Analytical Chemistry. 2011;83:7745-54.
[41] Lucas A. amap: Another Multidimensional Analysis Package. R package version 0.8-7 ed; 2011.
[42] Buchta C, Hahsler M. cba: Clustering for Business Analytics. R package version 0.2-9 ed; 2012.
[43] Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast Optimal Leaf Ordering for Hierarchical Clustering. Bioinformatics. 2001;17:S22 - S9.
[44] Wilkinson L, Friendly M. The History of the Cluster Heat Map. The American Statistician. 2009;63:179-84.
[45] Bak M, Silahtaroglu A, Moller M, Christensen M, Rath MF, Skryabin B, et al. MicroRNA expression in the adult mouse central nervous system. RNA. 2008;14:432-44.
[46] Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, et al. Modelling schizophrenia using human induced pluripotent stem cells. Nature. 2011;473:221-5.
[47] Duan Z, Andronescu M, Schutz K, Mcllwain S, Kim YJ, Lee C, et al. A Three-Dimensional Model of the Yeast Genome. Nature. 2010;465:363-7.
[48] Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nature Biotechnology. 2007;25:197-206.
[49] Key M. A Tutorial in Displaying Mass Spectrometry-Based Proteomic Data using Heat Maps. BMC Bioinformatics. 2012;13.
[50] Maze I, Covington III HE, Dietz DM, LaPlant Q, Renthal W, Russo SJ, et al. Essential Role of the Histone Methyltransferase G9a in Cocaine-Induced Plasticity. Science. 2010;327:213-6.

[51] McGregor LA, Gauchotte-Lindsay C, Nic Daéid N, Thomas R, Kalin RM. Multivariate statiscal methods for the environmental forensic classification of coal tars from former manufactured gas plants. Environmental Science & Technology. 2012;46:3744-52.
[52] Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene Expression of Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis. Journal of the National Cancer Institute. 2006;98:262-72.
[53] Taroni F, Aitken C, Garbolino P, Biedermann A. Bayesian Networks and Probabilistic Inference in Forensic Science. West Sussex, England: John Wiley & Sons, Ltd.; 2006.
[54] Gittelson SN. Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence. Lausanne, Switzerland: University of Lausanne; 2013.
[55] Puga JL, Krzywinski M, Altman N. Bayesian Networks: For making probabilistic inferences, a graph is worth a thousand words. Nature Methods. 2015;12:799-800.
[56] Barbas JT, Dabestani R, Sigman ME. A MECHANISTIC STUDY OF PHOTODECOMPOSITION OF ACENAPHTHYLENE ON A DRY SILICA SURFACE. J Photochem Photobiol A-Chem. 1994;80:103-11.
[57] Lucy D. Introduction to Statistics for Forensic Scientists. West Sussex, England: John Wiley & Sons, Ltd.; 2005.
[58] Stone JV. Bayes' Rule: A Tutorial Introduction to Bayesian Analysis. First ed. Sheffield, England: Sebtel Press; 2013.
[59] Scutari M. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software. 2010;35:1-22.
[60] Højsgaard S, Dethlefsen C. A Common Platform for Graphical Models in R: The gRbase Package. Journal of Statistical Software. 2005;14:1-12.
[61] Højsgaard S. Graphical Independence Networks with the gRain Package for R. Journal of Statistical Software. 2015;46:1-26.
[62] Højsgaard S, Edwards D, Lauritzen S. Graphical Models with R. gRim Package; 2012.
[63] Evett IW, Jackson G, Lambert JA. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. Science & Justice. 2000;40:3-10.
[64] A. Martin GD TK, M. Ordowski, M. Przybocki. The DET curve in assessment of decision task performance. Proceedings of Eurospeech. 1997;1:1895 - 8.

VI. Dissemination of Research Findings

Publications and Thesis

Dana-Marie K. Dennis, Mary R. Williams, Michael E. Sigman "Assessing the evidentiary value of smokeless powder comparisons", Forensic Sci. International (**2016**) Vol. 259, p179–187.

DANA-MARIE KARINE DENNIS "CHEMICAL ANALYSIS, DATABASING, AND STATISTICAL ANALYSIS OF SMOKELESS POWDERS FOR FORENSIC APPLICATION", A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the College of Sciences at the University of Central Florida, Orlando, Florida, Summer Term, 2015.

Dana-Marie K. Dennis; Mary R. Williams; Michael E. Sigman, “Investigative Probabilistic Predictions of Smokeless Powder Manufacturers utilizing a Bayesian Network”, Forensic Chemistry, **manuscript submitted, 2016**.

Databases

Smokeless Powders Database: http://www.ilrc.ucf.edu/powders/

Patents

None

Presentations

Preliminary Presentations

Dana-Marie K. Dennis, Michael E. Sigman, Mary R. Williams “Classification of Smokeless Powders by Cluster Analysis,” American Academy of Forensic Sciences 65th Anniversary Meeting, Washington, DC, February 18 – 23, 2013.

Dana-Marie K. Dennis, Erin E. Waddell, Mary R. Williams, Michael E. Sigman, “Classification of Smokeless Powders by Cluster Analysis”, 89th Annual Florida American Chemical Society Meeting and Exposition (FAME), May 9 – 11, 2013, Palm Harbor, FL.

Grant-funded Presentations

Dana-Marie Dennis, Michael E. Sigman, Mary R. Williams “Cluster Analysis of Smokeless Powders and Classification by Discriminant Analysis”, 90th Annual Florida American Chemical Society Meeting and Exposition (FAME), May 8 – 10, 2014, Palm Harbor, FL.

Quashanna Price; Mary R. Williams, MS; Michael E. Sigman, PhD, “Fourier-Transform Infrared Spectroscopy Investigations of Smokeless Powders”, American Academy of Forensic Sciences 67th Annual Scientific Meeting, Orlando, FL, February 16-21, 2015.

Dana-Marie K. Dennis, BS; Jessica L. Frisch-Daiello, PhD; Mary R. Williams, MS; Michael E. Sigman, PhD, “Smokeless Powders Database at the National Center for Forensic Science”, American Academy of Forensic Sciences 67th Annual Scientific Meeting, Orlando, FL, February 16-21, 2015.

Mary R. Williams MS* and Michael E. Sigman, “How Forensic Databases Aid Forensic Scientists”, 7th European Academy of Forensic Science Conference, Prague, Czech Republic, September 6 – 11, 2015.

Dana-Marie Dennis, Mary R. Williams and Michael E. Sigman, “Investigative Predictions of Smokeless Powder Manufacturers”, American Academy of Forensic Sciences 68th Annual Scientific Meeting, Las Vegas, NV, February 22 – 27, 2016.

Appendices

Appendix A: NCFS Smokeless Powders Analysis Method Year 2010

- Micrometry
    - A stage micrometer is used to calibrate the Leica Stereomicroscope micron bar in order to make measurements of the smokeless powder kernels. Measurements are made in millimeters rounded to one decimal place.
    - The smokeless powder kernels are placed on a white background, and a total of fifty kernels are photographed. The micrographs are taken at X10.1 magnification.
    - Double-sided tape is used to stand kernels, such as short tubed powders upright so that micrographs may be taken for measurement of thickness. Thickness measurements are not recorded for discs and flattened ball samples.
    - ImageTool software is used to calibrate the micron bar and make measurements of each of the kernels in the micrographs. Calibration of the micron bar is done when the micrograph image is at a zoom of 1:1 and measurements are taken when the micrograph image is at a zoom of 1:2. (The software may be downloaded from http://ddsdx.uthscsa.edu/dig/itdesc.html.)
    - Length and thickness measurements are taken from the images of the kernels in the micrographs. Thickness measurements of the short tube kernels are obtained from the micrographs taken using the double-sided tape.
    - The averages and standard deviations of the measurements are calculated and the maximum and minimum measurements noted of the 50 kernels measured.
- GC-MS
- Extraction

    - The smokeless powder kernels are cut, and approximately 10 mg of the cut kernels are placed into a previously weighed small vial. Research indicates that cutting the kernels is necessary for the optimal extraction of the analytes.

    - 300 µL of methylene chloride with 10 ppm undecane are added to the contents of the tube to facilitate extraction (undecane functions as an internal standard). The tube is allowed to sit for three hours.

    - Remove the supernatant from the vial and place into GC vial (with 200 µL vial insert) for analysis. A 1:10 dilution is also analyzed

- Analysis

    - Analyses of the samples are conducted using a Hewlett-Packard 6890 gas chromatograph interfaced to a 5973 mass spectrometer.

    - One micro liter injections of the sample are introduced using an Agilent ALS autosampler G2614A splitless injector.

    - The capillary column is an HP-5MS 5% phenyl methyl siloxane with a nominal diameter of 250.00 µm, a film thickness of 0.25 µm and a length of 30 meters.

    - The injection port temperature was maintained at 170°C.

    - Helium carrier gas is maintained at a flow rate of 1.2 mL/min on the column with an average velocity of 40 cm/sec.

    - The initial oven temperature of 40°C is held for 1 minute, followed by a temperature ramp of 25°C/min to a final temperature of 280°C, which is held for three minutes.

    - The mass spectrometer transfer line is maintained at 250°C, with a source temperature of 230°C and a quadrupole temperature of 150°C.

    - Mass spectra are scanned between 43.0 and 400.0 m/z. The solvent delay is 4.00 minutes.

    - Positive identification of an analyte is determined by a 3:1 signal-to-noise ratio, the presence of at least three major ions, and its retention time as compared to a standard.

- ATR-FTIR
  - Extraction
    - Approximately 2.5 mg of smokeless powder kernels are added to a small vial with 80 µL of acetone and allowed to sit for thirty minutes.
    - The supernatant is spotted on a microscope slide, the solvent allowed to evaporate, followed by another application of the supernatant (added as a second layer to the first spot) and subsequent solvent evaporation. The film is analyzed using ATR-FTIR.
  - Analysis
    - Analyses of the samples are conducted on an ATI Mattson Infinity Series FTIR and Spectra Tech IR Plan Advantage IR microscope.
    - The internal reflectance element used to analyze the sample is silicon.
    - Four spectra from four different areas of the film are collected in absorbance mode using OMNIC software with the number of scans totaling 32 and a resolution of 4. An ATR correction was applied using the OMNIC software which corrects for the variance in the penetration depth. The data image is reported in percent transmittance. The spectrum with the flattest baseline and best resolution is chosen and presented in transmittance.

Table 1. List of Analytes, CAS Numbers, and Three Major Ions (m/z) with Relative Intensities of the Target Analyte

Analyte	CAS Number	Mass-to-Charge	% Abundance*
Nitroglycerine	55-63-0	46	100
		76	15
		43	13
Camphor	21368-68-3	95	100
		81	64
		108	41
Dibutyl Phthalate	84-74-2	149	100
		223	6
		205	5
Diethyl Phthalate	84-66-2	149	100
		177	26
		105	7
Dipentyl Phthalate	131-18-0	149	100
		150	9
		237	8
Diphenylamine	122-39-4	169	100
		84	10
		51	7
Dioctyl Phthalate	117-81-7	149	100
		167	35
		57	17
Ethyl Centralite	85-98-3	120	100
		148	81
		268	46
Methyl Centralite	611-92-7	134	100
		240	62
		106	45
2,4-Dinitrotoluene	121-14-2	165	100
		89	50
		63	27
4-nitrodiphenylamine	836-30-6	214	100
		167	83
		184	26
2-nitrodiphenylamine	119-75-5	214	100
		167	82
		180	37
4-nitrosodiphenylamine	156-10-5	167	100
		198	72
		184	12
Undecane **	1120-21-4	57	100
		43	72
		71	55

* % Abundance determined using standards and in-house GC-MS method

**denotes internal standard

Appendix B: NCFS Smokeless Powders Analysis Method Year 2014

- Micrometry
    - A stage micrometer is used to calibrate the Leica Stereomicroscope micron bar in order to make measurements of the smokeless powder kernels. Measurements are made in millimeters rounded to one decimal place.
    - The smokeless powder kernels are placed on a white background, and a total of fifty kernels are photographed. The micrographs are taken at X10.1 magnification.
    - Double-sided tape is used to stand kernels, such as short tubed powders upright so that micrographs may be taken for measurement of thickness. Thickness measurements are not recorded for discs and flattened ball samples.
    - ImageTool software is used to calibrate the micron bar and make measurements of each of the kernels in the micrographs. Calibration of the micron bar is done when the micrograph image is at a zoom of 1:1 and measurements are taken when the micrograph image is at a zoom of 1:2. (The software may be downloaded from http://ddsdx.uthscsa.edu/dig/itdesc.html.)
    - Length and thickness measurements are taken from the images of the kernels in the micrographs. Thickness measurements of the short tube kernels are obtained from the micrographs taken using the double-sided tape.
    - The averages and standard deviations of the measurements are calculated and the maximum and minimum measurements noted of the 50 kernels measured.
- GC-MS
- Extraction
    - The smokeless powder kernels are cut, and approximately 10 mg of the cut kernels are placed into a previously weighed small vial. Research indicates that cutting the kernels is necessary for the optimal extraction of the analytes.

    - 300 µL of methylene chloride with 10 ppm undecane are added to the contents of the tube to facilitate extraction (undecane functions as an internal standard). The tube is allowed to sit for three hours.

    - Remove the supernatant from the vial and place into GC vial (with 200 µL vial insert) for analysis. A 1:10 dilution is also analyzed
  - Analysis

    - As of July 2013, analyses of the samples are conducted using an Agilent 7890A gas chromatograph interfaced to a 5975C inert XL EI/CI triple-axis mass spectrometer. Analyses are performed in EI.

    - As of July 2013, one micro liter injections of the sample are introduced using an Agilent ALS autosampler G4513A splitless injector.

    - As of July 2013, the capillary column is a Restek Rtx-5MS, 5% phenyl methyl siloxane with a nominal diameter of 250.00 µm, a film thickness of 0.25 µm and a length of 30 meters.

    - The injection port temperature was maintained at 170°C.

    - Helium carrier gas is maintained at a flow rate of 1.2 mL/min on the column with an average velocity of 40 cm/sec.

    - The initial oven temperature of 40°C is held for 1 minute, followed by a temperature ramp of 25°C/min to a final temperature of 280°C, which is held for three minutes.

    - The mass spectrometer transfer line is maintained at 250°C, with a source temperature of 230°C and a quadrupole temperature of 150°C.

    - Mass spectra are scanned between 43.0 and 400.0 m/z. The solvent delay is 4.00 minutes.

    - Positive identification of an analyte is determined by a 3:1 signal-to-noise ratio, the presence of at least three major ions, and its retention time as compared to a standard.

- ATR-FTIR

  - Extraction

  - Approximately 2.5 mg of smokeless powder kernels are added to a small vial with 80 μL of acetone and allowed to sit for thirty minutes.
  - The supernatant is spotted on a microscope slide, the solvent allowed to evaporate, followed by another application of the supernatant (added as a second layer to the first spot) and subsequent solvent evaporation. The film is analyzed using ATR-FTIR.
- Analysis
  - As of September 2014, analyses of the samples are conducted using a Bruker LUMOS (© Bruker Optik GmbH2012) FT-IR microscope.
  - The internal reflectance element used to analyze the sample is germanium.
  - Four spectra from four different areas of the film are collected in absorbance mode using OPUS software (version 7.2 Build: 7, 2, 139, 1294) with the number of scans totaling 64 at a resolution of 4 cm^{-1}. An ATR correction was applied within the software which corrects for the variance in the penetration depth. The data image is reported in percent transmittance. The data is collected within the 4000 – 600 cm^{-1} range. The spectrum with the flattest baseline and best resolution is chosen and presented in transmittance.

Table 1. List of Analytes, CAS Numbers, and Three Major Ions (m/z) with Relative Intensities of the Target Analyte

Analyte	CAS Number	Mass-to-Charge	% Abundance*
Nitroglycerine	55-63-0	46	100
		76	15
		43	13
Camphor	21368-68-3	95	100
		81	64
		108	41
Dibutyl Phthalate	84-74-2	149	100
		223	6
		205	5
Diethyl Phthalate	84-66-2	149	100
		177	26
		105	7
Dipentyl Phthalate	131-18-0	149	100
		150	9
		237	8
Diphenylamine	122-39-4	169	100
		84	10
		51	7
Dioctyl Phthalate	117-81-7	149	100
		167	35
		57	17
Ethyl Centralite	85-98-3	120	100
		148	81
		268	46
Methyl Centralite	611-92-7	134	100
		240	62
		106	45
2,4-Dinitrotoluene	121-14-2	165	100
		89	50
		63	27
4-nitrodiphenylamine	836-30-6	214	100
		167	83
		184	26
2-nitrodiphenylamine	119-75-5	214	100
		167	82
		180	37
4-nitrosodiphenylamine	156-10-5	167	100
		198	72
		184	12
Undecane **	1120-21-4	57	100
		43	72
		71	55

* % Abundance determined using standards and in-house GC-MS method

**denotes internal standard

Table 1. List of Analytes, CAS Numbers, and Three Major Ions (m/z) with Relative Intensities of the Target Analyte

Analyte	CAS Number	Mass-to-Charge	% Abundance*
Nitroglycerine	55-63-0	46	100
		76	15
		43	13
Camphor	21368-68-3	95	100
		81	64
		108	41
Dibutyl Phthalate	84-74-2	149	100
		223	6
		205	5
Diethyl Phthalate	84-66-2	149	100
		177	26
		105	7
Dipentyl Phthalate	131-18-0	149	100
		150	9
		237	8
Diphenylamine	122-39-4	169	100
		84	10
		51	7
Dioctyl Phthalate	117-81-7	149	100
		167	35
		57	17
Ethyl Centralite	85-98-3	120	100
		148	81
		268	46
Methyl Centralite	611-92-7	134	100
		240	62
		106	45
2,4-Dinitrotoluene	121-14-2	165	100
		89	50
		63	27
4-nitrodiphenylamine	836-30-6	214	100
		167	83
		184	26
2-nitrodiphenylamine	119-75-5	214	100
		167	82
		180	37
4-nitrosodiphenylamine	156-10-5	167	100
		198	72
		184	12
Undecane **	1120-21-4	57	100
		43	72
		71	55

* % Abundance determined using standards and in-house GC-MS method

**denotes internal standard