Final Report

Quantitative / Statistical Approach to Bullet-to-Firearm Identification with Consecutively Manufactured Barrels

PIs: Peter Striupaitis
Adjunct Lecturer in Forensic Science
University of Illinois at Chicago
Chicago IL
Contract Firearms Examiner / Trainer
Illinois State Police
Forensic Science Center at Chicago
Chicago IL

R.E. Gaensslen
Forensic Science Group
Dept of Biopharmaceutical Sciences
College of Pharmacy
University of Illinois at Chicago
Chicago IL

With the technical and laboratory assistance of:

Cole Goater, M.S.

Project Period: October 1, 2003 – January 30, 2005
This project consisted of three parts. Two were part of the original proposal, and the third was added following initial review.

For many years, it has been an article of faith among firearms examiners -- and by extension, other forensic scientists -- that the rifled barrels (of handguns and rifles) impart both class and individual characteristics to bullets fired through them. The individual characteristics on bullets take the form of striation markings imparted onto the bullet by its movement along the stationary barrel. The imparted markings are the result of machine tooling during barrel fabrication and rifling. When a "test" bullet fired from a barrel is compared side-by-side with a "questioned" bullet from a case, and the individual striations marking match up sufficiently, an identification can be made - it can be stated that the questioned bullet was fired from the same weapon as the test bullet.

In its decision in Daubert et ux et al vs. Merrell Dow Pharmaceuticals (1), and in subsequent decisions, primarily GE v. Joiner and Kumho Tire v. Carmichael (2) intended to clarify the intent of Daubert, the U.S. Supreme Court has made clear that a scientific basis or underpinning will be required if evidence is to be admitted into the federal courts as "scientific evidence" through an expert witness. The ruling did not mean that technical evidence introduced through experts was inadmissible, but that it was not admissible as "scientific evidence." The ruling applies to the federal courts, but a number of states have adopted Daubert standards as their own. Previously, the standard was "general acceptance in the scientific community to which it belongs," which had been established in Frye v. U.S. in 1923. Obviously, Frye is a more liberal standard. Not too long after Daubert was decided, lawyers challenged the admissibility of a fingerprint identification in a federal court in Philadelphia on the basis of Daubert. Initially, the judge decided in favor of the defendant, but later reversed himself. The case caused a furor in the latent fingerprint identification community. Latent print examiners and their adherents took the position that the validity of fingerprint individuality and unique identifications had long since been established through their practice, and that any challenges to this position were unthinkable - and could only be motivated by defense attorneys on judicial "fishing expeditions" encouraged by "defense" experts with political motives. That chapter ended without further ado. But no one thinks the matter is settled. Even the people who think the challenges are unwarranted and thus improper (probably most of the latent print examiners) know that there will be other challenges.

The recent misidentification of a fingerprint in an important anti-terrorism case by no less than the FBI Laboratory (3) has not helped to support the proposition that all is well. Briefly, the FBI experts misidentified a latent print from the Madrid train bombings as belonging to a Portland, Oregon, attorney who was a convert to Islam, and had represented muslim clients. After a few days, the FBI admitted that the fingerprint was not his and apologized.

The Chicago Tribune, which has recently been a relentless critic of forensic sciences (see below), jumped on the fingerprint story. Their slant - which supported their assertions in a recent five part series on how terrible forensic science is - was that the FBI examiners were under pressure from their superiors to make the identification, and as a result were unduly influenced (biased) by factors external to the examination itself.

The post-Daubert activity surrounding fingerprint identification follows an earlier, and ultimately more or less successful challenge to the individuality of handwriting (4). In Starzecpyzel, which was decided after Daubert, but before Joiner and Kumho Tire, the court ruled that handwriting identification was not science - did not have an appropriate scientific basis.
However, since the court had decided it was not science, it therefore did not fall under the Daubert standard, and was still admissible as technical or specialized knowledge. Since then, handwriting identifications by forensic document examiners have had mixed reviews in the federal district courts, some finding them admissible and others not. In Prime, where there was an extensive hearing on the motion in limine to exclude the testimony, it was ultimately admitted. But, as with fingerprint identification, no one thinks the matter is settled.

It is of considerable interest to note that the Chicago Tribune's five-part series noted earlier (6) comes down hard on what we would call “pattern” evidence – the criticism is primarily that pattern evidence identifications do not meet Daubert criteria, i.e., they do not have an appropriate scientific basis. Pattern evidence includes handwriting, fingerprints, firearms markings, tool marks, as well as footwear or fabric impressions, lip or ear impressions, tire tracks, bite marks, and other similar types of patterns.

There has not been a judicial challenge that we know of against a firearms or tool mark identification, but everyone thinks it is just a matter of time.

Looking at this whole picture from a research standpoint, a number of people have undertaken studies to establish a scientific basis for pattern evidence identifications, if not of outright individuality. If pattern evidence identifications are to meet Daubert standards, there need to be objective standards established for the examinations that can be followed universally. In addition, there need to be studies using objective measures, that show, in pair wise comparisons of patterns, that the likelihood of chance duplication is small. The trick is to try and design suitable “objective measures.” This term means something external to the examiner, but which can measure – at least in part – what the examiner sees with his eyes. These are not easy problems to solve, but approaches are possible.

In the fingerprint pattern arena, the sort of background studies that are needed to help establish a basis for “identifications” have already been done to some extent by biometrics specialists. The purpose of those studies is establishing bases for live-scan fingerprint readers to verify identity, instead of paper/document forms of identification. In the handwriting arena, the problem is more difficult because of the considerable intra-individual variation in handwriting. But approaches have been devised, built around the kind of technology the postal service uses to “read” handwriting (like zip codes off letters). These technological methods basically abstract class or individualizing features from the pattern, and assess them according to some pre-established algorithm. In the case of fingerprint biometrics, the algorithm is designed to test for “matches.” The operator of the technology can manipulate the range of features, match thresholds, etc. to come up with suitable operational match algorithms. But the ability to manipulate feature abstraction provides the tool that is needed to test the frequency of chance duplication, given some prior set of individualizing features and match criteria.

Another approach that has merit, at least in the case of fingerprint and firearms patterns (where intra-individual variation is not an issue) is to look into “objective” methods for differentiating very similar patterns – those expected to be the most similar. In the fingerprint case, these would be the prints of identical twins. In the firearms case, these would be the individualizing markings imparted onto bullets by consecutively manufactured barrels. This last formed the basis for our project.

We set out to obtain two sets of consecutively manufactured barrels, from different manufacturers, but of the same caliber (.45).
Bullets could then be fired from these barrels and subjected to imagings or measurements that might be “objective” mirrors of human examination/comparison. In part I, the bullets were imaged in a commercial IBIS system, and bitmap images produced. In part II, the bullets were subjected to measurements designed to “profile” the striation markings in a prototype instrument developed by a commercial enterprise. In part III, the bullets were assembled into validation “tests” and sent out to firearms examiners for comparison.

Consecutively Manufactured Barrels / Test Firings

We settled on .45 caliber bullets in part because they are large (and perhaps a little easier to examine as a result), and because no one had ever done a consecutively-manufactured barrel study using .45 caliber barrels/bullets. Such studies have been done before, using 9 mm caliber barrels/bullets (7).

The HiPoint firearms company in Galion OH agreed to furnish us with six consecutively manufactured .45 cal barrels gratis, as a service to the firearms examiner profession. We accepted this offer. The PIs went to the facility, and observed the cutting and rifling of the research barrels. As each was completed, we observed an operator stamp a number onto the barrel using a steel numbering punch. There was no opportunity for any mix-up or confusion of the barrels once the number was permanently stamped onto it.

A similar procedure was followed with Machine Solutions in Lenoir City TN, except that only PI Striupaitis was able to be present in person at this facility, and the procedure took longer. Machine Solutions makes .45 cal handgun barrels for Springfield. We had to purchase these barrels from the maker.

The barrels, and bullets fired from them, are called “HiPoint” and “Springfield” in the remainder of the report.

The barrels were mounted one by one onto pistol frames at the test firing range at the Illinois State Police Forensic Science Center – Chicago, and 230 gr. copper-jacketed bullets fired through each barrel into a water tank, from which they could be readily recovered. For Parts I and II of the project, two bullets were fired through each barrel, for a total of 12 HiPoint and 12 Springfield. For Part III of the project, bullets were fired through each of the barrels according to the scheme shown in Appendix 1. Sixty human examiner tests were constructed, thirty of which were HiPoint and thirty of which were Springfield. In tests 1-30, only one K (known) bullet from each barrel in the series was included, but in tests 31-60, duplicate K bullets for each barrel were sent. This matter is discussed further below. To avoid any specimen mix-ups, the bullets were physically marked with an identifier (like “1A”) as soon as they were collected from the water tank. In this way, even if a bullet got into the wrong-labeled envelope, the bullet itself was marked.

The same set of bullets was used for Parts I and II. The bullets were imaged at FTI in Montreal, then returned to Illinois, then furnished to IAI for Part II.
Part I. IBIS Imaging

IBIS technology has gone through some technological iterations and shakeouts, and is now the sole province of a company called FTI (Forensic Technology, Inc.) headquartered in Montreal, PQ Canada. FTI agreed to produce images of our test bullets, and provide those images. PI Striupaitis was present during the imaging procedures.

The primary purpose of IBIS is to collect and store images of the land impressions on bullets, and the firing pin indentation and other markings on cartridge cases. Given a barrel model with some number of lands, say six for example, there would be six images associated with an exemplar bullet from that barrel – one image for each land impression (sometimes called a land engraved area or LEA). A questioned bullet can then be imaged, and the image compared with existing image files. If the questioned bullet is in good condition, and six land impressions can be imaged, and they can each be pair-wise compared with the six on any exemplar. The system thus enables connecting fired evidence from different cases, and in this way assisting investigations. It is not and has never been intended to substitute for a human firearms examiner, but only to sort through large quantities of evidence and generate a manageable list of “possibles.”

The algorithms used by FTI in making these comparisons are proprietary, and FTI was clear that they were unwilling to share the details of their comparison algorithms. They did, however, provide the results of their comparison computations.

When two bullets are compared in the IBIS system, three different scores result. These are max phase, peak phase, and peak score. Max phase has, in effect to getting the bullets into “register.” On bullets with, for example, six twist and therefore six LEAs, there are 36 comparison scores: LEA1 of bullet 1 with LEA1 of bullet 2, then LEA2 of bullet 2, then LEA3 of bullet 2, and so on. The highest score here will indicate that the proper LEAs are being compared (highest peak phase). Then a peak score is computed, and the highest peak scores among a group of bullets being compared are noted. In this study, each bullet was compared with every other bullet in the study (but not with itself). Thus, 1A was compared with 1B (its sister bullet), with 2A, 2B, 3A, 3B etc. By “sister” in this study, we mean duplicate. Bullets 1A and 1B were fired from barrel 1, while 2A and 2B were fired from barrel 2, etc. Barrel 2 was the next one manufactured with the same rifling tool after barrel 1; barrel 3 was the next after barrel 2, etc.

Once all bullets are entered, a correlation is performed for each bullet against all other bullets in the test group. For each bullet a correlation result is produced. For each correlation result produced, the three score arrays noted above are generated. Each score array is sorted, and knowing which bullet is the “sister,” each array is consulted to determine its position. Out of the three arrays the lowest position is noted. And the lowest position occupied among all the arrays is recorded. From these data, a probability density vs position can be computed – the probability that the sister bullet of any given bullet being compared will be one of those in the top $n$ positions, where $n$ is the position number being plotted – indicating a perfect match. These plots are shown in Figures 1 and 2. In short, the Springfield bullets are properly matched 100% of the time.
Figure 1.

Figure 2.
So are all but two of the HiPoint bullets. But even in the two exceptional cases, the probabilities are 83 and 92 percent, respectively.

The reason for the better match results with Springfield isn’t apparent from any of the data. It is likely, however, that they are the result of the differences in rifling methods.

Firearms barrels of the type used in these studies can be rifled in one of several ways. The method used by HiPoint is called “button” rifling. Here, a more or less cylindrical-shaped plug called a button, and made of metal harder than the barrel, is forced down the barrel with a twisting action. The rifling is swaged into the barrel metal. The button does not cut the barrel metal. Compared with processes that do involve cutting of the barrel metal, the swaged rifling marks are shallower and not as distinct. As a result, the land impressions they make on bullets are not always as distinct as those made by barrels rifled by way of a cutting process. Machine Solutions, the barrel manufacturer for Springfield, uses “broach” rifling. Here, a hardened steel broach is drawn through the barrel, and introduces the rifling by cutting the barrel metal. We purposely chose these manufacturers because of the difference in rifling techniques.

It is interesting that the IBIS imaging system showed a small difference in the ability to identify a “sister” bullet that appears to be a result of the method used for barrel rifling in the two cases.

It is also fair to say that the IBIS imaging system correctly distinguished all the Springfield barrel bullets and nearly all the HiPoint barrel bullets 100% of the time. Even in the HiPoint cases where it did not, the probabilities were high. This result provides an objective demonstration of distinguishing between extremely similar, but slightly different, bullets from consecutively manufactured barrels from two different sources, using two different rifling methods.

On the accompanying CD, appended to this report, you will find the data summarized above in a folder labeled “FTI.” The HiPoint bullet (7AB through 12 AB) data is in the “High-Point” folder. The MS Excel spreadsheet file delineates and summarizes the numerical correlation data for each bullet. The spreadsheet also contains the probability vs position plot. A summary of the numerical correlation data for each bullet is in a set of text files named “RES02_NIJ_MRFC_PPS_FIXED_07_A_BUL_80042-0_0_Benchmark.txt,” where the bolded designators change with each bullet. Finally, in addition, is the complete set of bitmap images generated by the IBIS scans. There is a separate folder for each bullet. Within folder “7A,” for example, there are seven files named “FT141_NIJ_MRFC_PPS_FIXED_07_A_LEA_21375_1_1_Real.bmp,” where the bolded characters change for each LEA being imaged. In the HiPoint series, each bullet has seven images. In the Springfield series (bullets 1AB through 6AB), each bullet has six.

In addition you will find a folder called “NIJ Assembled Mosaics,” which in turn contains a “HiPoint” and a “Springfield” folder. Within each of these are 2-D and 3-D bitmap assembled mosaic images of each bullet. These are created by imaging a series of 1.5x1.5 mm patches along the optical axis of the bullet until the whole bullet circumference is captured. Overlap area is imaged, and then used to line up the patches to assemble the mosaic. The 3-D technology is rather newer, and was not available for use in the main study. It is capable (according to FTI) of capturing more data, as one might expect, than the 2-D system. The 3-D system was used to provide the assembled mosaics, but not for the comparative analyses referenced above.
Part II. SciClops™ “Profilimeter” Data

Bachrach reported in 2002 (8) on a novel, laser-based profilimeter that had been designed under grants from the NIJ (97-LB-VX-008) and the NSF (DMI-9801361). This device, which is trademarked, and could be commercialized by Intelligent Automation, Inc. (IAI) in Rockville MD, was constructed as a possible alternative to the much more expensive IBIS system, or perhaps as a screening device that could save time in reducing the number of possible striation pattern matches when numerous pair-wise comparisons were necessary. The device was benchmark tested on copper-jacketed 9 mm bullets fired through the barrels of Baretta Model 92 pistols. The evaluation used six bullets, two fired through each of three barrels. The data showed that the bullets fired from the same gun (barrel) could be distinguished from those fired from different guns (barrels).

Figure 3. The SciClops Set-up.

Figure 3 shows the set-up for this device. Scanning one “circumference” in an imaginary plane through the bullet (Figure 4) generates a large set of depth-measurement data. In theory, these metrics should accurately reflect the land impression area markings. Here, as in the IBIS case discussed above, each LEA is accounted for separately, and a scoring method is used to compute a comparison score for two bullets when they are in proper “register,” i.e., when LEA1 is compared to LEA1, LEA2 to LEA2, and so forth.

In the CD appended to this report, there is a folder called “IAI” that contains all the raw data provided to us by IAI. See the excel file called “ID keys” to translate between our bullet numbers and their arbitrarily assigned ones. The data actually abstracted from each bullet is contained in a large series of files in the folder “Bullet Data.” Both the zipped and unzipped versions of these files are there. The folders also have “normalized” data. This normalized data is the rawest form of the data available. It is abstracted from the bullet after employing some filters that correct for things like misalignment of the bullet axis with respect to the laser scanner, etc. It is basically depth data. There are \( n \) sets of this data for each bullet where \( n \) is the number of LEAs (six in Springfield, and seven in HiPoint). Although it is uninformative, you can get an idea of the nature of this data by plotting it (see Figure 5).

Figure 4.

Figure 5 shows a “depth” profile through one LEA of one bullet (in this case, LEA_0 of bullet 1674 (or 2A) from a Springfield barrel. You could construct a plot like this for every LEA of every bullet, but they are not helpful. There is too much similar data to be able to abstract any dissimilarities.

Using a different, more sophisticated plotter program, this normalized raw depth data is plotted on the same coordinates in Fig. 6A for the homologous LEA of “sister” bullets (2A and 2B) and in Fig 6B for the homologous LEA of non-sister bullets (2A vs 6A).
It is obvious that these plots do not provide useful differences between the cases. The normalized data for all the bullets are included on the CD in three excel spreadsheets.

Of more interest are the excel files called "HiPoint Data" and "Springfield Data," which hold the correlation scores for all the comparisons done. Here you see every permutation of LEA to LEA for every bullet in the series.

Figure 5. Normalized raw data

From the "best orientation," i.e., the one in which LEA1 is compared to LEA1, LEA2 to LEA2, etc., can be derived a best overall "correlation score." This is a number between zero and one, where zero represents no correlation and one would represent perfect correlation. The overall score can be ascertained for every possible pair-wise comparison in the series. These scores have been assembled into tables that are in Appendix 1 (Table 2A for HiPoint, and Table 2B for Springfield). In these tables, the sister bullets are highlighted in yellow. And, any correlation scores among non-sister bullets that exceed those for sister bullets are in red. Visual inspection of these tables will readily show that the SciClops system was not able to discriminate the minute, subtle differences between sister and non-sister bullets as well as IBIS. The data is shown in a more visually friendly format in Figures 7A and 7B.

IAI had never before profiled .45 cal bullets, nor had it ever attempted to profile (and discriminate) bullets that were maximally similar but from different barrels. IAI noted that its technology is consistently under improvement, and that newer versions might be able to do better than the version used to collect this data. Even though the results were not perfect, they were generally in the proper direction, and the system correctly discriminated two pairs of sister bullets from non-sisters in each manufacturer series.
Although it is speculative, there may be different reasons for SciClops performing less accurately than IBIS in the two different barrel manufacturer series. It is possible that, as noted above in the IBIS results discussion, the different rifling techniques of the two manufacturers cause markings that present differently to the technology. The depth of the striation markings in the LEAs is clearly a much greater issue for SciClops than it may be for IBIS, which is performing image analyses rather than determining depth. The differences in marking ability of the two sets of barrels are a function of the hardness of the barrel metals (Springfield is a harder, more expensive, barrel) and of the rifling technique (Springfield is broach rifled, which involves cutting and removal of some metal during the process).
Figure 7A. HiPoint Pair wise Correlation Scores

Figure 7B. Springfield Pair wise Correlation Scores
Part III. Human Firearms Examiner Validation Tests

As noted, the ability of firearms examiners to correctly identify bullets fired from consecutively manufactured barrels as being from different barrels has been verified before (7,9), although not with .45 cal. bullets.

Before proceeding with this phase of the project, the protocol was submitted to and approved by the UIC Institutional Review Board (Protocol # 2004-0067).

Firearms examiners for potential participation in the project were recruited by direct contact, through the AFTE (Association of Firearms and Toolmark Examiners) Directory, and by way of the AFTE web site. Ultimately, 55 separate examiners said they were willing to participate, and were provided with test sets. Some of these were examiners within the same lab, but we indicated that each examiner was to do the test set provided to him or her independently. The response rate was well under 100%, however, because of the difficulty associated with this particular validation and the correspondingly large amount of time it took to complete it.

Besides making an effort to issue approximately equal numbers of HiPoint and Springfield test sets, another issue arose after we had issued the first group of test sets. In many labs in real casework circumstances, it is routine to fire two bullets from a suspected weapon, so the examiner has two Ks that he knows are identical to use as a “benchmark” for “sameness.” When we designed the first group of tests, those numbered 1-15 (HiPoint) and 16-30 plus 46 (Springfield), we provided a single K for each barrel (i.e. K1 through K6, where K1 was fired from barrel 1, K2 was fired from barrel 2, etc.) plus the twelve Q bullets. When this issue was raised by someone participating in the test, we decided to make up the remainder of tests (i.e. those numbered 31-42 HiPoint, and those numbered 47-58 Springfield) with duplicate Ks from each barrel, i.e. K1A and K1B were fired through barrel 1, K2A and K2B were fired through barrel 2, etc. There were no differences in results obtained that could be definitely attributed to this slight, but potentially important, difference in the construction of the test comparison sets.

Of the 55 test sets issued, 16 examiners (29%) returned results. Table 3 below shows the details.

<table>
<thead>
<tr>
<th>Test Sets Issued</th>
<th>Test Sets – Results Returned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One K</td>
</tr>
<tr>
<td>HiPoint</td>
<td>15</td>
</tr>
<tr>
<td>Springfield</td>
<td>16</td>
</tr>
</tbody>
</table>

Although the number of test sets returned was small in comparison to the number issued, sixteen test sets, consisting of 6 Ks (ignoring the duplicate Ks) and 12 Qs required a total of 171 separate pairwise comparisons. In this set of results there was not a single instance of a misidentification – that is, not a single instance where an examiner reported a Q as matching an incorrect K. There were a few results that warrant separate comment. In one test, three Qs were reported as inconclusive. In another test from a European examiner, eight of the twelve Qs were reported as “strongly supported” matches for the Ks. All of these were, in fact, correct. In Europe, some examiners can follow such a less-than-definite reporting language protocol, but U.S. examiners do not.
In a third test set, three Qs were reported as having “no match,” when in fact they did have. One could characterize those results as “false negative,” failure to match a Q with a K in the set when a match in fact existed.

Summary

Efforts to use objective image comparison and bullet scanning technologies to distinguish bullets from consecutively manufactured handgun barrels from two manufacturers gave mixed results. The ability of a technology to reliably distinguish between matching and non-matching bullets, where the non-matching bullets were as close in pattern to the matching ones as is probably possible, would provide evidence that the distinctions could be made “objectively,” and independently of human eyes. That evidence is identical or very close to what seems to be needed to satisfy Daubert standards.

It is fair to say that the FTI IBIS image comparison technology correctly distinguished between all the Springfield barrel bullets, and between most but not all of the HiPoint barrel bullets. In the HiPoint cases that were not distinguished 100% of the time, they would be distinguished correctly at least 83% of the time. These results, although obviously limited to the materials used in the comparisons, provide strong evidence that barrel-to-bullet matching is objectively reliable.

The results with SciClops were less compelling. The results do not mean that bullet-to-barrel matching is not objectively reliable – rather, they mean that this version of the particular technology could not quite distinguish between these extremely similar yet different bullets as well as the image comparison technology did. In a number of cases, the numerical results made the correct distinctions, although they were close to one another. It is hard to say from this data that this technology differs in its ability to make distinctions between the manufacturers, because the results are very similar with both.

The human examiner results were as expected. We did not expect any misidentifications, and there were not any. It would have been preferable to have a higher return rate, and thus more comparisons in the overall sample. As noted, the “consecutively manufactured barrel exercise” has been done before, with the same outcome.
References


6. Chicago Tribune: “Forensics under the microscope.” Unproven techniques sway courts, erode justice, October 17, 2004; Arson myths fuel errors October 18, 2004; From the start, a faulty science, October 19, 2004; When labs falter, defendants pay, October 20, 2004; Scandal touches even elite labs, October 21, 2004.

7. Brundage DJ. The identification of consecutively rifled gun barrels. AFTE J 1998; 30(summer)


9. Miller J. An examination of two consecutively rifled barrels and a review of the literature. AFTE J 2000; 32(summer)