Forensic Laboratory Independence, Control, and the Quality of Forensic Testimony

Patrick Warren

May 10, 2014

Abstract

The relationship between forensic laboratories and the other institutions of law enforcement varies widely over space and time in the United States. Some jurisdictions have their own local lab within the police department, some locate the lab within the district attorney's office, others depend on a statewide lab system either independent or under the state police, and others still contract with a private lab to process their forensic evidence. These different organizational forms may shift the incentives lab technicians and managers have to provide timely and accurate analysis and testimony. In this paper, I investigate how the relationship the forensic lab bears to the rest of the structure of law enforcement correlates with the quality of forensic analysis and testimony in a sample of trials that eventually led to convictions and, later, exonerations. Both within evidence types and across all evidence types, the expert prosecution testimony provided by employees of public crime labs is consistently worse than expert prosecution testimony given by employees of private forensic labs working under contract for the state. These results are robust to controlling for case characteristics, defendant characteristics, time trends, and state fixed-effects. Furthermore, there do not seem to be significant differences among public crime lab types: validity rates for crime labs controlled by the local police are similar to those controlled by the state police and to those that are independent from the police, suggesting the dictating formal independence may not be enough to insulate public lab from whatever forces are driving their poor performance.

JEL Classification:

Keywords: Forensics, Law Enforcement, Trial Testimony, Public versus Private

1 Introduction

In the past decade, the forensic science community has been buffeted by a number of scandals. In Massachusetts in 2013, State chemist Sonja Farak pled guilty to tampering with drug evidence, potentially affecting 60,000 samples in 34,000 cases. In 2009, an audit of the Houston Police Department crime lab's fingerprint unit found irregularities in over half of the 548 cases reviewed. In 2010, lab workers in the North Carolina state forensic lab were found to have failed to turn over potentially exculpatory evidence, including in death penalty cases. The investigation of the State Bureau of Investigation blood serology unit yielded a total of 229 cases of misrepresentation of blood serology. Of the 229 cases, seven persons had been executed, others were on death row, and some had died in jail.

In 2009, the National Research Council of the National Academy of Science issued a report entitled "Strengthening Forensic Science in the United States: A Path Forward", in which a distinguished panel of scientists lays out a number of recommendations to improve the accuracy and reliability of forensic science in the United States. The first of these recommendations was to foster independent forensic organizations. In their words "[the current system] leads to significant concerns related to the independence of the laboratory and its budget. Ideally, public forensic science laboratories should be independent of or autonomous within law enforcement agencies. In these contexts, the director would have an equal voice with others in the justice system on matters involving the laboratory and other agencies." There has been no empirical analysis that could guide such an effort.

In fact, the relationship between forensic laboratories and the other institutions of law enforcement varies widely over space and time in the United States. Some jurisdictions have their own local lab within the police department, some locate the lab within the district attorney's office, others depend on a statewide lab system either independent or under the state police, and others still contract with a private lab to process their forensic evidence. These different organizational forms may shift the incentives lab technicians and managers have to provide timely and accurate analysis and testimony. That shift could affect outcomes throughout the criminal-justice process: crime rates, arrest rates, conviction rates, and the quality of forensic analysis and testimony.

In this paper, I investigate one particular outcome in this chain: the quality of forensic testimony. In the modern courtroom, forensic evidence plays a very important role. Trials for defendants facing probative forensic evidence are more likely to end in a conviction and result in longer sentences (McEwen 2011). But this impact depends on reliable testimony, and the anecdotes above suggest the high levels of reliability are not guaranteed. In each case, it was an employee or employees of a public crime lab who where implicated. Is there something about public crime labs that leads to unreliable testimony? Or is it simply a question of numbers, where the high number of public crime lab employees naturally leads to a greater number of extreme "bad apples"?

I investigate this question by comparing the rates of testimony validity across a sample of trials that eventually led to convictions and, later, exonerations. Both within evidence types and across all evidence types, the expert prosecution testimony provided by employees of public crime labs is consistently worse than expert prosecution testimony given by employees of private forensic labs working under contract for the state. These results are robust to controlling for case characteristics, defendant characteristics, time trends, and state fixed-effects. Furthermore, there do not seem to be significant differences among public crime lab types: validity rates for crime labs controlled by the local police are similar to those controlled by the state police and to those that are independent from the police, suggesting the dictating formal independence may not be enough to insulate public lab from whatever forces are driving their poor performance.

2 Forensic Testimony in Exonerations

2.1 Data

In 2009, Garrett and Neufeld (2009) performed a comprehensive analysis of the forensic evidence presented in the trials of defendants who were later exonerated by the Innocence Project through DNA evidence. There were 137 trials identified for which transcripts of the underlying testimony were uncovered. These trials were conducted between 1978 and 1999, with the mean and median year of 1987 and intra-quartile range of 1984 to 1990. In that study, the authors code the forensic testimony for the prosecution at the exonorees' original trials, for each type of forensic evidence and each forensic witness as "invalid" if the conclusion given was not supported by the empirical evidence.[1] These errors fell into one of six categories:

[1]Details on the coding can be found in the original article and the transcripts are posted at http://www.law.virginia.edu/html/librarysite/garrett_exoneree.htm

a) Non-Probative Evidence Presented as Probative: "The testimony disregarded that the evidence was non-probative, and instead the analyst provided a statistic purporting to include the defendant and implying that a percentage of the population was excluded." The most common error of this sort was when the defendent and victim shared a blood type, and a rape kit identified only that blood type. Testimony that excludes perpetrators of all other blood types is invalid, since the victim could have been the sole (or primary) donor, masking the blood type of the perpetrator.

b) Exculpatory Evidence Discounted: The most common error of this type was to provide speculative or entirely unsupported testimony "explaining away" evidence that did not fit with the prosecution's theory of the crime.

c) Inaccurate Frequency or Statistic Presented: These errors consisted primarily of basic mathematical errors in presenting population statistics.

d) Statistic Provided Without Empirical Support: Errors of this sort were unfounded guesses about population statistics. "For example, an analyst testifying that there was a 1 in 10,000 chance that the two hairs found at the crime scene could come from someone other than the defendant, despite there being no accepted population statistics for hair characteristics.

e) Non-numerical Statements Provided Without Empirical Support: "In the field of microscopic hair comparison, due to the lack of empirical data, the field adopted standards that the strongest statement of association that can be made by an analyst is that the hairs in question are "consistent" with the defendant's or "could have" come from the defendant." Adding qualifiers like "very probably" or "highly likely" is not empirically supportable.

f) Conclusion that Evidence Originated from Defendant: Claiming a unique match with no information about the population distribution. "Forensic disciplines involving impression evidence, such as bite mark and shoe print comparison, have not developed any objective criteria at all by which to judge assertions about the likelihood that crime scene evidence came from a particular defendant."

The validity rates in this (admittedly very selected) sample of trials were quite low. Of the 193 instances of forensic testimony (many trials had multiple evidence types), 47.7 percent were coded as valid. Serology (blood typing), the most common evidence type (100 instances), was even less valid than average, at exactly 40 percent valid. Hair evidence, the second most common type (60 instances), was valid 53.3 percent of the time.

From the trial transcripts, it is possible to find the identity of the forensic witnesses who provided the testimony and the organization who employed them at the time they provided the analysis. The employer of the witnesses providing the testimony falls into one of five categories: state labs (49 percent), public labs that are tied to local police (35 percent of forensic testimony units), private entities (8.3 percent), local public labs that are independent of the police (6.7 percent), and federal labs (1.5 percent). Again, the mix of employers is not identical across evidence types. For example, about 17 percent of dna evidence was provided by private lab employees, while 78 percent of finger print evidence came from local publics labs. At the extreme, bite-mark evidence is overwhelming provided by private entities (5 out of 6), all private dentists serving as consultants.

2.2 Results

Validity rates varied in this sample of trials across the identity of forensic witness. Figure 1 presents validity rates by lab type, pooling together all evidence types. The dots indicates mean validity rates, one the left axis, with error bars representing one standard deviation of the mean. The bars indicate the sample size, on the right axis. Pooling all evidence types together, there is no clear difference in validity rates across public labs. The federal labs seem better, but there are so few observations that it is difficult to come to firm conclusions. The private labs, in contrast, perform markedly better than all public counterparts. The differences are about 35 percentage points between the state/local public labs and the private labs– 45 percent valid versus over 80 percent.

These differences are not driven by forensic evidence type mix. Figure 2 breaks the analysis up by the four of the five most common evidence types: serology, hair, dna, and bite marks.[2] Across the two most common evidence types, serology and hair analysis, the pattern is nearly identical. The local and state public labs provide valid testimony about half the time, while the private lab testimony is valid in nine out of nine cases. In the handful of DNA cases, public labs get half the cases right

[2]Prints, the fourth most common type, is excluded, since there are no private labs presenting print evidence. But the print evidence is valid in 90 percent of cases, so its inclusion only reinforces the results in over pooled analysis

on average, while both private lab testimonies are valid.

Table 1 presents regression analysis to investigate whether these differences are statistically significant and robust to feasible control variables. In all regressions, the unit of observation is an individual testimony about a single evidence type, the outcome variable is equal to 1 for valid testimony, and local public labs make up the omitted class, so the coefficients on the other lab types should be interpreted as difference in validity rates between the indicated lab types and the validity rates at public labs affiliated with the local police. The columns represent the sort of forensic evidence included in the sample. The first column includes all observations, while the others are restricted to specific evidence types. Each panel is a different regression specification. The first panel includes only lab-control dummies. The second panel adds a dummy for whether the exonoree was convicted of a rape or sexual assault, convicted of murder, a linear control for year of initial conviction, and, in the case of the pooled sample, dummies for each of the forensic evidence types. The final panel then adds state-specific fixed effects. Samples sizes and mean validity rates are reported at the bottom of the table. Throughout, standard errors are clustered by the trial and given in parentheses, with asterisks in indicate significance at the 0.10 ($*$), 0.05 ($**$), and 0.01 ($***$) levels.

The basic results in the first panel mirror the graphical analysis, above. Across the board, there are large and significantly higher validity rates in private-lab testimony than in all types of public labs. The federal labs also perform better than the state/local labs, although the difference is only statistically significant in the case of serology.

In the second panel, the inclusion of the three available case-specific controls (and evidence-type dummies, for the pooled sample) only increases the correlation between ownership and validity. There also seems to be a significant improvement in validity over time, where testimony from trials one year closer to present has an increase in average validity of about three percentage points. There is no consistent relationship between type of conviction (Rape/Murder) and forensic validity rates.

Finally, the last panel includes state fixed-effects in addition to the control in Panel 2. The most important effect of this inclusion is to restrict the number of states that contribute to the estimation of the private-lab coefficient to those that have used both private and public labs: Arizona, Indiana, Louisiana, Massachusetts, New York, Oklahoma, Pennsylvania, and Texas. This specification should protect from unobserved state-specific drivers of both private-lab use and forensic validity.

If, for example, these states happened to have particularly high validity rates across all lab types, we might inappropriate attribute that difference to private lab use. In fact, the coefficients hardly move at all, suggesting the biases of these form are not driving our results.

3 Conclusion

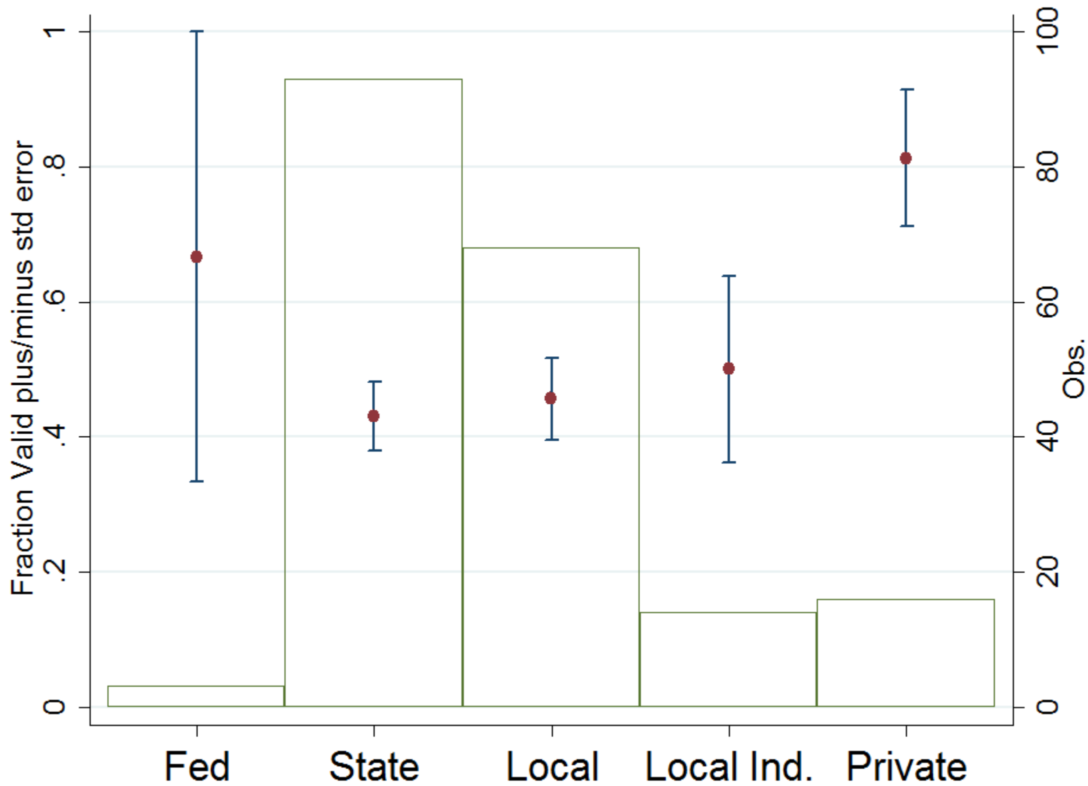
Taken together, these results demonstrate that in the set of cases for which defendants were convicted and exonerated by subsequent DNA testing, the validity of forensic testimony given by analysts from private labs was much higher than that of analysts from public labs, and the difference was much larger than the differences among public labs. These difference seem to extend across forensic evidence types and they are robust to controlling for some basic case characteristics, time, and any time-constant state-specific factors.

It is an open question whether these patterns of valid extend to non-exonorations or to more recent time periods. In the original study, Garrett and Neufeld (2009) analyze a set of otherwise similar cases that did not end in exoneration. They found that the overall rates of validity were indistinguishable from those in the exoneration cases. Unfortunately for the present questions, these non-exoneration cases were a much smaller sample drawn from only three states that make little or no use of private labs in this period, Virginia, Texas, and Missouri, so it is not possible to replicate the comparison made here. In fact only three cases used private labs, and they were all for DNA analysis. The validity rates for the local and state public labs in the non-exonoree sample, however, are equal or worse than rates in the exonoree sample, with 47 percent validity for the state labs and 35 percent for the local labs. Thus, to think the main results are non-representative, you would need to believe that selection into exonoree sample is related to exceptionally good performance by private labs.

References

Garrett, Brandon L. and Peter J. Neufeld, “Invalid Forensic Science Testimony and Wrongfuul Convictions,” *Virginia Law Review*, 2009.

Figure 1: Forensic Validity by Control of Lab



McEwen, Tom, "The Role and Impact of Forensic Evidence in the Criminal Justice System, Final Report," Technical Report, Institute for Law and Justice, Inc. 2011.

Figure 2: Forensic Validity by Control of Lab and Evidence Type

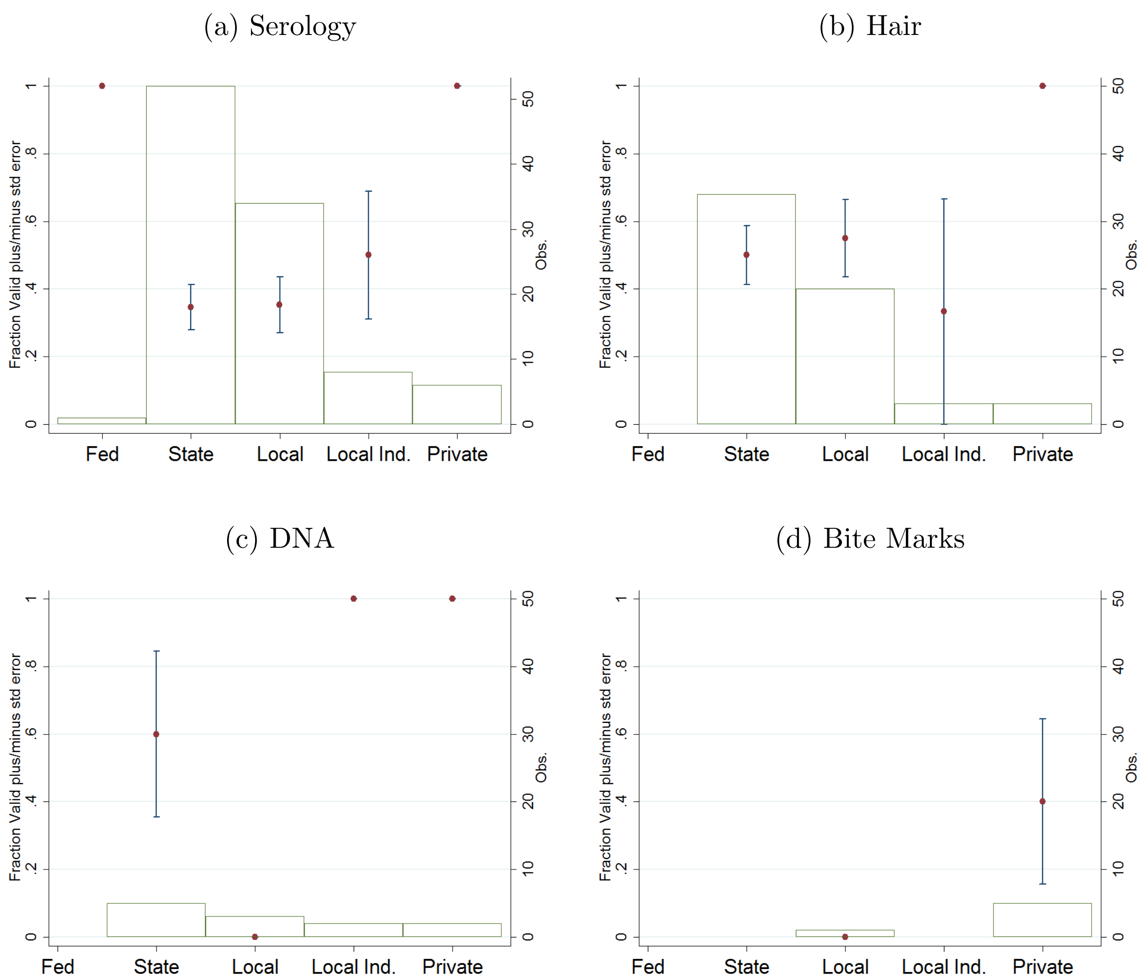


Table 1: Validity of Forensic Testimony by Control of Crime lab

	All Types	Serology	Hair	DNA	Bite
Panel 1: Simple Mean Differences					
Fed. Lab	0.21	0.65***	.	.	.
	(0.28)	(0.08)	.	.	.
State Lab	−0.03	−0.01	−0.05	0.60**	.
	(0.09)	(0.11)	(0.15)	(0.27)	.
Local Ind.	0.04	0.15	−0.22	1.00***	.
	(0.16)	(0.20)	(0.30)	(0.00)	.
Private	0.36***	0.65***	0.45***	1.00***	0.40
	(0.12)	(0.08)	(0.12)	(0.00)	(0.27)
Panel 2: + Rape Dummy, Year (+ Evidence-Type FE)					
Fed. Lab	0.40	0.72***	.	.	.
	(0.30)	(0.12)	.	.	.
State Lab	0.04	0.06	−0.04	0.72***	.
	(0.08)	(0.10)	(0.14)	(0.19)	.
Local Ind.	0.14	0.22	−0.20	1.95***	.
	(0.17)	(0.20)	(0.27)	(0.57)	.
Private	0.65***	0.79***	0.45***	1.71***	0.71
	(0.08)	(0.09)	(0.12)	(0.51)	(1.28)
Rape Conviction	0.06	−0.17	0.24	0.52**	0.59
	(0.15)	(0.22)	(0.25)	(0.18)	(1.09)
Murder Conviction	−0.04	−0.06	−0.18	−0.40	0.15
	(0.09)	(0.11)	(0.15)	(0.47)	(1.30)
Year of Conviction	0.03***	0.03***	0.02*	0.05**	0.04
	(0.01)	(0.01)	(0.01)	(0.02)	(0.14)
Panel 3: + State FE					
Fed. Lab	0.37	−0.02	.		
	(0.40)	(0.23)	.		
State Lab	0.02	−0.15	0.12		
	(0.09)	(0.13)	(0.25)		
Local Ind.	0.11	0.09	−0.57		
	(0.17)	(0.23)	(0.35)		
Private	0.66***	0.76***	0.44*		
	(0.12)	(0.18)	(0.23)		
Rape Conviction	0.14	0.05	0.32		
	(0.15)	(0.28)	(0.46)		
Murder Conviction	0.02	0.02	−0.02		
	(0.12)	(0.15)	(0.37)		
Year of Conviction	0.02***	0.03**	0.04		
	(0.01)	(0.01)	(0.03)		
Mean Validity Rate					
Sample Mean	0.48	0.41	0.53	0.58	0.33
n	193	100	60	12	6

Notes: