

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

2018

***Terry* Stops and Frisks: The Troubling Use of Common Sense in a World of Empirical Data**

David Rudovsky

University of Pennsylvania Carey Law School

David A. Harris

University of Pittsburgh

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Constitutional Law Commons](#), [Criminal Law Commons](#), [Fourth Amendment Commons](#), [Law and Race Commons](#), [Law and Society Commons](#), [Law Enforcement and Corrections Commons](#), and the [Public Law and Legal Theory Commons](#)

Repository Citation

Rudovsky, David and Harris, David A., "*Terry* Stops and Frisks: The Troubling Use of Common Sense in a World of Empirical Data" (2018). *Faculty Scholarship at Penn Law*. 1973.

https://scholarship.law.upenn.edu/faculty_scholarship/1973

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

Terry Stops and Frisks: The Troubling Use of Common Sense in a World of Empirical Data

DAVID RUDOVSKY* AND DAVID A. HARRIS†

The investigative detention doctrine first announced in Terry v. Ohio and amplified over the past fifty years has been much analyzed, praised, and criticized from a number of perspectives. Significantly, however, over this time period commentators have only occasionally questioned the Supreme Court’s “common sense” judgments regarding the factors sufficient to establish reasonable suspicion for stops and frisks. For years, the Court has provided no empirical basis for its judgments, due in large part to the lack of reliable data. Now, with the emergence of comprehensive data on these police practices, much can be learned about the predictive power of suspect conduct and other predicates for law enforcement interventions. And what has been learned calls into question a number of factors that have been cited in reasonable suspicion analyses over many years.

No observer of the legal system can fail to notice the growing role of data and empirical analysis in the courts. A disparate set of cases have turned in large part on rigorously analyzed data. Yet this trend has not taken root in an important set of cases involving the widely used practice of stop and frisk. When stop-and-frisk practices become the subject of litigation, courts generally either have no data to review or have failed to engage in empirical analysis of the data that are available and which could be used to test the claims of reasonable suspicion. Rather, the courts invoke the conventional wisdom that as a matter of common sense certain conduct, for example, furtive movement, flight, bulges in clothing, and suspect location, indicates criminal conduct.

We have no argument with common sense propositions; we have no aversion to clear, straightforward thinking. But what this phrase often reflects is a set of unexamined (even if widely held) assumptions. The proliferation of data on these basic questions provides the means for empirical analysis, and it is our argument that courts should use this newly available data in assessing reasonable suspicion factors in the

* Senior Fellow, University of Pennsylvania Law School; founding partner, Kairys, Rudovsky, Messing, Feinberg & Lin, LLP. Professor Rudovsky served as counsel in *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. filed Nov. 4, 2010), discussed *infra passim*.

† Sally Ann Semenko Endowed Chair and Professor of Law, University of Pittsburgh School of Law. Creator and host, “Criminal Injustice” podcast.

We thank David Abrams, Susan Lin, Paul Messing, Leonard Sosnov, and Seth Kreimer for their very useful comments and suggestions.

same manner that they have engaged in empirical judgments, using both big and targeted data, in other areas.

I.	INTRODUCTION	502
II.	THE <i>TERRY</i> DOCTRINE	506
III.	THE EMERGENCE OF BIG DATA AND EMPIRICISM	513
IV.	DATA, EMPIRICISM AND POLICE STOP-AND-FRISK PRACTICES	518
	A. <i>The Role of Empiricism and Data in the Courts</i>	518
	B. <i>Risk Assessment Tools</i>	524
	C. <i>Data, Empiricism, and Stops and Frisks</i>	527
	D. <i>Litigation and Studies</i>	531
	1. <i>New Jersey Turnpike and Racial Profiling</i>	531
	2. <i>New York City Stop-and-Frisk Practices</i>	532
	3. <i>Philadelphia Stop-and-Frisk Practices</i>	536
V.	PREDICTIVE ANALYTICS IN STOP-AND-FRISK PRACTICES	537
VI.	CONCLUSION.....	545

I. INTRODUCTION

When the United States Supreme Court decided *Terry v. Ohio* in 1968,¹ it created new constitutional rules to govern long-standing practices in American policing. The Fourth Amendment to the Constitution, fully applicable to the states since 1961,² required that officers have probable cause to suspect a person of involvement in a crime in order for police to conduct any search or seizure. Yet officers on the street had continued to focus on people who simply looked suspicious, perhaps for a concrete reason or just because of an officer’s gut feeling, and police had stopped these people, refused to let them go, and searched them—sometimes with less than probable cause, and sometimes with no basis whatsoever.³ For example, New York State permitted police to stop people and perform cursory searches, with less than probable cause but requiring some level of fact-based suspicion.⁴

¹ *Terry v. Ohio*, 392 U.S. 1 (1968).

² *Mapp v. Ohio*, 367 U.S. 643, 655 (1961).

³ *E.g.*, REMO FRANCESCHINI, *A MATTER OF HONOR* 35–36 (1993) (discussing the common practice of “giv[ing] him a toss”: stopping and detaining a person on a hunch and going through the person’s pockets and possessions in search of evidence or weapons).

⁴ N.Y. CODE CRIM. PROC. § 180-a, at issue in *Sibron v. New York*, 392 U.S. 40, 43–44 (1968), provides in full:

1. A police officer may stop any person abroad in a public place whom he reasonably suspects is committing, has committed or is about to commit a felony or any of the offenses specified in section five hundred fifty-two of this chapter, and may demand of him his name, address and an explanation of his actions.

Terry was the Supreme Court's opportunity to assess these practices. The Court would decide whether, as the defense in the case argued, all police-enforced stops, however temporary, and all searches, however cursory, required that police have probable cause,⁵ or whether a temporary stop and a frisk amounted to "a mere 'minor inconvenience and petty indignity'" which could be based on any suspicion by a police officer.⁶

The Court declined to adopt either of these opposing positions and limited its ruling to the question of the grounds required for a frisk of the suspect. A hunch or a gut feeling about possible criminal conduct would not suffice; an officer needed to have facts, and inferences drawn from them, that would raise a suspicion in the mind of a reasonable officer that the person under observation was armed and dangerous. If this standard was satisfied, the officer could perform a frisk—a limited pat down of the suspect's outer clothing—in order to avoid the danger posed by a weapon.⁷ Justice Harlan's concurrence emphasized that the use of stop and frisks constituted a two-step process, and likewise required a two-step analysis.⁸ The officer would need reasonable suspicion that crime was afoot, but this alone would not necessarily support a frisk.⁹ The frisk would require reasonable suspicion that the suspect was armed and dangerous.¹⁰ By the same token, the police could not justify a frisk unless reasonable suspicion first existed that would allow a temporary stop.¹¹

The *Terry* opinion told the story of how the defendant and two other men came to be observed, stopped, frisked, and then arrested in downtown Cleveland by Officer McFadden, a veteran police officer.¹² McFadden, having patrolled the downtown beat for decades, noticed two of the men walking up and down the same short stretch of a city block, pausing to stare into the same jewelry store window; the men did this, first one man, then the other, twenty-four times in all.¹³ At the end of each trip up and back, they conferred on a nearby corner, and were sometimes joined by the third man, who quickly walked away.¹⁴ Observing all of this, McFadden suspected that the men were "casing a job, a

2. When a police officer has stopped a person for questioning pursuant to this section and reasonably suspects that he is in danger of life or limb, he may search such person for a dangerous weapon. If the police officer finds such a weapon or any other thing the possession of which may constitute a crime, he may take and keep it until the completion of the questioning, at which time he shall either return it, if lawfully possessed, or arrest such person.

⁵ *Terry*, 392 U.S. at 11–12.

⁶ *Id.* at 10–11 (quoting *People v. Rivera*, 201 N.E.2d 32, 36 (N.Y. 1964)).

⁷ *Id.* at 19 n.16, 27.

⁸ *Id.* at 32–33 (Harlan, J., concurring).

⁹ *Id.* at 30 (majority opinion); *id.* at 33 (Harlan, J., concurring).

¹⁰ *Id.* at 30 (majority opinion); *id.* at 33 (Harlan, J., concurring).

¹¹ *Terry*, 392 U.S. at 32–33 (Harlan, J., concurring).

¹² *Id.* at 6–7 (majority opinion).

¹³ *Id.* at 6.

¹⁴ *Id.*

stick-up”—in other words, performing reconnaissance for a planned armed robbery of the store.¹⁵ McFadden stopped them, asked some questions, and then frisked suspect Terry—locating a gun.¹⁶ He found another gun on one of the other men as well, and arrested them both.¹⁷ This sort of police work, the Court said, constituted exactly the correct process: an officer, looking at the behavior of the suspects and interpreting what all of it meant, would have reasonable, fact-based suspicion to temporarily detain and question them, and then (because an armed robbery seemed to be the goal) to frisk them for weapons.¹⁸

By taking the facts as central to its conclusion that Officer McFadden had performed in exactly the way that would comply with the new *Terry* standard, the Court made clear an important concept: the reasonable suspicion determination turns in large part on the predictive value of a range of conduct and information made known to an officer. Based on this inference—in essence, making a prediction that the men, walking back and forth repeatedly with particular attention to a jewelry store’s display window, had plans for an immediate robbery of the store—McFadden may have intuited correctly.¹⁹ But a reviewing court operating under *Terry* need not necessarily find that the prediction was correct in order for the actions of the police to pass muster. Rather, the prediction, based on the observed facts and circumstances and the inferences McFadden could draw, need only have been reasonable.²⁰

Over the run of cases decided since *Terry*, the Court has put very little meat on the bones of the analysis necessary for determining the sufficiency of the predictive value in a particular cluster of facts to constitute reasonable suspicion. What substance the Court has supplied has come largely in the form of the ever-ready, ever-malleable phrase “common sense.”²¹ The Court instructs us to take a nontechnical, everyday lay person’s common sense view of the facts to ascertain whether the officer possessed the requisite reasonable suspicion necessary to perform a *Terry* stop and frisk. For example, in *United States v.*

¹⁵ *Id.*

¹⁶ *Id.* at 7.

¹⁷ *Terry*, 392 U.S. at 7.

¹⁸ *Id.* at 27–28.

¹⁹ The *Terry* Court acknowledged the possibility that racial bias could affect stop-and-frisk practices but apparently did not think the problem serious enough to restrict this practice. Some commentators addressed this issue and as later litigation and studies would show, see *infra* Part IV.D, racial bias is a serious matter. See Adina Schwartz, “Just Take Away Their Guns”: *The Hidden Racism of Terry v. Ohio*, 23 *FORDHAM URB. L.J.* 317, 323–24 (1996).

²⁰ *Terry*, 392 U.S. at 28 (“[Officer McFadden] had observed enough to make it quite reasonable to fear that [the suspects] were armed; and nothing in their response[s] . . . served to dispel that reasonable belief. We cannot say that [Officer McFadden’s] decision at that point to seize Terry and pat his clothing for weapons was the product of a volatile or inventive imagination . . .”).

²¹ *Infra* notes 32–34 and accompanying text.

Cortez,²² the Court stated a theme that echoes in many of its stop-and-frisk cases:

The process does not deal with hard certainties, but with probabilities. Long before the law of probabilities was articulated as such, practical people formulated certain common sense conclusions about human behavior; jurors as factfinders are permitted to do the same—and so are law enforcement officers. . . . [T]he evidence thus collected must be seen and weighed not in terms of library analysis by scholars, but as understood by those versed in the field of law enforcement.²³

The Court avoids empirical reasoning with respect to what meets the legal standard; rather, it looks at the question of whether the officer had reasonable suspicion through the experienced eyes of the police officer on patrol, with minimal consideration of whether these judgments are based on proven, reliable factors.²⁴ Further, the Court gives substantial deference to an officer's judgment based on the supposition that her experience provides unique expertise.²⁵

We do not write to oppose common sense or the usefulness of experience in decision-making. Neither of us have any aversion to clear, straightforward thinking. Rather, we believe that common sense as that concept informs the Court's criminal justice jurisprudence often means something different than what it sounds like: not obvious, homespun, solid judgment, but *unexamined assumptions* with which many people may agree (and which may be problematically reinforced by the fact that incriminating evidence was actually seized). However, wide-spread belief in the truth of what may seem obvious is not the same thing as actual proof. And where there is empirical data that can substantiate or call into question the predictive value of these "common sense" facts, it is our view that courts should adjust their perceptions accordingly.

In this Article, we will use empirical data to examine assumptions about the predictive value of observed facts in the context of the *Terry* stop-and-frisk jurisprudence. In a world where data and empiricism are becoming important metrics for determining the efficacy and reliability of a vast array of legal, scientific, and social science standards, including the use of data and algorithms in criminal justice decision-making, "common sense" resolutions may no longer be justified. Our project is informed by the growing literature examining the possibilities and implications of so-called "big data" and on the use of predictive analytics and algorithmic predictions for public safety and law enforcement.

We seek to answer the question of whether statistically driven analysis makes stop and frisk more efficient and fairer by paving the way to more productive stops, fewer encounters, and a reduction in racial and ethnic bias. In doing so, we make a modest claim: given the growing availability and use of

²² United States v. *Cortez*, 449 U.S. 411 (1981).

²³ *Id.* at 418; *see also* Illinois v. *Wardlow*, 528 U.S. 119, 123–25 (2000).

²⁴ *See Wardlow*, 528 U.S. at 123–25; *Cortez*, 449 U.S. at 418.

²⁵ *See Wardlow*, 528 U.S. at 123–25; *Cortez*, 449 U.S. at 418.

data in criminal justice, courts should consider data in deciding whether to accept the empirically testable claims by police and other criminal justice actors, instead of relying upon content-free, unsubstantiated assertions of “common sense.” And we conclude that the disturbingly low hit rates for weapons and other contraband in police frisks calls into question many of the court-approved justifications for this invasive police practice.

We proceed as follows. In Part II, we examine the *Terry* doctrine in greater depth, and in particular its concept of reasonableness under the Fourth Amendment. Part III discusses the emergence of “big data” and empiricism. Part IV brings us to what data analysis and empiricism have been able to tell us about stop-and-frisk practices, beginning with a discussion of how the availability of data and empirical judgments has grown in the law generally, with a focus on voting rights and the Supreme Court’s most recent case on abortion restrictions. Part V brings us to the heart of our argument: what data and empirical analysis can tell us about frisks, why courts must make use of data analysis instead of unexamined “common sense” assumptions, and what our analysis of predictive values and metrics means for the criminal law and for policing.

II. THE *TERRY* DOCTRINE

In *Terry v. Ohio* the United States Supreme Court ruled that a person who was stopped for investigation by the police could be frisked if the officer had reasonable suspicion that the person was “armed and dangerous.”²⁶ The Court had not previously authorized a seizure or search of a person on less than probable cause, but by focusing on the Fourth Amendment’s proscription of “unreasonable” searches or seizures, the Court determined that probable cause was not necessary given the limited intrusion into personal privacy posed by a stop and frisk.²⁷ While *Terry* did not directly involve the issue of the permissible grounds for a “stop,” the Court’s opinions made clear that it was approving forcible stops of persons based upon “reasonable suspicion” of involvement in serious criminal conduct.²⁸

The Supreme Court has expanded the *Terry* doctrine in several significant respects. First, the Court has permitted stops (and consequently possible frisks) of all persons who are reasonably suspected of *any* criminal activity, including possessory offenses, traffic violations, and “quality of life” summary offenses, thus significantly widening the scope of this practice.²⁹ These cases rejected arguments that expanding stop and frisk beyond investigations of violent and

²⁶ *Terry v. Ohio*, 392 U.S. 1, 27, 30 (1968).

²⁷ *Id.* at 21, 25, 29–31.

²⁸ *Id.* at 29–31; *see also id.* at 33 (Harlan, J., concurring); *id.* at 34–35 (White, J., concurring); *id.* at 37 (Douglas, J., dissenting). The Court expressly adopted this doctrine in *Adams v. Williams*, 407 U.S. 143, 146 (1972).

²⁹ *See, e.g.,* *Navarette v. California*, 134 S. Ct. 1683, 1687 (2014); *Arizona v. Johnson*, 555 U.S. 323, 327 (2009); *United States v. Arvizu*, 534 U.S. 266, 275–77 (2002); *Wardlow*, 528 U.S. at 124; *Adams*, 407 U.S. at 146–48.

other serious crimes would lead to arbitrary and excessive enforcement practices.³⁰ This development coincided with an era of proactive and order-maintenance policing that has generated broad debate regarding “broken windows,” “zero tolerance,” and “quality of life” policing methods.³¹ Designed to prevent serious crime and disorder by punishing even minor deviances, these controversial practices have become a mainstay of modern policing.

³⁰ Judge Friendly argued against applying stop and frisk in nonviolent crime investigations. See *Williams v. Adams*, 436 F.2d 30, 38–39 (2d Cir. 1970) (Friendly, J., dissenting), *rev'd*, 441 F.2d 394 (2d Cir. 1971) (en banc), *rev'd*, 407 U.S. 143 (1972); see also *Terry*, 392 U.S. at 32 (Harlan, J., concurring); Richard E. Myers II, *Challenges to Terry for the Twenty-First Century*, 81 *MISS. L.J.* 937, 943 (2012); Carol S. Steiker, *Terry Unbound*, 82 *MISS. L.J.* 329, 335 (2013). Data from jurisdictions that document investigatory stops show that the majority of these stops (and frisks) are for nonviolent offenses. See *Floyd v. City of New York*, 959 F. Supp. 2d 540, 623 (S.D.N.Y. 2013); Plaintiffs’ Seventh Report to Court and Monitor on Stop and Frisk Practices: Fourth Amendment Issues at 19, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. May 2, 2017) [hereinafter Plaintiffs’ Seventh Report to Court]; Plaintiffs’ Sixth Report to Court and Monitor on Stop and Frisk Practices: Fourth Amendment Issues at 19, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. Mar. 22, 2016) [hereinafter Plaintiffs’ Sixth Report to Court]; Bernard E. Harcourt & Tracey L. Meares, *Randomization and the Fourth Amendment*, 78 *U. CHI. L. REV.* 809, 819 (2011) (only 20% of arrests following stops in New York City were for felony investigations).

³¹ There is a vast literature on these issues, much of which is responsive to the early writings of James Q. Wilson. See JAMES Q. WILSON, *VARIETIES OF POLICE BEHAVIOR* 39–40 (1968); George L. Kelling & James Q. Wilson, *Broken Windows*, ATLANTIC (Mar. 1982), <https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/> [<https://perma.cc/SZ8B-VFB4>]. For a sample of responses that discuss stop-and-frisk practices in the context of proactive policing, see BERNARD E. HARCOURT, *ILLUSION OF ORDER* (2001); FRANKLIN E. ZIMRING, *THE CITY THAT BECAME SAFE: NEW YORK’S LESSONS FOR URBAN CRIME AND ITS CONTROL* (2012); Jeffrey Bellin, *The Inverse Relationship Between the Constitutionality and Effectiveness of New York City “Stop and Frisk,”* 94 *B.U. L. REV.* 1495, 1504–05 (2014); John E. Eck & Edward R. Maguire, *Have Changes in Policing Reduced Violent Crime? An Assessment of the Evidence*, in *THE CRIME DROP IN AMERICA* 207, 224–25, 228–45 (Alfred Blumstein & Joel Wallman eds., 2000); David F. Greenberg, *Studying New York City’s Crime Decline: Methodological Issues*, 31 *JUST. Q.* 154, 155 (2014); Rachel A. Harmon & Andrew Manns, *Proactive Policing and the Legacy of Terry*, 15 *OHIO ST. J. CRIM. L.* 49 (2017); Andrew Ingram, *Breaking Laws To Fix Broken Windows: A Revisionist Take on Order Maintenance Policing*, 19 *BERKELEY J. CRIM. L.* 112, 132–33 (2014); Steven D. Levitt, *Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not*, 18 *J. ECON. PERSP.* 163, 172–73 (2004); Tracey L. Meares & Bernard E. Harcourt, *Foreword: Transparent Adjudication and Social Science Research in Constitutional Criminal Procedure*, 90 *J. CRIM. L. & CRIMINOLOGY* 733, 779 (2000); Tracey L. Meares, *The Law and Social Science of Stop and Frisk*, 10 *ANN. REV. L. & SOC. SCI.* 335, 339–43 (2014); David Rudovsky, *Debate: Opening Statement, Stop-and Frisk: The Power of Data and the Decision in Floyd v. City of New York*, 162 *U. PA. L. REV. ONLINE* 117, 118 (2013); David Weisburd et al., *Could Innovations in Policing Have Contributed to the New York City Crime Drop Even in a Period of Declining Police Strength?: The Case of Stop, Question and Frisk as a Hot Spots Policing Strategy*, 31 *JUST. Q.* 129, 131–32 (2014); and John MacDonald et al., *The Effects of Local Police Surges on Crime and Arrests in New York City*, Article in *PLoS One*, PLOS 2 (2016), <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0157223&type=printable> [<https://perma.cc/RXT5-4TEG>].

Second, the Court has approved stops and frisks on the assumption that certain conduct is predictive of or associated with criminal behavior and weapon possession, but has not required an empirical basis for these judgments, many of which rest on conduct that is entirely consistent with innocence.³² The Court has repeatedly stated that it would apply “common sense” judgments and permit officers to make reasonable inferences from a suspect’s behavior in determining whether there were legal grounds for a stop and frisk.³³ Thus, in *Illinois v. Wardlow*, the Court permitted a stop and frisk where the suspect fled from police in a high-crime area even though it had no data showing that persons who fled from the police in high-crime areas were reasonably likely to be involved in criminal conduct.³⁴ And while the Court has cautioned against use of characteristics that would permit frequent stops of a “very large category of presumably innocent travelers,”³⁵ its cases over the years have provided police with vast stop-and-frisk discretion, permitting stops and frisks on vague and subjective standards such as nervousness, officer experience, furtive movements and suspicious activity.³⁶ This approach is similar to that followed by some

³² See, e.g., *United States v. Arvizu*, 534 U.S. 266, 277 (2002) (finding that conduct that appears to be innocent may still provide reasonable suspicion based on the “totality of the circumstances”).

³³ See, e.g., *id.* (discussing stop of car for immigration investigation based on location and reactions of occupants to police presence); *United States v. Cortez*, 449 U.S. 411, 418 (1981) (“[C]ommon sense conclusions about human behavior . . . weighed not in terms of library analysis by scholars, but as understood by . . . law enforcement.”).

³⁴ *Wardlow*, 528 U.S. at 124–25. As we discuss, *infra* note 64 and accompanying text, there was data that contradicted the Court’s analysis. See Meares & Harcourt, *supra* note 31, at 790 (citing New York State Attorney General’s Report on stop-and-frisk practice in New York City where only one in forty-five stops made on flight in a high-crime area resulted in an arrest). Moreover, as Justice Stevens stated in his dissent, innocent residents of minority neighborhoods might have good reason to avoid contacts with the police. *Wardlow*, 528 U.S. at 132–33 (Stevens, J., dissenting in part and concurring in part).

³⁵ *Reid v. Georgia*, 448 U.S. 438, 441 (1980) (per curiam).

³⁶ See, e.g., *Arvizu*, 534 U.S. at 277; *Florida v. Royer*, 460 U.S. 491, 494, 502 (1983) (plurality opinion) (noting nervousness as a factor contributing to a finding of reasonable suspicion); *United States v. Briggs*, 720 F.3d 1281, 1286, 1292 (10th Cir. 2013) (finding suspect’s evasive or erratic movements are part of the totality of the circumstances); *United States v. Logan*, 526 F. App’x 498, 503 (6th Cir. 2013) (presence in a high-crime area); *United States v. See*, 574 F.3d 309, 314 (6th Cir. 2009) (sitting in a car for an extended period may add to reasonable suspicion); *United States v. Himmelwright*, 406 F. Supp. 889, 892–93 (S.D. Fla. 1975) (finding suspect’s “unusually calm” demeanor supported the finding of reasonable suspicion), *aff’d*, 551 F.2d 991 (5th Cir. 1977); Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461, 505 (2015). On the question of the relevancy and characteristics of a “high-crime area,” see *United States v. Montero-Camargo*, 208 F.3d 1122, 1143 (9th Cir. 2000) (en banc) (Kozinski, J., concurring) (“Just as a man with a hammer sees every problem as a nail, so a man with a badge may see every corner of his beat as a high crime area.”), and Andrew Guthrie Ferguson, *Crime Mapping and the Fourth Amendment: Redrawing “High-Crime Areas,”* 63 HASTINGS L.J. 179, 181 (2011); see also DAVID COLE, NO EQUAL JUSTICE 47–52 (1999) (surveying airport drug courier profile stops that were based on virtually all characteristics of passenger travel).

courts in assessing airport “drug courier profiles.” As recounted by Professor Barry Friedman, the Drug Enforcement Agency and other law enforcement organizations used passenger behavior—often entirely innocent and often inconsistent—in their interdiction efforts. Thus, people were considered couriers if they “arrived at night [or] early in the morning,” were “[o]ne of the first to deplane[, o]ne of the last to deplane[, or d]eplaned in the middle,” “[u]sed a one-way ticket [or] a round-trip ticket,” “[t]raveled alone [or] with a companion,” “[a]cted too nervous [or] too calm,” etc.³⁷ Not surprisingly, when the full data set, as opposed to the very occasional successful stop, was examined, there was “less than a 2% hit rate” for persons carrying drugs.³⁸

Use of vague and indeterminate standards has been rejected in other contexts. The Court has ruled that vague statutory language such as “vagrancy,” “loitering,” and “suspicious conduct” is not a valid ground for arrest and prosecution under the void-for-vagueness doctrine.³⁹ More recently, the Court struck down as too indeterminate the “residual clause” of the Armed Career Criminal Act that punished unenumerated crimes by their likeness to specific criminal acts.⁴⁰ The Court has also extended the vagueness doctrine to other contexts, including standards for imposing costs on criminal defendants,⁴¹ rules regulating the bar,⁴² and agency regulations.⁴³

Yet, the Court has failed to explain why the maxim that “[n]o one should have to ponder the totality of the circumstances in order to determine whether

³⁷ BARRY FRIEDMAN, UNWARRANTED: POLICING WITHOUT PERMISSION 152–53 (2017). The Supreme Court did not validate drug courier profiles as grounds for a stop, but did permit the various factors to be part of the overall determination of the “evidentiary significance as seen by a trained agent.” *United States v. Sokolow*, 490 U.S. 1, 10 (1989).

³⁸ FRIEDMAN, *supra* note 37, at 153.

³⁹ *See* *Kolender v. Lawson*, 461 U.S. 352, 357, 361 (1983) (finding requirement that detained person provide “credible and reliable” identification is void for vagueness because law must provide “ordinary people” with notice to “understand what conduct is prohibited” so as to avoid “arbitrary and discriminatory enforcement”); *Papachristou v. City of Jacksonville*, 405 U.S. 156, 162 (1972) (holding vagrancy ordinance void for vagueness); *Shuttlesworth v. City of Birmingham*, 382 U.S. 87, 90 (1965) (holding loitering prohibition void for vagueness); *Winters v. New York*, 333 U.S. 507, 540 (1948) (Frankfurter, J., dissenting) (explaining definiteness is needed to prevent “the net to be cast at large, to enable men to be caught who are vaguely undesirable in the eyes of police and prosecution, although not chargeable with any particular offense”); *Desertrain v. City of Los Angeles*, 754 F.3d 1147, 1155–56 (9th Cir. 2014) (finding law prohibiting use of cars as “living quarters” void for vagueness); *Stahl v. City of St. Louis*, 687 F.3d 1038, 1040–41 (8th Cir. 2012) (finding conduct that “impede[s] traffic” is too vague a standard); Note, *The Void-for-Vagueness Doctrine in the Supreme Court*, 109 U. PA. L. REV. 67, 73–75 (1960).

⁴⁰ *Johnson v. United States*, 135 S. Ct. 2551, 2557–58 (2015); *cf.* *Beckles v. United States*, 137 S. Ct. 886, 894–95 (2017) (rejecting similar challenge to sentencing guidelines).

⁴¹ *Giaccio v. Pennsylvania*, 382 U.S. 399, 403 (1966).

⁴² *Gentile v. State Bar of Nevada*, 501 U.S. 1030, 1048–51 (1991).

⁴³ *FCC v. Fox Television Stations, Inc.*, 132 S. Ct. 2307, 2317–20 (2012).

his conduct is a felony,”⁴⁴ which precludes prosecutions based on vague statutes, does not apply to forcible stops and frisks of individuals. Indeed, there is good reason to believe that the investigative-detention rationale of *Terry* has become a substitute for vague loitering and vagrancy laws and has provided police with an even broader policing power.⁴⁵ The vagueness of the factors that may justify stops and frisks increases the risk that police will act on biases in deciding whether there is sufficient suspicion for forcible intervention. Thus, if police associate racial minorities with criminal conduct where the same actions by whites are not regarded as suspicious, or if they consider race or ethnicity as surrogates for criminal conduct, even on a subconscious level, bias may become a causative factor.⁴⁶ For this reason alone, courts should scrutinize the reasons provided for stops and frisks to ensure that they have predictive value on the issue of “reasonable suspicion,” and where empirical evidence bears on this question, should use that evidence in the constitutional calculus.

Third, stop-and-frisk doctrine has developed primarily as a function of the exclusionary rule, which means that the cases in which courts adjudicate the issue are those in which contraband or other evidence of criminal conduct has already been uncovered in the investigative detention. In this context, fact-finding and legal analysis can be compromised by false or misleading police officer testimony and the court’s reluctance to suppress otherwise probative evidence.⁴⁷ Indeed, where evidence of a crime is uncovered in a stop-and-frisk

⁴⁴ *Bond v. United States*, 134 S. Ct. 2077, 2097 (2014) (Scalia, J., concurring); *see also* *McCullen v. Coakley*, 134 S. Ct. 2518, 2543 (2014) (Scalia, J., concurring in the judgment) (asserting that abortion protesters should not be subject to seizure based on an offense of “follow[ing] and harass[ing]” (alterations in original)).

⁴⁵ *See* Tracey Meares, *This Land Is My Land?*, 130 HARV. L. REV. 1877, 1880 (2017) (book review).

⁴⁶ There is a growing body of literature, judicial rulings, and studies regarding the concept of implicit bias. *See, e.g.*, *United States v. Mateo-Medina*, 845 F.3d 546, 553 (3d Cir. 2017) (using arrest records for sentencing impermissible); REBECCA C. HETLEY ET AL., STANFORD SPARQ, DATA FOR CHANGE 33 (2016); Geoffrey P. Alpert et al., *Police Suspicion and Discretionary Decision Making During Citizen Stops*, 43 CRIMINOLOGY 407, 415–17 (2005); Samuel R. Bagenstos, *Implicit Bias’s Failure*, BERKELEY J. EMP. & LAB. L. (forthcoming 2018) (manuscript at 4–5); Jennifer L. Eberhardt et al., *Seeing Black: Race, Crime, and Visual Processing*, 87 J. PERSONALITY & SOC. PSYCHOL. 876, 889–91 (2004); Kent Greenawalt, *Commentary, Probabilities, Perceptions, Consequences and “Discrimination”*: One Puzzle About Controversial “Stop and Frisk,” 12 OHIO ST. J. CRIM. L. 181, 182–84 (2014); Justin D. Levinson & Robert J. Smith, *Systemic Implicit Bias*, 126 YALE L.J.F. 406, 415 (2017); Myers, *supra* note 30, at 954–55; L. Song Richardson, *Arrest Efficiency and the Fourth Amendment*, 95 MINN. L. REV. 2035, 2037–39 (2011) [hereinafter Richardson, *Arrest Efficiency*]; L. Song Richardson, *Implicit Racial Bias and Racial Anxiety: Implications for Stops and Frisks*, 15 OHIO ST. J. CRIM. L. 73, 75–88 (2017); Rob Voigt et al., *Language from Police Body Camera Footage Shows Racial Disparities in Officer Respect*, 114 PNAS 6521, 6524 (2017).

⁴⁷ *See, e.g.*, Bernard Chao et al., *Why Courts Fail To Protect Privacy: Race, Age, Bias, and Technology*, 106 CALIF. L. REV. 263, 279–86 (2018) (discussing judges’ reluctance to suppress incriminating evidence because of hindsight bias); David N. Dorfman, *Proving the*

encounter, the weapon or drugs or other evidence provides a kind of “self-validating” reason for an officer’s intervention.⁴⁸

Similar criticisms can be made of the Court’s “probable cause” jurisprudence, as the Fourth Amendment probable cause standard also involves inferential judgments from facts that may be innocent in nature. Probable cause requires a showing of a “fair probability” that the suspect has committed a crime or that evidence will be found at a certain location.⁴⁹ The courts have not quantified the probability standard, but judges and commentators have suggested a 35–50% probability of criminal conduct.⁵⁰ Reasonable suspicion is satisfied by “considerably less” proof of wrongdoing than a preponderance of the evidence,⁵¹ and some commentators have suggested a significantly lower range.⁵²

The differences in assessing probabilities, however, are more than one of mere degree. Probable cause analysis is usually anchored to a specific, completed criminal act and the police investigation is based on information provided by witnesses, personal observations by the police, forensics, and culpable behavior of the suspect.⁵³ Accordingly, in most cases there are

Lie: Litigating Police Credibility, 26 AM. J. CRIM. L. 455, 458–65 (1999) (discussing both “exclusion of illegally obtained evidence” and false police testimony); Stanley Z. Fisher, “*Just the Facts Ma’am*”: *Lying and the Omission of Exculpatory Evidence in Police Reports*, 28 NEW ENG. L. REV. 1, 9–12 (1993); Laurie L. Levenson, *Unnerving the Judges: Judicial Responsibility for the Rampart Scandal*, 34 LOY. L.A. L. REV. 787, 790–94 (2001) (discussing police perjury); Myers, *supra* note 30, at 965–66; Christopher Slobogin, *Testilying: Police Perjury and What To Do About It*, 67 U. COLO. L. REV. 1037, 1041 (1996) (discussing the development of Fourth Amendment jurisprudence through the exclusionary rule).

⁴⁸ Occasionally, courts expressly recognize the pressures to validate searches that uncover evidence, regardless of the Fourth Amendment limitations. *See, e.g.*, *United States v. Hussain*, 835 F.3d 307, 309 (2d Cir. 2016) (“Appeals based on the Fourth Amendment from denied motions to suppress evidence of illegal weapons or contraband (drugs, etc.) are often difficult because the Government is in a sense proven right. Whatever prompted the search (a hunch, suspicion, luck, reasonable belief, or probable cause), incriminating evidence was found.”).

⁴⁹ *Illinois v. Gates*, 462 U.S. 213, 238–39 (1983).

⁵⁰ *See City of Indianapolis v. Edmond*, 531 U.S. 32, 34–35, 47–48 (2000) (finding invalid under Fourth Amendment, highway drug checkpoint that resulted in a 7.4% drugs-related arrest rate, and 8.96% overall arrest rate); *Texas v. Brown*, 460 U.S. 730, 742 (1983) (holding probable cause does not require proof by a preponderance of the evidence); Bambauer, *supra* note 36, at 468; Orin Kerr, *Why Courts Should Not Quantify Probable Cause*, in *THE POLITICAL HEART OF CRIMINAL PROCEDURE* 131, 131–32, 132 n.6 (Michael Klarman et al. eds., 2012); Kit Kinports, *The Dog Days of Fourth Amendment Jurisprudence*, 108 NW. U. L. REV. COLLOQUY 64, 68 (2013); Max Minzner, *Putting Probability Back into Probable Cause*, 87 TEX. L. REV. 913, 925 (2009); Jeffrey J. Rachlinski et al., *Probable Cause, Probability, and Hindsight*, 8 J. EMPIRICAL LEGAL STUD. 72, 86–88 (2011); Richardson, *Arrest Efficiency*, *supra* note 46, at 2075–77.

⁵¹ *United States v. Sokolow*, 490 U.S. 1, 7 (1989).

⁵² Greenawalt, *supra* note 46, at 187; Rudovsky, *supra* note 31, at 119.

⁵³ *Gates*, 462 U.S. at 241–46.

relatively clear factors on which a court can determine the “fair probability” of guilt standard. If a robbery has been committed, the description of the suspect, method of flight, and other distinguishing factors can be assessed. Similarly, where police are searching for evidence, the reliability of a witness or informant, the personal observations of the officer, as well as other information related to the specific crime or evidence will determine whether probable cause exists.⁵⁴

Some stops are based on similar, if less detailed factors (e.g., report of recent robbery with description of suspect), but more often the reasonable suspicion analysis rests solely on observed behavior of the suspect.⁵⁵ The discretion given to the police in assessing what is “suspicious” conduct is far more malleable than the discretion exercised in the probable cause context. It is not surprising, therefore, that the hit rates for stops and frisks (where police actually seize weapons or contraband, effectuate an arrest, or issue a summons) are substantially lower than for arrests or searches conducted on full probable cause.⁵⁶

To date, courts have rarely employed an “evidence based” approach to police stop, frisk, or search practices.⁵⁷ But with newly emerging

⁵⁴ *Id.*

⁵⁵ See, e.g., *Illinois v. Wardlow*, 528 U.S. 119, 124 (2000); *Pennsylvania v. Mimms*, 434 U.S. 106, 107 (1977) (per curiam); *infra* note 255 and accompanying text.

⁵⁶ See *infra* note 89 and accompanying text. Indeed, where police have engaged in random stops without any particularized suspicion, for example at highway checkpoints, the hit rates have in some cases been higher than hit rates for stops and frisks where police are supposedly acting on objective evidence of suspicious conduct. *City of Indianapolis v. Edmond*, 531 U.S. 32, 35 (2000) (noting drug checkpoint yielded drug seizures in 4.7% of stops); see also Jeffrey Fagan, *Terry’s Original Sin*, 2016 U. CHI. LEGAL F. 43, 54–55 (2016); Minzner, *supra* note 50, at 925.

⁵⁷ See Bambauer, *supra* note 36, at 501–02; Chao et al., *supra* note 47, at 276–79; Lee Epstein et al., *Foreword: Testing the Constitution*, 90 N.Y.U. L. REV. 1001, 1017–18, 1023–24 (2015); David A. Harris, *Across the Hudson: Taking the Stop and Frisk Debate Beyond New York City*, 16 N.Y.U. J. LEGIS. & PUB. POL’Y 853, 861–62 (2013). Scholarly analysis and critiques of the *Terry* doctrine that have been published without reference to this data have ranged from high accolades, e.g., John Q. Barrett, *Deciding the Stop and Frisk Cases: A Look Inside the Supreme Court’s Conference*, 72 ST. JOHN’S L. REV. 749, 839 (1998) (“Many thus think of *Terry* . . . as a . . . sensible balancing of public interests in law enforcement against relatively lesser intrusions on personal freedom . . .”); I. Bennett Capers, *Rethinking the Fourth Amendment: Race, Citizenship, and the Equality Principle*, 46 HARV. C.R.-C.L. L. REV. 1, 30–35 (2011); Frank Rudy Cooper, *The Un-Balanced Fourth Amendment: A Cultural Study of the Drug War, Racial Profiling, and Arvizu*, 47 VILL. L. REV. 851, 884–89 (2002); Richardson, *Arrest Efficiency*, *supra* note 46, at 2059–72; Lawrence Rosenthal, *Pragmatism, Originalism, Race, and the Case Against Terry v. Ohio*, 43 TEX. TECH L. REV. 299, 325–30 (2010); and Stephen A. Saltzburg, *Terry v. Ohio: A Practically Perfect Doctrine*, 72 ST. JOHN’S L. REV. 911, 952–62 (1998), to strong criticism, e.g., Paul Butler, *Stop and Frisk and Torture-Lite: Police Terror of Minority Communities*, 12 OHIO ST. J. CRIM. L. 57, 59–66 (2014); Tracey Maclin, *When the Cure for the Fourth Amendment Is Worse than the Disease*, 68 S. CAL. L. REV. 1, 35–36 (1994); Scott E. Sundby, *An Ode to Probable Cause: A Brief Response to Professors Amar and Slobogin*, 72 ST. JOHN’S L. REV. 1133, 1134 (1998).

comprehensive and reliable data sets, there is an opening to new perspectives on Fourth Amendment issues, including the scope of the programs, analysis of compliance performance, and “hit rates,” all of which are critical to the development of reliable standards for balancing public safety with privacy and personal dignity.⁵⁸

III. THE EMERGENCE OF BIG DATA AND EMPIRICISM

Over the last decade, a new phrase has emerged: big data. The idea is that there now exists an almost inexhaustible stream of data concerning a wide array of human activities—business and commercial, financial, scientific, medical, industrial, governmental, and social. Using powerful analytical computing, we can examine these vast troves of data to discover patterns that might otherwise remain hidden.

The criminal justice system is capable of generating substantial data, from the earliest phases of an investigation and proactive policing tactics through final adjudications, sentencing, and parole and probation. With these data, scholars from multiple disciplines have begun to examine the potential, the pitfalls, and the contradictions to existing practice that analysis provides. More specifically, data collection on police stops of pedestrians over the past fifteen to twenty years has led to research and studies on the relationship of the data to doctrinal issues in stop-and-frisk practices.⁵⁹

Even before “big data” became a central focus point, Professors Tracey Meares and Bernard Harcourt addressed the problem of “common sense” judgments and the predictive value of human behavior.⁶⁰ In particular, they were critical of Supreme Court assertions of fact, rarely supported by data analysis—they called them “pseudo-empirical statements”⁶¹—about consequential and case determinative issues. Rather, “these are . . . purely rhetorical statements intended to render authoritative the Court’s decisions.”⁶² As an example, in *Illinois v. Wardlow*,⁶³ Chief Justice Rehnquist conceded the empirical nature of the question of whether flight from a police officer raises reasonable suspicion of wrongdoing, given the many other legitimate reasons to flee from an officer but stated that in the absence of “empirical studies dealing with inferences

⁵⁸ For example, in New York City, from 2004 to 2012, the police made 4.4 million pedestrian stops. *Floyd v. City of New York*, 959 F. Supp. 2d 540, 573 (S.D.N.Y. 2013). In Philadelphia, there were approximately 253,000 stops in 2009, Plaintiffs’ Third Report to Court and Monitor on Stop and Frisk Practices at 4, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. Mar. 19, 2013) [hereinafter Plaintiffs’ Third Report to Court], 204,000 in 2011, Plaintiffs’ Second Report to Court and Monitor on Stop and Frisk Practices at 13, *Bailey*, No. 10-5952 (July 2012), and 215,000 in 2012, Plaintiffs’ Third Report to Court, *supra*, at 4.

⁵⁹ See *infra* Part IV.D.

⁶⁰ Meares & Harcourt, *supra* note 31, at 784.

⁶¹ *Id.* at 739.

⁶² *Id.* at 740.

⁶³ *Illinois v. Wardlow*, 528 U.S. 119 (2000).

drawn from suspicious behavior,” the Court need do no more than look to its own “commonsense judgements and inferences about human behavior.”⁶⁴

Meares and Harcourt argued that courts cannot justify an approach that ignores “contested empirical claims that are hotly debated in legal and social scientific circles”⁶⁵ and for which data and studies exist that could help to resolve the debates. They proposed “a mode of judicial decision-making and academic debate that treats social scientific and empirical assessment as a crucial element in constitutional decision-making”⁶⁶

With the advent of big data, even more attention is being paid to potential uses (and abuses) of data mining and empirical determinations. For example, Professor Ric Simmons has expressed faith in the ability of “predictive algorithms” to “revolutionize” the criminal justice system.⁶⁷ Algorithms will do this by, among other things, making “far more accurate determinations of reasonable suspicion and probable cause.”⁶⁸ This will increase both efficiency—police will spend time and energy much more selectively, on a more promising set of targets—and also fairness, because “fewer innocent people will be stopped and searched.”⁶⁹

In the same vein, Professor Andrew Guthrie Ferguson, one of the most prolific scholars of big data and criminal justice, asks: What if big data analytics could replace the typical “small data”—that is, observable actions and activities of individuals—upon which police now rely to decide whether they believe reasonable suspicion of criminal activity exists?⁷⁰ Ferguson promotes the use of networked data sources, as this will increase knowledge and will allow for more accurate and targeted judgments of whether police have reasonable suspicion.⁷¹ Now, “[w]ith little effort,” Ferguson says, “officers can now identify most unknown suspects, not through their observations, but by accessing a web of information containing extensive personal data about suspects.”⁷²

⁶⁴ *Id.* at 124–25. As Meares and Harcourt noted, there was empirical data that called into question the inferential judgments that the Court made about flight in a high-crime area. Meares & Harcourt, *supra* note 31, at 790.

⁶⁵ Meares & Harcourt, *supra* note 31, at 750–51.

⁶⁶ *Id.* at 735.

⁶⁷ Ric Simmons, *Quantifying Criminal Procedure: How To Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947, 1016 (2016).

⁶⁸ *Id.* at 947.

⁶⁹ *Id.*

⁷⁰ Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 350–51 (2015) [hereinafter Ferguson, *Big Data*]; see also ANDREW GUTHRIE FERGUSON, *THE RISE OF BIG DATA POLICING 2* (2017).

⁷¹ *Id.* at 360–65.

⁷² Ferguson, *Big Data*, *supra* note 70, at 329–30. In a similar vein, Professor Andrew Manuel Crespo argues that data analytics can improve the criminal justice process in a more big-picture way: by bringing broad institutional contours into focus. Andrew Manuel Crespo, *Systemic Facts: Toward Institutional Awareness in Criminal Courts*, 129 HARV. L. REV. 2049, 2053 (2016). As currently structured, the system of constitutional criminal procedure, in particular its one-off, one-case-at-a-time nature, impedes any attempt to understand the

Both Simmons and Ferguson present uses of big data that could make the system fairer, but both recognize troubling issues. Simmons prudently warns of “obstacles” to overcome, first and foremost attaining the assurance that neither the algorithms nor the data rely on “improper factors, such as the race of the suspect.”⁷³ He also cautions that predictive algorithms must still take individualized suspicion into account, and that the legal standard must be quantified—something the Supreme Court has resisted.⁷⁴ Ferguson’s proposal for making known an unknown suspect relies on an ability to identify the person, for example through biometric identification like facial recognition.⁷⁵ But sufficiently accurate facial recognition is a ways off, and other types of biometric identification, for example fingerprint or retina scans, which seem closest to field deployment, would require a seizure of the individual, and that would require reasonable suspicion.

In other works on the subject,⁷⁶ Ferguson recognizes that big data and its predictions necessitate a careful understanding of the technology, because some flaws, while invisible, are baked into the finished product.⁷⁷ In other words, if the data reflect actions that have their bases in biased policing practices, the data will carry this forward, in the guise of clean, neutral metrics. While big-data-based predictive policing “has been promoted as a data-driven, race neutral, objective solution to the failed policing policies of the past,”⁷⁸ Ferguson says, it deserves a sustained examination and critique that scrutinizes the collection of

system or to regulate the actors in it from a systemic point of view. The nature of the system “den[ies] courts an opportunity to ‘see’ the systemic features of law enforcement behavior,” creating a “mismatch” between courts’ “institutional task and their institutional capacity.” *Id.* at 2050. Other commentators have urged similar use of data relevant to particular aspects of the criminal justice system. *See, e.g.*, Chao et al. *supra* note 47, at 318; Mary D. Fan, *Panopticism for Police: Structural Reform Bargaining and Police Regulation by Data-Driven Surveillance*, 87 WASH. L. REV. 93, 125–30 (2012); Pamela Metzger & Andrew Guthrie Ferguson, *Defending Data*, 88 S. CAL. L. REV. 1057, 1061–62 (2015).

⁷³ Simmons, *supra* note 67, at 947. Other commentators have made similar arguments regarding the integrity of algorithms. *See, e.g.*, Jane Bambauer, *Other People’s Papers*, 94 TEX. L. REV. 205, 253–57, 253 n.247 (2015); Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 805–06 (2014); *see also* Rebecca Wexler, Opinion, *When a Computer Program Keeps You in Jail*, N.Y. TIMES (June 13, 2017), <https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html> (on file with *Ohio State Law Journal*) (discussing problem of assertions of confidentiality of proprietary information that block challenges to probabilistic algorithms used in sentencing and parole decisions).

⁷⁴ Simmons, *supra* note 67, at 987.

⁷⁵ Ferguson, *Big Data*, *supra* note 70, at 330.

⁷⁶ Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109 (2017) [hereinafter Ferguson, *Policing Predictive Policing*]; Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259 (2012) [hereinafter Ferguson, *Predictive Policing and Reasonable Suspicion*].

⁷⁷ Ferguson, *Predictive Policing and Reasonable Suspicion*, *supra* note 76, at 314.

⁷⁸ Andrew Guthrie Ferguson, *Policing Predictive Policing: Abstract*, SSRN (Apr. 18, 2016), <https://ssrn.com/abstract=2765525> [<https://perma.cc/4Q6N-SXUU>].

data, the methodology used, transparency, accountability, and implementation, at the very least.⁷⁹ Moreover, even with full data sets, there can be serious disputes over analysis and predictive accuracy. For example, in a 2003 opinion upholding an Alaska sex offender registration law, Justice Kennedy relied on a Department of Justice (DOJ) study to assert that the risk of re-offending among sex offenders is “frightening and high.”⁸⁰ But it turns out that this assertion, which has been repeated in lower court cases since that decision, was based on single source in a compendium of papers reviewed by the DOJ that claimed up to an 80% re-offending rate.⁸¹ Recent research shows that this figure, taken from an article in *Psychology Today*, is highly dubious and most likely false; to the contrary, sex offenders re-offend at much lower rates.⁸²

But whatever the merits and limits of big data, there is good reason to examine even more limited data sets reflecting specific police practices. In a series of articles, Professor Sharad Goel and various colleagues employ data analysis to suggest ways in which stop-and-frisk practices can be fairer, rarer, and more productive.⁸³ Using data from New York City’s Stop, Question and Frisk program, these scholars show how “a particular kind of calculation made possible by modern, large-scale datasets—determining the likelihood that stopping and frisking a particular pedestrian will result in the discovery of contraband . . . could be used to reduce the racially disparate impact of pedestrian searches and to increase their effectiveness.”⁸⁴ Thus, the data show

⁷⁹ See Ferguson, *Policing Predictive Policing*, *supra* note 76, at 1144–87; see also Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 878–79 (2016) (finding that automated suspicion algorithms do not fit well into existing Fourth Amendment doctrine, so accuracy and effectiveness must be assured through “extrajudicial means”).

⁸⁰ *Smith v. Doe*, 538 U.S. 84, 103 (2003) (quoting *McKune v. Lile*, 536 U.S. 24, 34 (2002)).

⁸¹ *Id.* See generally RAYMOND C. BROWN & JOHN E. MOORE, U.S. DEP’T OF JUSTICE, A PRACTITIONER’S GUIDE TO TREATING THE INCARCERATED MALE SEX OFFENDER xiii (1988) (“[T]he recidivism rate of untreated [sex] offenders is around 80%.”).

⁸² Ira Mark Ellman & Tara Ellman, “Frightening and High”: *The Supreme Court’s Crucial Mistake About Sex Crime Statistics*, 30 CONST. COMMENT. 495, 495–508 (2015); Melissa Hamilton, *Briefing the Supreme Court: Promoting Science or Myth?*, 67 EMORY L.J. ONLINE 2021, 2022–23, 2022 n.5 (2017); Melissa Hamilton, Essay, *Constitutional Law and the Role of Scientific Evidence: The Transformative Potential of Doe v. Snyder*, 58 B.C. L. REV. E. SUPP. 34, 39 (2017); Adam Liptak, *Did the Supreme Court Base a Ruling on a Myth?*, N.Y. TIMES (Mar. 6, 2017), <https://www.nytimes.com/2017/03/06/us/politics/supreme-court-repeat-sex-offenders.html> (on file with *Ohio State Law Journal*).

⁸³ Sharad Goel et al., *Combating Police Discrimination in the Age of Big Data*, 20 NEW CRIM. L. REV. 181, 188 (2017) [hereinafter Goel et al., *Combating Police Discrimination*]; Sharad Goel et al., *Personalized Risk Assessments in the Criminal Justice System*, 106 AM. ECON. REV. 119, 119–23 (2016) [hereinafter Goel et al., *Personalized Risk Assessments*]; Sharad Goel et al., *Precinct or Prejudice? Understanding Racial Disparities in New York City’s Stop-and-Frisk Policy*, 10 ANNALS OF APPLIED STAT. 365, 383 (2016) [hereinafter Goel et al., *Precinct or Prejudice?*].

⁸⁴ Goel et al., *Combating Police Discrimination*, *supra* note 83, at 181.

that around 40% of the stops made by the police in New York City in the period studied were based on observations or other factors that had only a 1% chance of finding a weapon, usually a knife, and that police have used these low-percentage stops disproportionately on blacks and Latinos.⁸⁵ By contrast, where more reliable factors were the basis of stops, it was determined that 6% of all stops were far more likely to result in the discovery of weapons, and simultaneously to mitigate racial disparities.⁸⁶

Professor Ferguson and Professor Damien Bernache capture the idea in the context of another claim based on resort to common sense reasoning: police testimony that a particular location constitutes a high-crime area.⁸⁷ Using the *Wardlow* case as a prototypical example, they suggest a framework for courts to use to identify a high-crime area, based on “objective and verifiable evidence.”⁸⁸ The judgment of whether or not the location is a high-crime area should follow from real data based on the actual experience of police officers on the ground.

Max Minzner makes a similar argument in contrasting the assertions that police officers make in support of probable cause in applications for search warrants—which, he finds, are correct in about 80% of the cases—with those made in quick, on-the-street judgments, which are no more than 12% accurate.⁸⁹ In the latter category, which includes *Terry* stops and frisks, the judicial assessment of probable cause or reasonable suspicion should include evidence regarding the officers’ “differential success rates”—in other words, proof of their hit rates or rates of success over the great run of stops and frisks. “These success rates,” Minzner says, “capture information not currently analyzed in the search process and their addition would improve the accuracy of probable-cause decisions.”⁹⁰

Professor Jeffrey Fagan argues that the low standard of reasonable suspicion for *Terry* stops and frisks—what Fagan calls *Terry*’s “original sin”—has allowed police to take action against members of the public that the probable

⁸⁵ Goel et al., *Precinct or Prejudice?*, *supra* note 83, at 374–75, 382.

⁸⁶ *Id.* at 382. This study is referenced in more detail, *infra* notes 261–72 and accompanying text.

⁸⁷ Andrew Guthrie Ferguson & Damien Bernache, *The “High-Crime Area” Question: Requiring Verifiable and Quantifiable Evidence for Fourth Amendment Reasonable Suspicion Analysis*, 57 AM. U. L. REV. 1587, 1629 (2008).

⁸⁸ *Id.*

⁸⁹ Minzner, *supra* note 50, at 914, 921.

⁹⁰ *Id.* at 913. Professor Anna Lvovsky addresses this issue from a different angle: examining the ways in which courts have accepted “police expertise” in a range of judgments relating to the detection and investigation of criminal conduct. She notes that the Court frequently assumes that police possess “rare and reliable ‘expert’ insight,” without supporting empirical evidence. Anna Lvovsky, *The Judicial Presumption of Police Expertise*, 130 HARV. L. REV. 1995, 2081 (2017). Similar criticism has been made of police forensic science, much of which has been shown to be unreliable by DNA testing. *See, e.g.*, Adam B. Shniderman, *Prosecutors Respond to Calls for Forensic Science Reform: More Sharks in Dirty Water*, YALE L.J.F. 348, 348 (2017).

cause standard would have disallowed, all without evidence of any effectiveness.⁹¹ Using large datasets from the *Floyd* litigation in New York City, Fagan shows that the lower standard of reasonable suspicion makes policing less, not more, effective at controlling crime.⁹² Stated differently, the data prove that stops police made using the probable cause standard actually result in reductions in crime; by contrast, stop activity utilizing the more subjective and vague *Terry* reasonable suspicion standard may have very little effect on crime.⁹³

IV. DATA, EMPIRICISM AND POLICE STOP-AND-FRISK PRACTICES

These analytical articles and related research are dependent on the availability of comprehensive and reliable data. Indeed, as data collection and analysis has grown, so has empiricism, a process greatly enabled by big (and little) data. We turn therefore to an overview of empirical evidence in administrative and judicial decision-making.

A. *The Role of Empiricism and Data in the Courts*

The United States Supreme Court has given inconsistent signals regarding the role of statistics, economic analysis, and empirical evidence in judicial adjudications. In some areas, the Court has emphasized the need for empiricism. For example in *Daubert v. Merrell Dow Pharmaceuticals*,⁹⁴ the Court rejected the long-standing *Frye* test of “general acceptance” with respect to the introduction of expert testimony.⁹⁵ Interpreting Federal Rule of Evidence 702, which imposes a standard for the use of expert testimony based on scientific or

⁹¹ Fagan, *supra* note 56, at 45.

⁹² *See id.* at 84–95; *see also* *Floyd v. City of New York*, 959 F. Supp. 2d 540, 575 (S.D.N.Y. 2013).

⁹³ The recent data regarding stops and frisks and violent crime rates in New York City support this thesis. In 2016, police reported approximately 12,500 pedestrian stops (representing a huge and precipitous drop from the peak of stop-and-frisk policing of close to 700,000 stops in 2011), yet the violent crime rate continued to fall. *Stop-and-Frisk Data*, N.Y.C.L. UNION, <https://www.nyclu.org/en/stop-and-frisk-data> [<https://perma.cc/6Z4T-SR8Z>]; *see also Latest Data: Stop-and-Frisk and Crime Both Lowest in Years*, N.Y.C.L. UNION (Oct. 11, 2016), <https://www.aclu.org/news/latest-data-stop-and-frisk-and-crime-both-lowest-years> [<https://perma.cc/AKJ7-M55P>] [hereinafter N.Y.C.L. UNION] (finding there were 7,636 stops for the first six months). However, the court-appointed monitor in *Floyd v. City of New York*, Peter Zimroth, has found a pattern of undocumented stops by police officers. Al Baker, *City Police Officers Are Not Reporting All Street Stops, Monitor Says*, N.Y. TIMES (Dec. 13, 2017), <https://www.nytimes.com/2017/12/13/nyregion/nypd-stop-and-frisk-monitor.html> [<https://perma.cc/25B4-YNGD>].

⁹⁴ *Daubert ex rel. Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

⁹⁵ *Id.* at 597 (rejecting test set forth in *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923)).

technical expertise, the Court opted for a test with increased scientific rigor.⁹⁶ Rule 702 requires that “all scientific testimony or evidence” must be “not only relevant but reliable.”⁹⁷ Reliability is based on sound scientific processes, and “scientific knowledge” refers to “known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds.”⁹⁸ “[I]n order to qualify as ‘scientific knowledge,’ an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—*i.e.*, ‘good grounds,’ based on what is known.”⁹⁹ The *Daubert* standard has produced uneven results, but at its core, it supports the use of empirical evidence and analysis in situations presenting the relevance of testable factual propositions.¹⁰⁰

Yet, just a few years before *Daubert*, the Court had rejected what seemed to be decisive statistical proof of racial bias in capital sentencing practices. In *McCleskey v. Kemp*,¹⁰¹ a challenge to the Georgia death penalty system rested upon statistical studies by Dr. David Baldus that examined 2,000 Georgia murder cases from the 1970s in an attempt to discern patterns in the jury decisions on the death penalty.¹⁰² Professor Baldus showed that “defendants charged with killing white persons received the death penalty in 11% of the cases, but defendants charged with killing blacks received the death penalty in only 1% of the cases.”¹⁰³ The numbers also indicated a “reverse racial disparity according to the race of the defendant: 4% of the black defendants received the

⁹⁶ *Id.* at 589 (“That the *Frye* test was displaced by the Rules of Evidence does not mean, however, that the Rules themselves place no limits on the admissibility of purportedly scientific evidence. To the contrary, under the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.” (footnote omitted)).

⁹⁷ *Id.*

⁹⁸ *Id.* at 590 (citing WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 1252 (1986)).

⁹⁹ *Id.*

¹⁰⁰ *Daubert*, 509 U.S. at 589 (“[U]nder the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.”); *see also* *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 147 (1999) (describing Rule 702, as construed in *Daubert*, as requiring the trial court to take on a “basic gatekeeping obligation” to ensure that scientific and technical testimony are not only relevant but reliable and based on good grounds). A number of commentators have pointed to the lack of *Daubert* standards in criminal litigation where untested police “science” is not infrequently admitted. *See, e.g.*, D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99, 99–100, 149 (2000) (noting in the aftermath of *Daubert*, federal judges seemed inclined to take a much more active role in keeping questionable science out of the courtroom in civil cases, in which that scientific work came to court as proposed evidence from plaintiffs seeking to prove damages, but much less inclined to do so when questionable science originated with the prosecution in criminal cases).

¹⁰¹ *McCleskey v. Kemp*, 481 U.S. 279 (1987).

¹⁰² *Id.* at 286–87.

¹⁰³ *Id.* at 286.

death penalty, as opposed to 7% of the white defendants.”¹⁰⁴ The results of Baldus’s more sophisticated analysis, controlling for thirty-nine nonracial variables, were even more damning. “[D]efendants charged with killing white victims were 4.3 times as likely to receive a death sentence as defendants charged with killing blacks. According to this model, black defendants were 1.1 times as likely to receive a death sentence as other defendants.”¹⁰⁵ Thus, the Court said, “the Baldus study indicates that black defendants, such as McCleskey, who kill white victims have the greatest likelihood of receiving the death penalty.”¹⁰⁶

Even with this dramatic statistical evidence of racial disparity in a high-stakes process, the Court found the empirical evidence beside the point as it did not show that “the decisionmakers in *his* case acted with discriminatory purpose.”¹⁰⁷ The Court allowed that it had considered statistical proof in other areas of the law, for example, statistical disparities as proof of an equal protection violation in the selection of the jury venire in a particular district, and “multiple-regression analysis to prove statutory violations under Title VII of the Civil Rights Act of 1964.”¹⁰⁸ There seems little doubt but that the Court believed that the statistical evidence of disparity as evidence of discrimination, even if the strength and rigor of that evidence seemed unassailable, would prove too powerful in ways that the Court did not welcome. Not only would such a ruling jeopardize the institution of capital punishment, but given the evidence of race discrimination throughout the entire criminal justice system, the impact could be even broader, a concern that Justice Brennan rightly captured in his dissenting opinion as a fear of “too much justice.”¹⁰⁹

The Court turned to empiricism again in *Whole Woman’s Health v. Hellerstedt*,¹¹⁰ a case that examined a Texas law that imposed new requirements on medical facilities in which women could obtain abortions.¹¹¹ The Court reviewed the statute under the *Planned Parenthood v. Casey* “undue burden” test.¹¹² The Texas law required that any doctor performing abortions have “admitting privileges” at a hospital not more than thirty miles from the location of the abortion clinic and required that the abortion facility meet the minimum standards set for “ambulatory surgical centers.”¹¹³ The plaintiffs challenged these requirements as undue burdens on the constitutional right to obtain

¹⁰⁴ *Id.*

¹⁰⁵ *Id.* at 287.

¹⁰⁶ *Id.*

¹⁰⁷ *McCleskey*, 481 U.S. at 292.

¹⁰⁸ *Id.* at 294.

¹⁰⁹ *Id.* at 339 (Brennan, J., dissenting).

¹¹⁰ *Whole Woman’s Health v. Hellerstedt*, 136 S. Ct. 2292 (2016).

¹¹¹ *Id.* at 2300.

¹¹² *Planned Parenthood of Se. Pa. v. Casey*, 505 U.S. 833, 878 (1992) (plurality opinion). The plurality opinion stated that a law creates an undue burden on a woman’s right to have a legal abortion, making the law unconstitutional, if the “purpose or effect” of the law is to “place a substantial obstacle in the path of a woman” seeking a previability abortion. *Id.*

¹¹³ *Whole Woman’s Health*, 136 S. Ct. at 2300.

abortions, alleging that physicians could not obtain the required admitting privileges, and that because the cost of retrofitting existing clinics to comply with the ambulatory surgical center standards would prove so great that many clinics would close, leaving huge areas of Texas without any abortion providers or putting them so far away that the right to an abortion would exist in name only.¹¹⁴

In the view of the majority, in the absence of relevant legislative findings, the factual record showed no medical need for the law. Peer-reviewed studies and expert testimony showed “no significant health related problem [for the admitting privileges requirement] to cure” and the state’s evidence failed to show that the new law advanced any interest in women’s health beyond what the prior law already provided.¹¹⁵ “At the same time,” the Court said, the evidence showed that the admitting privileges requirement created a “substantial obstacle” in the path of a woman wanting a legal abortion, in the form of a “dramatic” drop in the number of clinics, fewer available doctors to perform the procedures, and thus longer wait times.¹¹⁶ Similarly, the ambulatory-surgical-center standards imposed requirements that provided “no benefit when complications arise,” and that statistically, “abortions taking place in abortion facilities are . . . safer than numerous procedures that take place outside hospitals and to which Texas does not apply its surgical-center requirements.”¹¹⁷ The empirical facts in the record, the Court said, supported the trial court’s “ultimate legal conclusion that the . . . requirement is not necessary.”¹¹⁸

The use of data analysis has also come to the fore in cases challenging voter identification statutes. These statutes are premised on voter fraud by persons voting in the place of others. A robust debate and many investigations have found no evidence for these claims, but legislators persist in these enactments.¹¹⁹ The legal challenges to these laws are based on allegations that they disenfranchise poor and minority voters, because they are typically less likely to have acceptable forms of identification. To prove these claims, the challengers have presented experts from various disciplines who have used statistical tools to examine, compare, and analyze the effect of voter identification laws on voting populations, in order to measure just how many voters, from what demographic groups, the laws would disenfranchise.

¹¹⁴ *Id.* at 2310–18.

¹¹⁵ *Id.* at 2311.

¹¹⁶ *Id.* at 2312–18.

¹¹⁷ *Id.* at 2315.

¹¹⁸ *Id.* at 2316.

¹¹⁹ See Wendy Underhill, *Voter Identification Requirements: Voter ID Laws*, NAT’L CONF. ST. LEGISLATURES (Jan. 5, 2018), <http://www.ncsl.org/research/elections-and-campaigns/voter-id.aspx> [<https://perma.cc/DA7D-6N6P>].

For example, Pennsylvania required particular kinds of state-issued identification bearing a photograph of the person attempting to vote.¹²⁰ In *Applewhite v. Commonwealth*, the Commonwealth Court looked closely at the statistical, data-based evidence on the issue of how many people the law would disenfranchise.¹²¹ The court credited statistical analysis that showed that 511,415 registered voters in the state lacked a form of identification acceptable for voting.¹²² On the related question of voter awareness of the voter identification requirement, plaintiffs presented data on the results of inquiries based on telephone interviews that showed minority voter disenfranchisement would result from the law.¹²³

In *Veasey v. Perry*,¹²⁴ a challenge to the Texas voter identification law was based on claims that the law constituted a substantial burden on the right to vote, discriminated against minority voters, and constituted a poll tax.¹²⁵ Empirical evidence was presented by “an expert in quantitative and qualitative historical analysis of voting, political, and statistical data,” whose report showed “intentional discrimination against minorities to achieve a partisan political advantage.”¹²⁶ For example, the “provision allowing the use of concealed handgun permits favor[ed] Anglos because they are disproportionately represented among those permit holders,” and that was true as well for military veterans’ identification because “Anglos are a disproportionate share of Texas’s military veterans of voting-age population relative to African-Americans.”¹²⁷ The evidence also showed that “[w]hen the legislature rejected student IDs, state government employee IDs, and federal IDs, they rejected [forms of identification] that are disproportionately held by African Americans and Hispanics.”¹²⁸ In addition, two different experts used database matching (as in the *Applewhite* case in Pennsylvania) and other statistical methods “to determin[e] the number of registered voters who might lack [acceptable forms of identification], along with their demographic characteristics.”¹²⁹ Based on their work, the court found that “approximately 608,470 registered voters in Texas, representing approximately 4.5% of all registered voters, lack qualified [acceptable forms of identification], and of these, 534,512 voters do not qualify for [an] exemption. Moreover, a disproportionate number of African-Americans and Hispanics populate that group of potentially disenfranchised voters.”¹³⁰ The

¹²⁰ *Applewhite v. Commonwealth*, No. 330 M.D.2012, 2014 WL 2619590, at *1 (Pa. Commw. Ct. Apr. 28, 2014).

¹²¹ *Id.* at *6–7.

¹²² *Id.* at *6.

¹²³ *Id.* at *9.

¹²⁴ *Veasey v. Perry*, 71 F. Supp. 3d 627 (S.D. Tex. 2014), *aff’d in part, vacated in part, and judgment rendered in part*, 830 F.3d 216 (5th Cir. 2016) (en banc).

¹²⁵ *Id.* at 633.

¹²⁶ *Id.* at 658.

¹²⁷ *Id.*

¹²⁸ *Id.*

¹²⁹ *Id.* at 659.

¹³⁰ *Veasey*, 71 F. Supp. 3d at 659.

statistical evidence formed the foundation for the court's finding that the law disenfranchised and discriminated against minority voters.¹³¹

Yet, for each area in which the courts have looked to statistical and other empirical evidence to inform their fact-finding process, in others there continues to be reliance on intuition and common sense, even where science and social science provides more advanced metrics. For example, in *Manson v. Braithwaite*,¹³² the Court ruled that in determining reliability of identifications, a court should consider the opportunity of the witness to view the perpetrator, his degree of attention, the accuracy of the witness's prior description, level of the witness's certainty, and the time between the crime and confrontation.¹³³ This due process analysis has remained unchanged for over forty years, notwithstanding a multitude of studies that demonstrate that the factors that the Court has credited in assessing reliability where suggestive practices may have been used are not valid.¹³⁴ And, there are continuing debates about empirical data and normative constitutional principles in other areas.¹³⁵

¹³¹ Statistical analysis also played a significant role in the ruling of a three-judge district court in North Carolina finding unconstitutional the 2016 North Carolina redistricting plan on grounds that the plan discriminated against voters who support Democratic Party candidates. *Common Cause v. Rucho*, 279 F. Supp. 3d 587, 608, 629 (M.D.N.C. 2017), *stay granted*, 138 S. Ct. 923 (2018) (mem.). In *Common Cause*, the plaintiffs presented a range of expert testimony showing that the plan is an "extreme statistical outlier" in its disfavoring of non-Republican candidates, and that there is a large "partisan symmetry" resulting in a translation of votes into a Republican Party advantage. *Id.*

¹³² *Manson v. Braithwaite*, 432 U.S. 98 (1977).

¹³³ *Id.* at 114.

¹³⁴ See, e.g., *Dennis v. Sec'y, Pa. Dep't of Corr.*, 834 F.3d 263, 314 (3d Cir. 2016) (en banc) (McKee, C.J., concurring); *Young v. State*, 374 P.3d 395, 414–16 (Alaska 2016); *State v. Henderson*, 27 A.3d 872, 928, *modified*, *State v. Chen*, 27 A.3d 930, 938 (N.J. 2011); *State v. Lawson*, 291 P.3d 673, 695 (Or. 2012); U.S. DEP'T OF JUSTICE, EYEWITNESS EVIDENCE: A GUIDE FOR LAW ENFORCEMENT 1–4 (Oct. 1999); Gary L. Wells & Deah S. Quinlivan, *Suggestive Eyewitness Identification Procedures and the Supreme Court's Reliability Test in Light of Eyewitness Science: 30 Years Later*, 33 LAW & HUM. BEHAV. 1 (2008) (criticizing continued use of criteria for assessing reliability of identifications that have been shown to be faulty by social science studies). The Supreme Court had the opportunity to revise these criteria in *Perry v. New Hampshire*, 565 U.S. 228, 232–33 (2012), but avoided the issue in ruling that that there was no "state action" in the particular identification procedure used in that case. *Id.* at 233.

¹³⁵ See, e.g., David L. Faigman, *The Supreme Court's Confused Empirical Jurisprudence*, Article in *Expert Evidence Report*, BLOOMBERG BNA (July 14, 2015), <https://uconsortiumssl.files.wordpress.com/2015/08/faigman-bna-glossip-g2i.pdf> [https://perma.cc/4QUB-AGG3] (criticizing decision in *Glossip v. Gross*, 135 S. Ct. 2726 (2015), for applying the "clearly erroneous" standard review in a lethal injection case). In cases presenting the issue of whether a person has a reasonable expectation of privacy sufficient to trigger Fourth Amendment review, the doctrinal question of whether the expectation is objectively reasonable has been made in many instances without regard as to how a majority of persons assess their privacy interests in various locations and contexts. Public surveys and polls show a large disconnect between what many persons view as private areas and conduct and what the Supreme Court has recognized as an objectively reasonable expectation of

B. Risk Assessment Tools

Risk assessment tools are data-based instruments that allow judgments and decisions to be made on an empirical measure of the actual, as opposed to an assumed or intuited risk. These tools could assist judges and others to assess the risks inherent in a criminal justice decision, based on data reflected in thousands of individual cases with respect to bail, sentencing, and parole. For example, under most governing statutes and rules, the bail/pretrial release decision centers upon the question of whether the arrestee, if released, will appear at later court proceedings. In practice, however, courts regularly place significant weight on their assessment of whether the suspect, if released, would pose a danger to the community. Where a court has serious concerns about the dangerousness of the suspect, bail will be set at an amount the defendant is unlikely to make.

In some jurisdictions, courts do not do any risk balancing; instead, there is a bail “schedule” laying out the preset amounts of bail, based solely on the current charge and criminal history information. Many have noted the irrational aspects of these approaches.¹³⁶ A judge considering a defendant for possible release will almost always have some concerns about the possibility that the suspect will offend while on release, perhaps violently. But when judges set bail, a suspect who has sufficient resources can secure release by posting money or property even if she represents a true risk to the public. By contrast, an indigent suspect in a low-level, non-dangerous offense, no matter how sure to show up in court, will remain in custody if a money bond is set.

Might there be empirically based alternatives? Several years ago, the Laura and John Arnold Foundation designed a tool to give judges the benefit of “data-driven, objective risk assessment[s]” of the risks individuals posed.¹³⁷ The goal was to enable the courts to release those defendants who did not pose a risk of committing crimes, especially violent crimes, or who were not serious flight risks.¹³⁸ Some jurisdictions used risk assessment tools, and though they were imperfect, there was some evidence that these jurisdictions had “been able to spend less on pretrial incarceration, while at the same time enhancing public safety.”¹³⁹ The Foundation’s researchers created a tool that included nine factors, all available from the suspects’ record, as the most pertinent risk

privacy. *See, e.g.*, Christopher Slobogin & Joseph E. Schumacher, *Reasonable Expectations of Privacy and Autonomy in Fourth Amendment Cases: An Empirical Look at “Understandings Recognized and Permitted by Society,”* 42 DUKE L.J. 727, 772–74 (1993); *cf.* Williams v. Sec’y Pa. Dep’t of Corr., 848 F.3d 549, 567–69 (3d Cir. 2017) (discussing empirical studies showing the impact of long-term solitary confinement on physical and mental health of inmates), *cert. denied*, 138 S. Ct. 357 (2017) (mem.).

¹³⁶ *See, e.g.*, Richard Fausset, *Bail Was \$500, Money He Didn’t Have. Atlanta Faces Calls for Change.*, N.Y. TIMES (Jan. 11, 2018), <https://www.nytimes.com/2018/01/11/us/atlanta-bail-courts-reform.html> (on file with *Ohio State Law Journal*).

¹³⁷ LAURA & JOHN ARNOLD FOUND., DEVELOPING A NATIONAL MODEL FOR PRETRIAL RISK ASSESSMENT 5 (Nov. 2013).

¹³⁸ *Id.* at 1.

¹³⁹ *Id.* at 2.

predictors for “new crime, new violence, and failure to appear [for court].”¹⁴⁰ Their studies suggested that defendants failed at similar rates in each of the three risk categories, regardless of race or gender, which would support the assertion “that the assessment [tool] does not over-classify non-whites’ risk levels, which has been a concern in some other areas of risk assessment.”¹⁴¹

When testing began, preliminary results showed that the Arnold Foundation tool—called PSA-Court—“successfully predict[s] criminal reoffending and failing to return to court.”¹⁴² By July 2015, PSA-Court had been rolled out nationally, in courts across the United States “in 29 jurisdictions, including three entire states—Arizona, Kentucky, and New Jersey—as well as three of the largest cities in the country—Charlotte, Chicago, and Phoenix.”¹⁴³ As these jurisdictions implemented their use of the tool, results seemed promising. In Kentucky, both jail populations and pretrial crime rates fell; in Mecklenburg County, North Carolina (the location of the city of Charlotte), the jail population fell by almost 20%.¹⁴⁴ However, these reported trends, as well as the causative relationship between the risk assessment tools and releases of pretrial defendants, have been called into question in more recent studies.¹⁴⁵

But with the increasing use of data-based risk assessment across the criminal justice system, criticism has arisen. One such system, created by a private company, Northpointe, is used to predict the likelihood of re-offending for purposes of sentencing decisions.¹⁴⁶ An investigation by ProPublica of seven thousand cases in Broward County, Florida, for 2013 and 2014 provided a disquieting picture.¹⁴⁷ ProPublica found that “[o]nly 20[%] of the people

¹⁴⁰ *Id.* at 3–4.

¹⁴¹ *Id.* at 5.

¹⁴² *Id.*

¹⁴³ Press Release, Laura & John Arnold Found., More than 20 Cities and States Adopt Risk Assessment Tool To Help Judges Decide Which Defendants To Detain Prior to Trial (June 26, 2015), <http://www.arnoldfoundation.org/more-than-20-cities-and-states-adopt-risk-assessment-tool-to-help-judges-decide-which-defendants-to-detain-prior-to-trial/> [<https://perma.cc/NDB4-BLXJ>].

¹⁴⁴ *Id.*

¹⁴⁵ See John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. (forthcoming 2018) (manuscript at 24–39, 49–51), <https://ssrn.com/abstract=3041622> [<https://perma.cc/384S-WUC8>] (explaining that the data jurisdictions are using for risk assessment are outdated and will limit the potential positive impact of reform); Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. (forthcoming 2018) (manuscript at 28–50), <https://ssrn.com/abstract=3016088> [<https://perma.cc/5D4N-YVY5>] (using data from Kentucky, after the state’s adoption of pretrial risk assessment, to show that it had little impact on pretrial release and cost effectiveness).

¹⁴⁶ Northpointe, *Practicioners Guide to COMPAS 14*, EQUIVANT (Aug. 17, 2012), http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf [<https://perma.cc/M2V4-NK3W>].

¹⁴⁷ Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<https://perma.cc/4JS5-ZBFF>].

predicted to commit violent crimes actually went on to do so.”¹⁴⁸ With respect to re-offending at any level, including lower level, misdemeanor crimes, the tool worked somewhat better: “Of those deemed likely to re-offend, 61[%] were arrested for any subsequent crimes within two years.”¹⁴⁹ More disturbing, ProPublica’s examination appears to have uncovered racial bias in the process. “The [Northpointe] formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants.”¹⁵⁰ Northpointe, which did not disclose its entire process and formula for arriving at the risk assessment scores—the company says that information is proprietary—denied ProPublica’s findings, disagreeing with the methodology used and the accuracy of ProPublica’s testing.¹⁵¹ Others have suggested that the bias findings may have more to do with biases in the underlying criminal justice system and data, and not with Northpointe’s tool.¹⁵²

The lessons seem clear. Risk assessment tools may well provide more reliable and predictive information than current methodologies, but they present their own problems. Validation by uninterested sources should be a universal requirement, and as Professor Christopher Slobogin has stated, “[r]isk assessments should be impermissible unless both parties get to see all the data that go into them It should be an open, full-court adversarial proceeding.”¹⁵³

¹⁴⁸ *Id.*

¹⁴⁹ *Id.*

¹⁵⁰ *Id.*; see also Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> [<https://perma.cc/622Y-45WG>]. There has also been scholarly analysis and criticism. See, e.g., John Monahan, *Risk Assessment in Sentencing*, in 4 REFORMING CRIMINAL JUSTICE 77, 77–78 (Erik Luna ed., 2017); Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537 (2015); Starr, *supra* note 73, at 805–08, 838.

¹⁵¹ Angwin et al., *supra* note 147 (“In a letter, [Northpointe] criticized ProPublica’s methodology and defended the accuracy of its test”).

¹⁵² Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There’s Software Used Across the Country To Predict Future Criminals. And It’s Biased Against Blacks,”* 80 FED. PROB. 38, 38 (2016); Max Ehrenfreund, *The Machines that Could Rid Courtrooms of Racism*, WASH. POST (Aug. 18, 2016), <https://www.washingtonpost.com/news/wonk/wp/2016/08/18/why-a-computer-program-that-judges-rely-on-around-the-country-was-accused-of-racism/> [<https://perma.cc/9WU8-33CX>] (explaining that ProPublica’s statistics “are misleading” because they “could reflect broader racial disparities” both in the criminal justice system and society at large).

¹⁵³ Angwin et al., *supra* note 147; see also *Malenchik v. State*, 928 N.E.2d 564, 575 (Ind. 2010) (upholding the use of a risk assessment tool in the sentencing context); *State v. Loomis*, 881 N.W.2d 749, 761–64 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017) (mem.) (examining the use of risk assessment at sentencing); Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 566–67 (2018); Monahan, *supra* note 150, at 93–94; Wexler, *supra* note 73. Predictive analytics are being employed in a wide range of disciplines and industries, and these analytics continue to show the inherent weaknesses in common sense

C. Data, Empiricism, and Stops and Frisks

Even with considerable data regarding stop-and-frisk practices potentially available to police departments and ultimately to the courts, only a few departments actually collect and maintain this data in a comprehensive and usable form. Nothing in *Terry*, in any other Supreme Court decisions, or in any federal law required or incentivized the collection of this data. The few police departments that have collected data—most prominently, New York City and Philadelphia—have done so as a result of litigation.¹⁵⁴

But there are signs of change. In a 2014 article, David Harris published the results of a survey of fifty-five of the largest police departments in the country.¹⁵⁵ The survey, conducted in 2011 and 2012, garnered forty-four responses, and out of those, twenty-three reported requiring police officers to collect at least some data on stops and frisks.¹⁵⁶ Three others made data collection discretionary; eighteen had no requirement.¹⁵⁷ Of the twenty-three departments requiring some data collection, twenty-one reported that the data included racial and ethnic characteristics of those stopped; thirteen reported that the data were available to the public in some form.¹⁵⁸ The departments included some of the largest law enforcement agencies in the country, thus encompassing a significant number of police officers.¹⁵⁹

No doubt some aspects of this change—while too little and too slow, but a noticeable shift nonetheless—are motivated by supervisors’ desire to effectively oversee day-to-day precinct-level operations. The old saw from the business world, “you can’t manage what you don’t measure,” applies as well to police

judgments not tested by reliable data. See, e.g., Joseph B. Treaster, *Will You Graduate? Ask Big Data*, N.Y. TIMES (Feb. 2, 2017), <https://www.nytimes.com/2017/02/02/education/edlife/will-you-graduate-ask-big-data.html> (on file with *Ohio State Law Journal*) (finding early course performance as the best predictor of graduation rates).

¹⁵⁴ See, e.g., Settlement Agreement, Class Certification, and Consent Decree at 3–7, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. June 21, 2011) [hereinafter *Philadelphia Consent Decree*]; *infra* notes 171–85 and accompanying text. Missouri is the only state that statutorily requires collection of stop data, but even there it is limited to traffic stops. MO. ANN. STAT. § 590.650 (West 2011). The ABA and NAACP Legal Defense Fund have remarked: “We cannot understand what we cannot measure, and we cannot change what we cannot understand.” Press Release, Am. Bar Ass’n & LDF, American Bar Association and NAACP Legal Defense Fund Joint Statement on Eliminating Bias in the Criminal Justice System (July 2015), <http://www.naacpldf.org/files/about-us/ABA-LDF%20Joint%20Statement%20on%20Eliminating%20Bias%20in%20the%20Criminal%20Justice%20System.pdf> [<https://perma.cc/QYY2-7KLQ>]; see also Rachel Harmon, *Why Do We (Still) Lack Data on Policing?*, 96 MARQ. L. REV. 1119, 1123–24 (2013) (detailing the complexities of local politics and their impact on police department data transparency).

¹⁵⁵ Harris, *supra* note 57, at 856.

¹⁵⁶ *Id.* at 870–71.

¹⁵⁷ *Id.* at 871.

¹⁵⁸ *Id.*

¹⁵⁹ *Id.*

management.¹⁶⁰ In other departments, pressure to collect data about stop-and-frisk practices has come from the public. Stops and frisks have long constituted a frequent point of friction and contention between police officers and members of the public, particularly with members of African American and other minority communities, a fact noted in the *Terry* opinion.¹⁶¹

The stop-and-frisk crime suppression strategy of the New York City police department became a major political and legal issue and ultimately played a significant role in the 2013 New York mayoral election.¹⁶² Thereafter, the combination of a change of mayoral administrations and the limits imposed by the *Floyd* court has led to a precipitous reduction in the number of stops and frisks.¹⁶³ Similarly, in Chicago, in March 2015, months before the city erupted after the release of the video of the police shooting of Laquan McDonald,¹⁶⁴ the American Civil Liberties Union (ACLU) of Illinois had turned a spotlight on the Police Department's unlawful use of stops and frisks, especially against African Americans and other people of color.¹⁶⁵ The Illinois ACLU and others demanded considerably improved data collection for all stops and frisks.¹⁶⁶ In Boston, complaints that stops and frisks targeted African Americans led to changes in the collection of data in that city. A paper-based system, that a high-ranking police department spokesperson conceded had "a lot of potential for data errors," was changed into an electronic system.¹⁶⁷ According to Boston City Council President Michelle Wu, who made more transparency in city government one of her top issues, "the goal is to say [to the public,] here is the

¹⁶⁰ Daniel Denvir, *The Key Ingredient in Stop-and-Frisk Reform: Open Data*, CITYLAB (Aug. 24, 2015), <http://www.citylab.com/crime/2015/08/the-missing-ingredient-in-stop-and-frisk-accountability-open-data/402026/> [<https://perma.cc/NGN4-QVGJ>].

¹⁶¹ *Terry v. Ohio*, 392 U.S. 1, 14–15 (1968).

¹⁶² Annie Karni, *How Bill de Blasio's Nimble 2013 Mayoral Election Campaign Got Democratic Party Back into Gracie Mansion*, N.Y. DAILY NEWS (Nov. 5, 2013), <http://www.nydailynews.com/news/election/bill-de-blasio-nimble-campaign-made-mayor-article-1.1507754> (on file with *Ohio State Law Journal*) ("[De Blasio] campaigned relentlessly against . . . stop-and-frisk tactics that have resulted in police stopping hundreds of thousands of innocent minorities.").

¹⁶³ N.Y.C.L. UNION, *supra* note 93, at 2; Fifth Report of the Independent Monitor: Analysis of NYPD Stops Reported, 2013–2015, *Floyd v. New York*, 959 F. Supp. 2d 540 (S.D.N.Y. 2013) (No. 08 Civ. 1034).

¹⁶⁴ See Christy Gutowski, *2 Years Later, Laquan McDonald Shooting Leaves a Trail of Change*, CHI. TRIB. (Oct. 20, 2016), <http://www.chicagotribune.com/news/laquanmcdonald/ct-laquan-mcdonald-shooting-anniversary-met-20161020-story.html> [<https://perma.cc/3SQC-VAE5>].

¹⁶⁵ ACLU OF ILLINOIS, STOP AND FRISK IN CHICAGO 2–3 (Mar. 2015).

¹⁶⁶ *Id.* ("Currently, officers are not required to record when they frisk someone. . . . Absent a record, supervisors and the public have no means to determine whether officers' searches are lawful. Officers should record frisks, the reason for the frisk (which must be separate from the reason for the stop), and the results of the search (i.e., whether there was a weapon or other contraband and if so, what type.")).

¹⁶⁷ Jordan Graham, *Boston Police To Update Traffic Stop Database*, GOV'T TECH. (Jan. 20, 2016), <http://www.govtech.com/public-safety/Boston-Police-to-Update-Traffic-Stop-Database.html> [<https://perma.cc/26PE-WLM2>].

information we're collecting, can you help us out, look for trends we haven't seen."¹⁶⁸

In Newark, New Jersey, then-Mayor (currently Senator) Cory Booker came to an agreement with citizens and advocacy groups to make all police reports of stops and frisks available to the public.¹⁶⁹ The city pledged to make stop-and-frisk reports, “detailing the race, gender, age, force[] used, and arrests made,” available monthly.¹⁷⁰ The idea was to improve police–community relations by building trust with the community, since it would have all the information necessary to see what the police did and whether any abuses had taken place, and simultaneously to help police to focus less on the wrong targets by indicating those searches yielding fewer arrests and seizures of contraband than others.

The extensive data collection efforts in New York and Philadelphia came about as a result of settlements of lawsuits brought by private parties represented by nongovernmental organizations. In New York, the police department had kept internal records of stops and frisks for years, requiring officers to record the details of the encounters on a form known as UF-250.¹⁷¹ In a 1999 class action lawsuit, *Daniels v. City of New York*,¹⁷² filed by plaintiffs represented by the Center for Constitutional Rights (CCR), a court required that the New York Police Department (NYPD) provide all UF-250s completed by officers to CCR, thus creating a database of stop-and-frisk data from the largest police department in the country, which was making increasingly intense use of stops and frisks.¹⁷³ In Philadelphia, the ACLU of Pennsylvania's 2010 lawsuit, *Bailey v. City of Philadelphia*,¹⁷⁴ resulted in a consent decree that required the collection and electronic maintenance of data on all stops and frisks to be shared with an independent monitor and the ACLU lawyers for use in monitoring the stop-and-frisk program under Fourth and Fourteenth Amendment standards.¹⁷⁵

The federal government has also imposed data-collection requirements on police departments. Pursuant to authority granted by Congress in a 1994 statute known as the “pattern or practice” law, the DOJ may now investigate and, if necessary, bring suit in federal court upon findings that a state or local police department has engaged in a pattern of conduct or consistently engages in practices that deprive people of their constitutional rights.¹⁷⁶ The Civil Rights

¹⁶⁸ *Id.*

¹⁶⁹ Gregory Ferenstein, *Cory Booker Opens Stop-and-Frisk Data to the Public. Here's Why It Might Help.*, TECHCRUNCH (Aug. 27, 2013), <https://techcrunch.com/2013/08/27/corybooker-opens-stop-and-frisk-data-to-the-public-heres-why-it-might-help/> [<https://perma.cc/RQ3T-A4CE>].

¹⁷⁰ *Id.*

¹⁷¹ *Daniels v. City of New York*, 138 F. Supp. 2d 562, 563 n.2 (S.D.N.Y. 2001).

¹⁷² *Id.* at 565.

¹⁷³ *Id.* at 563–65.

¹⁷⁴ *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. filed Nov. 4, 2010).

¹⁷⁵ Settlement Agreement, Class Certification, and Consent Decree at 3–7, *Bailey*, No. 10-5952.

¹⁷⁶ 42 U.S.C. § 14141 (2012).

Division of the DOJ has investigated a number of U.S. police departments, starting with Pittsburgh in 1997.¹⁷⁷ In the usual course of events, the Special Litigation Section conducts an in-depth investigation which often includes search and seizure practices, use of force, training, and policies and practices concerning the use of police custody, restraints, and particular weapons.¹⁷⁸ If, upon completion of the investigation, the Special Litigation Section finds that the agency has engaged in a pattern or practice of violating constitutional rights, it proposes reforms to the agency; these reforms often become the basis for a court-supervised consent decree.¹⁷⁹ A number of these consent decrees have required that police departments collect stop-and-frisk data, including those in effect in Los Angeles¹⁸⁰ and New Orleans.¹⁸¹

Recently, the DOJ reached a settlement agreement with the City of Cleveland over various aspect of police conduct, including use of force and search and seizure practices.¹⁸² The agreement mandates the collection and analysis of data on a variety of police conduct that the DOJ investigated, and specifically requires the collection of specified data on all stops and searches by police, including stops and frisks of pedestrians.¹⁸³ And, in the closing days of the Obama Administration, the DOJ entered into agreements with the Baltimore Police Department (BPD) that required the BPD to collect and maintain data on all stops, searches, and arrests to ensure that officers' enforcement activities promote public safety and are consistent with constitutional and legal standards.¹⁸⁴ The DOJ attempted to reach a similar agreement with the City of Chicago.¹⁸⁵

¹⁷⁷ Consent Decree at 1–2, *United States v. City of Pittsburgh*, No. 97-0354 (W.D. Pa. Feb. 26, 1997).

¹⁷⁸ The best description of this process can be found in the Civil Rights Division's January 2017 report. CIVIL RIGHTS DIV., U.S. DEP'T OF JUSTICE, THE CIVIL RIGHTS DIVISION'S PATTERN AND PRACTICE POLICE REFORM WORK: 1994-PRESENT 9 (Jan. 2017).

¹⁷⁹ *Id.* at 20–21.

¹⁸⁰ Consent Decree at 42–46, *United States v. Los Angeles*, No. 00-cv-11769 (C.D. Cal. June 15, 2001).

¹⁸¹ Consent Decree Regarding the New Orleans Police Department at 42–43, *United States v. City of New Orleans*, No. 2:12-cv-01924-SM-JCW (E.D. La. Jan. 11, 2013).

¹⁸² Settlement Agreement at 39–43, *United States v. City of Cleveland*, No. 1:15-cv-01046 (N.D. Ohio May 26, 2015).

¹⁸³ *Id.* at 60–61, 100.

¹⁸⁴ Consent Decree, *United States v. Police Dep't*, No. 1:17-cv-00099-JKB (D. Md. Jan. 12, 2017). The district court approved the Consent Decree and denied the motion of the DOJ for additional time for the Trump Administration to review this agreement. *See United States v. Balt. Police Dep't*, No. JKB-17-99, at 4, 7 (D. Md. Apr. 17, 2017) (order approving consent decree and denying governments motion to delay entry).

¹⁸⁵ *See* U.S. DEP'T OF JUSTICE, AGREEMENT IN PRINCIPLE BETWEEN THE UNITED STATES DEPARTMENT OF JUSTICE AND THE CITY OF CHICAGO REGARDING THE CHICAGO POLICE DEPARTMENT (2017), <https://www.justice.gov/opa/file/925901/download> [<https://perma.cc/8XWZ-Z3LU>]; *see also* U.S. DEP'T OF JUSTICE, INVESTIGATION OF THE CHICAGO POLICE DEPARTMENT 158–61 (Jan. 2017). In light of the failure of the city and the DOJ to follow through with adopting a Consent Decree, a lawsuit by persons alleged to have been harmed

Of course, there is never a pure linear progression or assurance of continued best practices. And as the Trump Administration took control of the DOJ in 2017, there is good reason to believe that much of the work of the Civil Rights Division with regard to local law enforcement agencies will be reduced or terminated.¹⁸⁶

D. *Litigation and Studies*

Over the past twenty years, a number of cases and studies have employed statistical analysis to assess the legality, racial distribution, and efficacy of stop-and-frisk practices, with significant debate about the proper “benchmarks” and standards to be applied on all of these issues. In this Part, we summarize a number of the leading studies and court cases and then turn our attention to the largely overlooked significance of hit-rate data in the assessment of the predictive value of many of the generally accepted grounds for stop-and-frisk interventions.

1. *New Jersey Turnpike and Racial Profiling*

In one of the first cases presenting the issue of racial disparities in police stop practices, *State v. Soto*, the court found that car stops and searches on the New Jersey Turnpike disproportionately targeted African Americans.¹⁸⁷ Relying on traffic and violator surveys and a large set of data regarding traffic stops on the Turnpike, the court determined that approximately 14% of all cars on the Turnpike had at least one black occupant, that virtually all drivers violated the traffic laws (in particular for speeding), and that blacks and whites violated traffic laws at almost exactly the same rate.¹⁸⁸ Yet 46.2% of stops were of black motorists, representing a statistically significant disparity of 16.35 standard deviations.¹⁸⁹ Further, the court found that while radar stops made by the radar unit were relatively consistent with the percentage of black violators,

by past practices and policies has been filed. Civil Complaint at 1–2, *Campbell v. City of Chicago*, No. 17-cv-02984 (N.D. Ill. Apr. 20, 2017).

¹⁸⁶ See, e.g., Eric Lichtblau, *Sessions Indicates Justice Department Will Stop Monitoring Troubled Police Agencies*, N.Y. TIMES (Feb. 28, 2017), <https://www.nytimes.com/2017/02/28/us/politics/jeff-sessions-crime.html> [<https://perma.cc/3ZW6-VPV6>]; Jon Swaine, *Jeff Sessions Could Weaken Police Reforms on Excessive Force and Racial Bias*, GUARDIAN (Jan. 10, 2017), <https://www.theguardian.com/us-news/2017/jan/10/jeff-sessions-police-reform-excessive-force-racial-bias-justice-department> [<https://perma.cc/TEV7-2Z4H>] (reporting Sessions would not rule out changes to existing consent decrees and stated that consent decrees and other DOJ actions had “unfairly maligned” police.).

¹⁸⁷ *State v. Soto*, 734 A.2d 350, 360 (N.J. Super. Ct. Law Div. 1996) (noting that the statistical disparities between African-American and white motorists stopped for traffic offenses were “stark”).

¹⁸⁸ *Id.* at 352–53.

¹⁸⁹ *Id.* at 353.

discretionary stops made by the patrol division—a unit involved in drug interdiction—resulted in double the percentage of blacks stopped.¹⁹⁰

A study conducted by Peter Verniero, Attorney General of New Jersey, expanded upon these findings.¹⁹¹ Verniero determined that searches of cars on the Turnpike were even more racially disparate than the initial stops: 77.2% of all “consent” searches were of Hispanics and blacks.¹⁹² Verniero concluded that the use of arrest statistics could not justify racially disparate stops and searches, as those arrests were the product of racially discriminatory law enforcement practices.¹⁹³

2. New York City Stop-and-Frisk Practices

In 1999, New York Attorney General Eliot Spitzer, conducted a study of 175,000 pedestrian stops by the NYPD over a fifteen-month period in 1998–1999 and found a highly disproportionate rate of stops of minorities.¹⁹⁴ Spitzer

¹⁹⁰ *Id.* at 354, 356.

¹⁹¹ PETER VERNIERO, ATTORNEY GEN. OF N.J., INTERIM REPORT OF THE STATE POLICE REVIEW TEAM REGARDING ALLEGATIONS OF RACIAL PROFILING 1–2 (Apr. 1999).

¹⁹² *Id.* at 26–28. Seizures of contraband or arrests following these stops on the Turnpike were made at a rate of 10.5% from white drivers and 13.5% from African-American drivers. *Id.* at 28.

¹⁹³ *Id.* at 67–68. There is a broad range of studies and litigation on the issue of the racial distribution of traffic stops. *E.g.*, AM. CIVIL LIBERTIES UNION OF ARIZ., DRIVING WHILE BLACK OR BROWN: AN ANALYSIS OF RACIAL PROFILING IN ARIZONA 3 (Apr. 2008) (finding that Native Americans were 3.25 times more likely, and African Americans and Hispanics 2.5 times more likely to be subjected to searches than whites after being stopped); AMY FARRELL ET AL., NE. UNIV. INST. ON RACE & JUSTICE, RHODE ISLAND TRAFFIC STOP STATISTICS ACT FINAL REPORT 1 (June 2003) (“In most communities in Rhode Island non-white drivers are stopped disproportionately to their presence in the driving population.”); STEPHEN M. HAAS ET AL., DIV. OF CRIMINAL JUSTICE SERVS., WEST VIRGINIA TRAFFIC STOP STUDY: FINAL REPORT, at i, 7–9 (Feb. 2009) (“State-level results indicate that black drivers are 1.64 times more likely . . . [and] Hispanics were 1.48 times more likely to be stopped compared to white drivers.”); ALEXANDER WEISS & DENNIS P. ROSENBAUM, THE UNIV. OF ILL. AT CHI. CTR. FOR RESEARCH IN LAW & JUSTICE, ILLINOIS TRAFFIC STOPS STATISTICS ACT 2010 ANNUAL REPORT: EXECUTIVE SUMMARY 10 (July 2011) (“[V]ehicles driven by African-Americans and Hispanics are twice as likely to be the subject of a consent search tha[n] those driven by Caucasians.”); *see also* Rodriguez v. Cal. Highway Patrol, 89 F. Supp. 2d 1131, 1141 (N.D. Cal. 2000); Md. State Conference of NAACP Branches v. Md. Dep’t of State Police, 72 F. Supp. 2d 560, 563–64 (D. Md. 1999). Dr. John Lamberth employed regression analysis and found racial disparities in traffic stops in Kalamazoo, Michigan. JOHN C. LAMBERTH, TRAFFIC STOP DATA ANALYSIS PROJECT 29 (Sept. 2013). Traffic stops present some different issues than pedestrian stops flowing primarily from the fact that almost all drivers violate traffic laws at the same rate, thus permitting pretextual stops. *See* Whren v. United States, 517 U.S. 806, 818–19 (1996).

¹⁹⁴ N.Y. STATE ATTORNEY GEN., REPORT TO THE PEOPLE OF THE STATE OF NEW YORK FROM THE OFFICE OF THE ATTORNEY GENERAL 88–89 (Dec. 1999) [hereinafter REPORT], https://ag.ny.gov/sites/default/files/pdfs/bureaus/civil_rights/stp_frsk.pdf [<https://perma.cc/7CGK-DPXU>].

determined that (1) African Americans were stopped six times more frequently than whites;¹⁹⁵ (2) in precincts where African American constituted 10% or less of the population, stops of African Americans constituted 30% of all stops, more than ten times their percentage of the population;¹⁹⁶ (3) stops of African Americans were less likely to result in arrests than stops of whites;¹⁹⁷ and (4) adjusting for crime rates by race, the differences in stops of minorities compared to stops of whites was statistically significant, with African Americans stopped more than twice as often as whites for suspected violent crimes and weapons offenses.¹⁹⁸ Spitzer also reported that where the police provided a full factual statement concerning the stop, 15.4% of the stops failed to comply with Fourth Amendment standards.¹⁹⁹ In addition, 23.5% of the stops failed to provide a sufficient factual basis to determine whether the stop was constitutionally proper.²⁰⁰ However, these findings were largely overshadowed by the racial analysis of the stops.

*Floyd v. City of New York*²⁰¹ was a lawsuit built on the foundations of the N.Y. Attorney General investigation of 1999 and on data on stops and frisks in New York City for the period 2004–2012.²⁰² During that time period, the NYPD made 4.4 million pedestrian stops, of which over 80% were of African Americans or Latinos.²⁰³ More than half of those stopped were also subjected to a frisk.²⁰⁴ *Floyd* presented a challenge to these stop-and-frisk practices on Fourth and Fourteenth Amendment grounds.²⁰⁵

Crediting a number of statistical methodologies, including regression analysis, that had been presented by the plaintiffs' expert, Professor Jeffrey Fagan, the court determined that the stop-and-frisk program in New York City was marked by intentional race discrimination in violation of the Fourteenth Amendment.²⁰⁶ Professor Fagan determined that the best predictor of stops was the racial composition of the precinct; that blacks and Hispanics were more likely to be stopped (even in areas with predominantly white populations); that blacks and Hispanics were 30% more likely to be arrested (as opposed to receiving a summons) after being stopped, and 14% more likely to be subjected to the use of force during the stop; and that the hit rate for blacks and Hispanics (as measured by recovery of contraband, arrests made, or summonses issued

¹⁹⁵ *Id.* at 95.

¹⁹⁶ *Id.* at 106.

¹⁹⁷ *Id.* at 93.

¹⁹⁸ *Id.* at 126.

¹⁹⁹ *See id.* at xiv.

²⁰⁰ REPORT, *supra* note 194, at xiv.

²⁰¹ *Floyd v. City of New York*, 959 F. Supp. 2d 540 (S.D.N.Y. 2013).

²⁰² *Id.* at 560–61.

²⁰³ *Id.* at 556.

²⁰⁴ *Id.*

²⁰⁵ *Id.* at 557.

²⁰⁶ *Id.* at 559–60, 572–75.

following a stop and/or frisk) was 8% lower for blacks and Hispanics than for whites.²⁰⁷

On the Fourth Amendment issue, based on an analysis of the reasons provided on the NYPD stop forms for stops and frisks, the court ruled that a significant number of stops were made without the requisite reasonable suspicion.²⁰⁸ Further, the court determined that certain factors frequently cited by police for their interventions were poor indicators of crime.²⁰⁹ Thus, “Furtive movements” and/or presence in “High Crime Areas” were each cited in over 40% of the stops, yet they rarely produced any evidence of criminal conduct.²¹⁰ The court further found that “52% of all stops were followed by a protective frisk for weapons,” but that a weapon was found in only 1.5% of these frisks and that “6% of all stops resulted in an arrest, and 6% resulted in a summons.”²¹¹ “The remaining 88% of the 4.4 million stops resulted in no further law enforcement action.”²¹²

²⁰⁷ *Floyd*, 959 F. Supp. 2d at 558, 589. Beyond the statistical evidence, the court found evidence of intentional discrimination through an examination of “institutional evidence,” and specifically the deliberate indifference of the NYPD to patterns of racial discrimination in stop-and-frisk practices. *Id.* at 590. Patterns of race discrimination were known to the NYPD as early as 1999 when the State Attorney General issued a report on stop-and-frisk practices that documented unexplained racial disparities in stops. *Id.* at 560, 590–91. Further, the NYPD put great pressure on commanders and others in the chain of command (down to patrol officers) to increase the number of stops (from 97,000 in 2002 to 686,000 in 2011), but that the NYPD failed to audit the stops in a manner that would examine possible racial discrimination. *Id.* at 591–96. And there was evidence that officers were encouraged to make stops based on racial characteristics or stereotypes and direct orders to the target the “right people,” which was explained as blacks and Hispanics since it was young men of color who were committing violent crimes most often. *Id.* at 603–04.

Similar findings regarding racial bias have been made with respect to other major police departments. For Los Angeles, see IAN AYRES & JONATHAN BOROWSKY, A STUDY OF RACIALLY DISPARATE OUTCOMES IN THE LOS ANGELES POLICE DEPARTMENT 27 (Oct. 2008). The authors stated that “[i]t is implausible that the higher frisk and search rates were justified by higher minority criminality . . . when these frisks and searches were less likely to uncover . . . contraband.” *Id.*; see also CHRISTOPHER STONE ET AL., POLICING LOS ANGELES UNDER A CONSENT DECREE: THE DYNAMICS OF CHANGE AT THE LAPD 22–24 (May 2009). For Boston, see ACLU OF MASS., BLACK, BROWN AND TARGETED: A REPORT ON BOSTON POLICE DEPARTMENT STREET ENCOUNTERS FROM 2007–2010, at 1–2 (Oct. 2014). For Chicago, see ACLU FOUND. OF ILL., STOP AND FRISK IN CHICAGO 1–2 (May 2015), and see also ARLANDER KEYS, THE CONSULTANT’S FIRST SEMI-ANNUAL REPORT ON THE INVESTIGATORY STOP AND PREVENTATIVE PAT DOWN AGREEMENT FOR THE PERIOD JANUARY 1, 2016–JUNE 30, 2016, at 139–40 (Mar. 2017), and ARLANDER KEYS, THE CONSULTANT’S SECOND SEMI-ANNUAL REPORT: INVESTIGATORY STOP & PROTECTIVE PAT DOWN AGREEMENT 9–14 (Mar. 2018).

²⁰⁸ *Floyd*, 959 F. Supp. 2d at 559.

²⁰⁹ *Id.* at 579–83.

²¹⁰ *Id.* at 581–83.

²¹¹ *Id.* at 573.

²¹² *Id.*

The low hit rates in New York City (and as we discuss, *infra*, in other departments as well) present a host of issues regarding the reasonable suspicion legal standard.²¹³ In particular, the low rate of gun seizure raised serious questions regarding the validity of the factors upon which police rely in frisking for guns. In 2011, the NYPD made 524,873 more stops than in 2003, an increase of over 300%, but recovered only 176 more guns, an incremental recovery rate of 0.03%.²¹⁴ And in 2012, the police conducted approximately 297,000 frisks (56% of all stops) and weapons were found in only 2% of these cases.²¹⁵ Blacks and Latinos were more likely to be frisked than whites, even though whites were much more likely to be found in possession of weapons.²¹⁶

Moreover, with the sharp drop in the number of stops in New York City following *Floyd* and the election of Mayor DeBlasio, it is possible to conduct at least a preliminary comparative analysis of the relationship between and among stop-and-frisk practices, crime control, and constitutional policing. Notwithstanding the dispute as to the size of the drop in stops due to questions that police may not have been reporting some stops,²¹⁷ the reduction from the nearly 700,000 stops in 2011,²¹⁸ is dramatic.²¹⁹ And, with far fewer stops, the hit rates have increased and violent crime has continued to decline.²²⁰

²¹³ See Kinports, *supra* note 50, at 64; Craig S. Lerner, *Reasonable Suspicion and Mere Hunches*, 59 VAND. L. REV. 407, 414–17 (2006); Minzner, *supra* note 50, at 920–22; Christopher Slobogin, *The Exclusionary Rule: Is It on Its Way out? Should It Be?*, 10 OHIO ST. J. CRIM. L. 341, 352–53 (2013).

²¹⁴ N.Y. CIVIL LIBERTIES UNION, STOP-AND-FRISK 2011, at 2, 6–12 (May 2012); see also *United States v. McCrae*, No. 07-CR-772(JG), 2008 WL 115383, *1 (E.D.N.Y. Jan. 11, 2008) (fifty stops made on “telltale signs” of possession of guns as reported by officer making stops, resulted in a single gun seizure). A study by the New York Attorney General on the outcome of the cases in which arrests were made after a stop in New York City for the years 2009–2012 (6% of the total stops) showed that close to a half did not result in a conviction, fewer than one in four (or about 1.5% of all stops) “resulted in a jail or prison sentence,” one in fifty (or about 0.1% of all stops) resulted in a conviction of a crime of violence or for possession of a weapon, and close to one-quarter of these arrests were dismissed before arraignment or downgraded to an infraction or violation. N.Y. STATE OFFICE OF THE ATTORNEY GEN., A REPORT ON ARRESTS ARISING FROM THE NEW YORK CITY POLICE DEPARTMENT’S STOP-AND-FRISK PRACTICES 1 (Nov. 2013). In cases with random stops, the rate of seizures of guns and drugs has been appreciably higher. See, e.g., *City of Indianapolis v. Edmond*, 531 U.S. 32, 35 (2000) (noting that of 1,161 random road-block stops, 4.7% led to arrests for drug-related offenses).

²¹⁵ CHRISTOPHER DUNN, N.Y. CIVIL LIBERTIES UNION, STOP-AND-FRISK 2012, NYCLU BRIEFING 8 (May 2013).

²¹⁶ *Id.* at 10 (“[A] weapon was found in only 1.8% of blacks and Latinos frisked, as compared to a weapon being found in 3.9% of whites frisked.”).

²¹⁷ See Seventh Report of the Independent Monitor at 38, *Floyd v. City of New York*, 959 F. Supp. 2d 540 (S.D.N.Y. Dec. 13, 2017).

²¹⁸ N.Y.C.L. UNION, *supra* note 93 (documenting 685,724 stops in 2011).

²¹⁹ See Sixth Report of the Independent Monitor at 3 n.1, *Floyd v. City of New York*, 959 F. Supp. 2d 540 (S.D.N.Y. June 29, 2017).

²²⁰ N.Y.C.L. UNION, *supra* note 93.

3. Philadelphia Stop-and-Frisk Practices

In 2011, in a class action lawsuit filed against the City of Philadelphia, alleging Fourth and Fourteenth Amendment violations in stop-and-frisk practices,²²¹ the parties agreed to a consent decree that required the City to conduct stops and frisks within constitutional limits, prohibited stops and frisks without reasonable suspicion (specifying certain conduct that did not establish reasonable suspicion, such as “loitering,” presence in “high crime areas,” acting “suspiciously,” and making “furtive movements”), prohibited the use of race as a basis for a stop except in cases of suspect identifications by race, mandated the creation of an electronic database of all stops and frisks with relevant information as to each stop, provided for the appointment of an independent monitor, and established a monitoring and auditing process under which plaintiffs’ counsel and the monitor would have access to all relevant data and information.²²²

The long-standing problems of stops and frisks without reasonable suspicion and large racial disparities in stop practices have proven to be slow to remedy, even with court oversight. Thus, an analysis of stops and frisks in the first half of 2015 showed that approximately 33% of all stops were made without the requisite reasonable suspicion and 56% of all frisks were made without reasonable suspicion or were the “fruit” of an illegal stop.²²³ Indeed, the data suggests that the unreasonable frisk rate may even be higher, as only 13.6% of stops resulted in a reported frisk and of “159 stops in which guns or gun-related activity [were] referenced as a basis for the stop, there were no frisks recorded on 55 stops, or 35% of the total.”²²⁴ Following the implementation of new internal accountability measures, some progress was seen in 2016. As documented in Plaintiffs’ Eighth Report to the Court, the percentage of stops without reasonable suspicion was reduced to 21%, and frisks without reasonable suspicion occurred in 41% of the incidents reviewed.²²⁵

On the Fourteenth Amendment racial distribution issue, the 2015 data showed patterns of racial bias. For each police service area (subunits of police districts) there were an average of 1,251 stops of black pedestrians, 375 of white pedestrians, and 136 of Latino pedestrians.²²⁶ The stop rate by race per 10,000 residents was 1,611 for blacks, 747 for whites, and 583 for Latinos.²²⁷ A regression analysis that controlled for possible nonracial causative factors, including demographic makeup and crimes rates by district, showed that

²²¹ Philadelphia Consent Decree, *supra* note 154, at 1.

²²² *Id.* at 3–5.

²²³ Plaintiffs’ Sixth Report to Court, *supra* note 30, at 31.

²²⁴ *Id.* at 18.

²²⁵ Plaintiffs’ Eighth Report to Court and Monitor on Stop and Frisk Practices: Fourth Amendment Issues at 18, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. Dec. 7, 2017) [hereinafter Plaintiffs’ Eighth Report to Court].

²²⁶ Plaintiffs’ Sixth Report to Court, *supra* note 30, at 24.

²²⁷ *Id.*

nonracial factors did not explain the deep racial disparities in stop-and-frisk practices.²²⁸ Further, of all frisks, 79% were of blacks, 11% were of whites, and 10% were of Latinos, and 1 in 6.4 stops of blacks resulted in a frisk while the rate for whites was 1 for every 15.2 stops.²²⁹ There was also significant variation by race in the share of stops lacking reasonable suspicion, which ranged from 31% each for whites and Latinos to 35% for blacks.²³⁰ For frisks without reasonable suspicion, the rates were 47% for whites, 57% for blacks, and 62% for Latinos.²³¹ The 2017 data show some limited improvement with respect to racial disparities (driven in part by the reduction in the overall number of stops and in the number of stops without reasonable suspicion).²³²

V. PREDICTIVE ANALYTICS IN STOP-AND-FRISK PRACTICES

In the ongoing debate over the most appropriate Fourth and Fourteenth Amendment standards for assessing stops and frisks, relatively little attention has been paid to the significance of stop-and-frisk data on the predictive value of the kinds of conduct and other factors that have been recognized by courts as sufficient to establish reasonable suspicion of criminal conduct.

We believe that data analysis provides a highly reliable basis for empirical evaluation of the factors that courts have identified as relevant to the issue of reasonable suspicion. As discussed above, the Supreme Court has resorted to “common sense” and “reason” in determining the inferential power of these factors.²³³ But even if there seems to be a basis to believe that certain types of

²²⁸ *Id.* at 27–29.

²²⁹ *Id.* at 25.

²³⁰ *Id.* at 30–31.

²³¹ *Id.* at 31. Among other reports on stop-and-frisk practices is a 2014 Department of Justice investigation of the Newark Police Department (NPD). The DOJ reported that for the period from January 2009 through June 2012, there were approximately 39,000 pedestrian stops documented by the NPD. DEP’T OF JUSTICE, INVESTIGATION OF THE NEWARK POLICE DEPARTMENT 8 (July 2014). Of these stops, 6,200 (15.8%) had no reason provided by the officer, and a study of a random sample of the rest of the stops showed that 75% were without legal justification based on the officer’s stated reasons for the stop. *Id.* at 8–9. Thus, 93% of the stops appeared to be in violation of the Fourth Amendment. *Id.* at 9 n.7. The population of Newark is 54% black, but they comprised 80% of all stops or a stop rate of 2.5 times higher than whites. *Id.* at 16. Further, when stopped, blacks were 2.7 times more likely to be searched and 3.1 times more likely to be frisked than whites, yet the hit rates for these searches and frisks were no greater for blacks than for whites. *Id.* at 20.

²³² See Plaintiffs’ Eighth Report to Court, *supra* note 225, at 4.

²³³ See *supra* note 33 and accompanying text; see also *Ornelas v. United States*, 517 U.S. 690, 700 (1996) (taking into account the fact that the arresting officer had searched approximately 2000 cars for narcotics in determining whether his search satisfied the Fourth Amendment, but without reference to his hit rate); *Illinois v. Gates*, 462 U.S. 213, 244 (1983) (“The corroboration of [a] letter’s predictions . . . suffice[d] for the practical, common-sense judgment called for in making a probable-cause determination.”); *United States v. Cortez*, 449 U.S. 411, 418 (1981) (“[A] trained officer draws inferences and makes deductions . . . that might well elude an untrained person.”).

behavior and observations provide a rational ground for reasonable suspicion, where empirical evidence does not support the unexamined “common sense,” courts should carefully review the predictive quality of these factors. This is particularly the case where categorical judgments for assessing Fourth Amendment individualized suspicion standards are based on conduct that can often be entirely innocent in nature.²³⁴

The same is true for the Supreme Court’s view that police experience leads to a special expertise regarding judgments as to reasonable suspicion and probable cause. In *Terry* and many other cases, the Court has stressed the “experience” factor and has given great deference to officers who testify that their experience informed their decision-making.²³⁵ To be sure, experience can provide insight and special knowledge, but where data demonstrates that the professed expertise does not yield the expected results, there is no basis for broad deference. Indeed, the Court’s earlier insight into the dangers of deferring to officers “engaged in the often competitive enterprise of ferreting out crime”²³⁶ may provide a more accurate benchmark for analysis.²³⁷

These empirical judgments will require comprehensive and reliable data. As discussed above, there is an expanding universe of stop-and-frisk data, particularly in large urban police departments, and as demands for transparency and accountability of policing agencies grow, it is likely that the data base will continue to expand. These data have provided significant insight on the issues of whether police have the requisite reasonable suspicion to support stops and frisks and whether racially disproportionate stop-and-frisk practices are explainable on grounds other than race—in other words, on the critical issue of intentional race discrimination barred by the Equal Protection Clause.²³⁸

Beyond the use of data to determine overall compliance of stop-and-frisk programs with constitutional standards, some courts have recognized the relevance of empirical evidence in providing a more particularized Fourth Amendment analysis of the reasons provided for stops and frisks. Thus, the Supreme Judicial Court of Massachusetts ruled that flight in a high-crime area in Boston was not sufficient grounds for a *Terry* stop because that stop data in Boston showed substantial and intentional racial disparities in stop-and-frisk rates of black men in that city.²³⁹ As the court stated:

²³⁴ See David A. Harris, *Particularized Suspicion, Categorical Judgments: Supreme Court Rhetoric Versus Lower Court Reality Under Terry v. Ohio*, 72 ST. JOHN’S L. REV. 975, 976, 989 (1998).

²³⁵ *Terry v. Ohio*, 392 U.S. 1, 27 (1968); see also, e.g., *Brown v. Texas*, 443 U.S. 47, 52 n.2 (1979).

²³⁶ *Johnson v. United States*, 333 U.S. 10, 14 (1948).

²³⁷ For a comprehensive analysis of the police expertise issue, see Lvovsky, *supra* note 90, at 2026, and see also Josh Bowers, *Annoy No Cop*, 166 U. PA. L. REV. 129, 204–05 (2017).

²³⁸ See *Floyd v. City of New York*, 959 F. Supp. 2d 540, 570–72 (S.D.N.Y. 2013); see also AYRES & BOROWSKY, *supra* note 207, at 27–28; LAMBERTH, *supra* note 193, at 7–9.

²³⁹ *Commonwealth v. Warren*, 58 N.E.3d 333, 342–43 (Mass. 2016); see also *People v. Horton*, 78 N.E.3d 489, 504 (Ill. App. Ct. 2017) (adopting the analysis of the *Warren* court

[T]he finding that black males in Boston are disproportionately and repeatedly targeted for [stop-and-frisk] encounters suggests a reason for flight totally unrelated to consciousness of guilt. . . . Given this reality for black males in the city of Boston, a judge should, in appropriate cases, consider the report’s findings in weighing flight as a factor [of the reasonable suspicion issue].²⁴⁰

In another setting, the Seventh Circuit Court of Appeals determined that the fact that a car had been “re-painted” does not provide reasonable suspicion that it had been stolen.²⁴¹ The court noted that because the State provided no empirical data to support its contention that re-painted cars are likely stolen vehicles, the color discrepancy between the car and vehicle registration was insufficient to show probable cause.²⁴² In large part, however, courts without access to empirical data accept subjective judgments regarding behavioral patterns to find reasonable suspicion.²⁴³

Hit-rate data has the power of informing the inquiry into the predictive value a number of the factors that courts have credited as relevant on the issue of whether a stop and/or frisk was supported by reasonable suspicion.²⁴⁴ Hit rates are available with respect to stops and separately with respect to frisks, but for the reasons that follow we focus on the frisk data as the most reliable. The effectiveness of stops can be measured in part by the percentage of cases in which police seize contraband, make arrests, or issue citations or summonses. And using these metrics, the hit rates in various cities is low as an absolute numerical matter. Thus, in New York City for the years 2004–2012, during which the police made over four million stops, they effectuated arrests or issued citations in approximately 12% of the stops.²⁴⁵ Similar results are shown in Philadelphia, where in the first half of 2014 approximately 7.5% of the stops

in Chicago), *appeal denied and judgment vacated*, No. 122461, 2017 WL 5635931 (Ill. Nov. 22, 2017).

²⁴⁰ *Warren*, 58 N.E.3d at 342.

²⁴¹ *United States v. Uribe*, 709 F.3d 646, 652 (7th Cir. 2013).

²⁴² *Id.*

²⁴³ *See, e.g., United States v. Arvizu*, 534 U.S. 266, 277 (2002); *Florida v. Royer*, 460 U.S. 491, 494, 502 (1983) (plurality opinion) (counting nervousness as a factor contributing to a finding of reasonable suspicion); *United States v. Briggs*, 720 F.3d 1281, 1286, 1292–93 (10th Cir. 2013) (holding that a suspect’s evasive or erratic movements are part of the totality of the circumstances in determining probable cause); *United States v. Logan*, 526 F. App’x 498, 503 (6th Cir. 2013) (considering presence in a high-crime area); *United States v. See*, 574 F.3d 309, 314 (6th Cir. 2009) (suggesting that seeing individuals sitting in a car for an extended period may add to reasonable suspicion); *United States v. Himmelwright*, 406 F. Supp. 889, 892–93 (S.D. Fla. 1975), *aff’d*, 551 F.2d 991 (5th Cir. 1977) (holding that a suspect’s “unusually calm” demeanor supported the finding of reasonable suspicion); *Bambauer*, *supra* note 36, at 505.

²⁴⁴ *See Fagan*, *supra* note 56, at 84 (arguing that factors that would establish probable cause under the Fourth Amendment are more likely to be predictors of criminal conduct than those just sufficient for reasonable suspicion).

²⁴⁵ *Floyd v. City of New York*, 959 F. Supp. 2d 546, 558, 575 (S.D.N.Y. 2013).

resulted in arrests or citations.²⁴⁶ Far fewer cases resulted in the seizure of weapons or other contraband.²⁴⁷

While this data suggest that the predictive power of court-validated stop-and-frisk factors is very low, the data can be misleading from two different directions. On the one hand, since the hit rate includes arrests or citations for *post-stop* conduct or new information unrelated to the stop (for example, the existence of an outstanding warrant, or criminal conduct such as an assault on the officer after the stop), a hit rate measured by whether the stop resulted in an arrest would be artificially high. On the other hand, the lack of an arrest or seizure of contraband does not necessarily prove that the stop or frisk was illegal. First, police may exercise discretion, particularly for low level offenses, simply to issue a warning and not to take any official action even where there are grounds to do so. Second, in many stops supported by reasonable suspicion, there may be no permissible sanction, as it turns out that the suspicious conduct was not criminal in nature. And third, many legitimate stops will not be expected to yield contraband or weapons (for example, quality of life stops, disturbances, curfew stops). Thus, a low hit rate for stops does not necessarily mean that the officers did not have good cause for their intervention. To be sure, low hit rates raise a number of questions about the benefits and costs of stop-and-frisk practices and ought to be considered very carefully by police departments as to the impact of stops and frisks both on crime control and on police–community relations.²⁴⁸

The same problems do not arise in the hit-rate analysis for frisks. Under the *Terry* doctrine, frisks can only be conducted where the officer has reasonable suspicion that the person they have stopped is armed and dangerous. This standard allows the officer to consider the reason for the stop, other information known about the suspect, and observations made at or after the stop that would provide grounds for a reasonable officer to believe that the person was armed and dangerous.²⁴⁹ Reasonable suspicion to stop does not automatically provide grounds for a frisk; indeed, many stops are for quality of life offenses that would not justify a frisk without more observations indicating that the person was armed.²⁵⁰ Thus, assuming that the officer asserts that the person stopped is

²⁴⁶ Plaintiffs' Fifth Report to Court and Monitor on Stop and Frisk Practices at 18, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. Feb. 24, 2015) [hereinafter Plaintiffs' Fifth Report to Court].

²⁴⁷ *Floyd*, 959 F. Supp. 2d at 558; Plaintiffs' Seventh Report to Court, *supra* note 30, at 7–9.

²⁴⁸ The issue of the “effectiveness” of stop-and-frisk practices in terms of crime reduction and police–community relations, is much disputed. *See, e.g.*, Bellin, *supra* note 31, at 1504–05; Rudovsky, *supra* note 31, at 126.

²⁴⁹ *Terry v. Ohio*, 392 U.S. 1, 24–27 (1968).

²⁵⁰ *See* 4 WAYNE R. LAFAVE, SEARCH AND SEIZURE § 9.6(a) (5th ed. 2012); *see also Terry*, 392 U.S. at 32 (Harlan, J., concurring) (“In the first place, if the frisk is justified in order to protect the officer during an encounter with a citizen, the officer must first have constitutional grounds to insist on an encounter, to make a *forcible* stop. Any person, including a policeman, is at liberty to avoid a person he considers dangerous. If and when a

armed and dangerous, the hit rate is a powerful tool in assessing the factors cited by officers in support of this belief. One would expect that truthful assertions would yield weapons on more than a very occasional basis, and, if not, there is good reason for courts to use the empirical data, as opposed to common sense, to determine which factors can be used by officers.²⁵¹

Existing data provides strong grounds for doubting the predictive value of many of the factors regularly cited by officers in support of their belief that a suspect is armed and dangerous. First, on a macro level, the hit rate for all frisks in departments that have collected comprehensive data on stop-and-frisk practices shows a hit rate that rarely exceeds 1–2%.²⁵² Thus, in Philadelphia, data for the years 2012–2016 show a frisk hit rate for weapons of approximately 1%.²⁵³ In New York City, the hit rate for frisks over the period 2004–2012 was 1.5%.²⁵⁴

On a more particularized basis, examination of the reasons that officers give for engaging in frisks can yield significant insight when viewed in relationship to hit rates. Since frisks may target weapons only, one would think that officers would rely on factors that most frequently result in the seizure of a weapon. And, no doubt, officers using “common sense” have relied on factors they believe are strongly predictive. Thus, an observation of a suspicious weapon-shaped bulge under a suspect’s clothing seems unlikely to indicate anything else. But this is where actual data—as opposed to “common sense” assumptions—are perhaps most surprising. The data show that certain factors regularly reported by police, such as observation of a “bulge,” a suspect not being cooperative, a suspect having their hands in their pockets, presence in a high-crime neighborhood, acting nervous or making furtive movements, and “flight” are poor predictors of whether one is armed and dangerous, yet the courts have regularly credited these explanations in sustaining police frisks.²⁵⁵ In almost all cases, “‘bulges’ turn out to be cell phones or wallets,” and the

policeman has a right instead to disarm such a person for his own protection, he must first have a right not to avoid him but to be in his presence.”).

²⁵¹ We stress the importance of understanding the significance of this empirical data only on the assumption that the officer has accurately reported the basis for the frisk. No doubt, there are instances of false allegations of conduct or observations that could support a frisk.

²⁵² See, e.g., Plaintiffs’ Seventh Report to Court, *supra* note 30, at 19.

²⁵³ Plaintiffs’ Seventh Report to Court, *supra* note 30, at 19; Plaintiffs’ Sixth Report to Court, *supra* note 30, at 19; Plaintiffs’ Fifth Report to Court, *supra* note 246, at 32; Plaintiffs’ Fourth Report to Court and Monitor on Stop and Frisk Practices at 17, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. Dec. 3, 2013). The hit rate is likely even less than this figure given that a number of frisks go unreported. See Plaintiffs’ Seventh Report to Court, *supra* note 30, at 19.

²⁵⁴ *Floyd v. City of New York*, 959 F. Supp. 2d 540, 615 (S.D.N.Y. 2013).

²⁵⁵ See, e.g., *Illinois v. Wardlow*, 528 U.S. 119, 124 (2000) (flight and presence in a high-crime area); *Pennsylvania v. Mimms*, 434 U.S. 106, 107 (1977) (per curiam) (bulges); see also *LAFAYE*, *supra* note 250, § 9.6(a).

other triggering factors are also very weak indicators of criminal activity.²⁵⁶ Thus, in audits conducted in 2014–2016, of 220 frisks based on a “bulge,” only one weapon was seized, a hit rate of less than 0.5%.²⁵⁷

Frisks conducted where officers reported that suspects failed to take their hands out of their pockets, were not “cooperative,” engaged in furtive movements, or were stopped in high-crime areas were similarly unproductive. In the 2016 audit, frisks based on these factors in 288 cases yielded only a single weapon.²⁵⁸ The fact that so few frisks lead to the recovery of a weapon (in the second half of 2016, 722 frisks yielded only 14 weapons)²⁵⁹ raises serious questions as to whether the police are accurately reporting what they observe and, if so, whether the grounds that the courts have regularly approved for conducting frisks are reliable indicators of weapon possession. The data show otherwise.

The data from New York City are strikingly similar.²⁶⁰ A perceptive study of the NYPD stop-and-frisk practices focused on stops that were based on reports of weapons to determine the “stop-level hit rate” (SHR) for these encounters.²⁶¹ For the period 2008–2010, there were close to 475,000 such stops, with only a small number of stops resulting in a seizure of a weapon.²⁶² The study sought to determine whether there were certain factors that were of a significantly higher predictive quality in terms of actual weapon possession. The study employed logistic regression analysis to identify the factors from the 2008–2010 stops that were most productive.²⁶³ These included the demographics of the person stopped, location of the stop, date and time of day, the “recorded reason” for the stop (e.g., furtive movements, bulge, high-crime area), officer observations, and the reliability of third-party sources.²⁶⁴ This regression showed that in 80% of the stops, there was less than a 3% chance of finding a weapon, and in 43% of the stops there was less than 1% chance of discovering a weapon.²⁶⁵ Not surprisingly, the least helpful indicators of weapon possession were amorphous factors like “furtive movements.”²⁶⁶ Moreover, the stops in which there was a very low SHR were also highly racially

²⁵⁶ See Plaintiffs’ Seventh Report to Court, *supra* note 30, at 19.

²⁵⁷ Plaintiffs’ Seventh Report to Court, *supra* note 30, at 19 (no weapons); Plaintiffs’ Sixth Report to Court, *supra* note 30, at 19–20 (no weapons); Plaintiffs’ Fifth Report to Court, *supra* note 246, at 19 (one weapon).

²⁵⁸ Plaintiffs’ Seventh Report to Court, *supra* note 30, at 19.

²⁵⁹ *Id.* at 9.

²⁶⁰ See *Floyd v. City of New York*, 959 F. Supp. 2d 540, 615 (S.D.N.Y. 2013).

²⁶¹ Goel et al., *Combatting Police Discrimination*, *supra* note 83, at 187; see also Goel et al., *Personalized Risk Assessments*, *supra* note 83, at 120.

²⁶² Goel et al., *Combatting Police Discrimination*, *supra* note 83, at 211–13.

²⁶³ *Id.* at 211–12.

²⁶⁴ *Id.* at 212.

²⁶⁵ *Id.* at 188, 214, 218.

²⁶⁶ *Id.* at 188.

disparate: 49% of these stops were of blacks, 34% of Latinos, and 19% of whites.²⁶⁷

Reliance on more reliable factors significantly increased the rate of recovery: 66% of all stops were responsible for 90% of the weapon seizures, and 8% of all stops were responsible for half of these seizures.²⁶⁸ Using this data, the authors of the study created a model to estimate the *ex ante* likelihood that the stops in 2011 and 2012 would turn up a weapon.²⁶⁹ The results strongly supported the authors' thesis: the factors that were present in cases in the 2008–2010 data base that more often led to a weapon seizure were predictive of a similar SHR in 2011–2012.²⁷⁰ For example, the factors in 7,310 cases in the base study that led to a 5% SHR led to the same 5% SHR for the later years.²⁷¹ The point, of course, was that data analysis could lead the NYPD (and other departments) to implement better stop-and-frisk practices. But regardless of whether police departments will use data to enhance their programs (and there is good reason to be skeptical on this point), our view is that courts must engage in this analysis to properly define the boundaries of *Terry* stops. Surely, if empirical evidence shows that certain factors that the courts have previously credited as “common sense” grounds for believing one is armed and dangerous have no or very minimal predictive value, reconsideration of their validity is in order.²⁷²

It may be the case that, even with very low hit rates, certain conduct (for example, suspicion of violent criminal conduct or of weapon possession based on credible sources) should continue to be a basis for a frisk, but with respect to more circumstantial evidence such as bulges, hands in pockets, furtive movements, and high-crime areas, common sense should yield to empirical data. As discussed, in numerous other areas, courts have updated their validation analysis given new data and new information, and there is very good reason to do so here, where so many persons are adversely affected by practices that may bear little or no relation to their purported goals.

Of course, if empirical data is considered, there is a very important normative question regarding the appropriate hit rate for weapons that would satisfy the standard of reasonable suspicion. The debate regarding quantification

²⁶⁷ *Id.* at 215.

²⁶⁸ Goel et al., *Combatting Police Discrimination*, *supra* note 83, at 219.

²⁶⁹ *Id.* at 187.

²⁷⁰ *Id.* at 212–13.

²⁷¹ *Id.* at 213.

²⁷² While we suggest that the data provide grounds for reconsidering the factors that the courts have justified as grounds for stops and frisks, there are other ways in which the data might impact Fourth Amendment doctrine and litigation. *See, e.g.*, Andrew Guthrie Ferguson, *Constitutional Culpability: Questioning the New Exclusionary Rules*, 66 FLA. L. REV. 623, 625–26, 674 n.294 (2014) (discussing that data can provide stronger arguments that an officer's impermissible actions should result in suppression under the *Herring v. United States* rationale limiting suppression to “deliberate, reckless, or grossly negligent conduct, or in some circumstances recurring or systemic negligence.” (quoting *Herring v. United States*, 555 U.S. 135, 144 (2009))).

or not for probable cause and reasonable suspicion determinations has centered on both the theoretical issue of whether such analysis is reliable under the Fourth Amendment and the more data-driven question of determining statistical minimums in assessing probable cause and reasonable suspicion issues.²⁷³

The Supreme Court has resisted quantifying the concepts of reasonable suspicion or probable cause and even where it has considered hit rate data, the Court has provided no consistent set of controlling metrics.²⁷⁴ With respect to probable cause, the Court has required a showing of a “fair probability” that the facts alleged show criminal conduct or the presence of relevant evidence²⁷⁵ but has not specified where on the spectrum of probability that showing is located.²⁷⁶ Commentators, including judges (in surveys), have suggested a range of metrics.²⁷⁷

Reasonable suspicion, by definition, requires something less in terms of probabilities. But whether that is in the 20–25% range or even as low as 5–10%, current data on frisks consistently show a far lower hit rate in many studies of no more than 1–2%, a rate we believe is so low as to be outside any reasonable

²⁷³ See, e.g., Bruce A. Antkowiak, *Saving Probable Cause*, 40 SUFFOLK L. REV. 569, 570 (2007); Ronald J. Bacigal, *Making the Right Gamble: The Odds on Probable Cause*, 74 MISS. L.J. 279, 295 (2004); Erica Goldberg, *Getting Beyond Intuition in the Probable Cause Inquiry*, 17 LEWIS & CLARK L. REV. 789, 794–95 (2013); Jon B. Gould & Stephen D. Mastroski, *Suspect Searches: Assessing Police Behavior Under the U.S. Constitution*, 3 CRIMINOLOGY & PUB. POL’Y 315, 316–17 (2004); Orin S. Kerr, *An Economic Understanding of Search and Seizure Law*, 164 U. PA. L. REV. 591, 608–10 (2016); Kerr, *supra* note 50, at 143; Kinports, *supra* note 50, at 78; Craig S. Lerner, *The Reasonableness of Probable Cause*, 81 TEX. L. REV. 951, 995–97 (2003); Meares & Harcourt, *supra* note 31, at 735; Minzner, *supra* note 50, at 951–52; Rachlinski et al., *supra* note 50, at 87; Lawrence Rosenthal, *The Crime Drop and the Fourth Amendment: Toward an Empirical Jurisprudence of Search and Seizure*, 29 N.Y.U. REV. L. & SOC. CHANGE 641, 669 (2005); Rudovsky, *supra* note 31, at 119.

²⁷⁴ This may be the result of different analysis for different Fourth Amendment contexts, and in particular with respect to the degree of intrusiveness of the police intervention. See, e.g., *Florida v. Harris*, 133 S. Ct. 1050, 1059 (2013) (permitting use of drug detecting dogs without a showing of field accuracy or error rates); *City of Indianapolis v. Edmond*, 531 U.S. 32, 35, 48 (2000) (finding unconstitutional the use of checkpoints to interdict drugs on highways, even though there was a 9% hit rate in these random stops); *Michigan Dep’t of State Police v. Sitz*, 496 U.S. 444, 455 (1990) (sustaining DUI check points where only 1.6% of drivers stopped were arrested for DUI); *United States v. Martinez-Fuerte*, 428 U.S. 543, 554, 566–67 (1976) (immigration check point was permissible with only a 0.12% hit rate (171 of 146,000 cars)).

²⁷⁵ *Illinois v. Gates*, 462 U.S. 213, 238 (1983).

²⁷⁶ See Goldberg, *supra* note 273, at 801 & n.62.

²⁷⁷ See, e.g., Minzner, *supra* note 50, at 915 (advocating use of historic hit rates in assessing individual officer allegations of probable cause); C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293, 1327–28 (1982) (citing a 1981 survey of 166 federal judges who quantified probable cause at an average of 46% (but ranging from 10–90%) and 164 federal judges who quantified reasonable suspicion at an average of 31%).

quantification of suspicion. Even if a rate as low as 5% was deemed acceptable, the current rates of 1–2% (and sometimes lower) are clearly not sufficient.²⁷⁸

All of this takes us back to our original focus. Collection, analysis, and wide use of data as a basis for public policy and legal decision-making has become more common. Empirical bases for legal and policy choices have become the norm in a growing number of areas including the criminal justice arena, as shown by the proliferation of predictive risk assessments tools for setting bail and assessing sentences.

In this era, the continued use of unexamined “common sense” assumptions to validate police practices when data can be used to empirically test factual propositions is anachronistic. Data testing of frisk hit rates in which officers use “common sense” signs of the presence of weapons to justify frisks do not support the predictive efficacy of those “common sense” indicators; something more profound than “the way we’ve always done it” is at play. Courts should not participate in this type of willful avoidance of reality.

VI. CONCLUSION

We have seen how data analysis has become a valuable tool across many aspects of American law. This makes it all the more striking that the Supreme Court and lower courts have failed to grasp the significance of the available empirical data and the role it can play regarding judicial judgments of reasonable suspicion and particularly *Terry* frisks. Hit-rate data on frisks is but one striking example. There is strong evidence that the hit rate analysis discussed here can provide police departments with critical information regarding the most salient predictors of criminal conduct and weapon possession. Properly used, this would lead to fewer civilian stops and frisks and higher rates of weapon seizures and other legitimate law enforcement actions.

Equally important, big data regarding police law enforcement practices and crime rates can provide the government and the citizenry with information integral to the most basic questions of law enforcement methods and democratic community participation.²⁷⁹ In jurisdictions in which comprehensive data is maintained, the debate about stop-and-frisk practices, crime control, and racial and procedural justice takes place in a fact-based world. The legal standards

²⁷⁸ In *United States v. Jones*, 565 U.S. 400, 412–13 (2012), the Court ruled that the use of a GPS device that tracked the defendant’s car movements for twenty-eight days, 24/7, was unconstitutional in the absence of a valid search warrant. Justice Alito wrote a concurring opinion (joined by four Justices) that supported a reasonable expectation of privacy with respect to such long range and intrusive surveillance and left open the question as to the threshold point where privacy interests prevailed. *Id.* at 429–31 (Alito, J., concurring). Presumably, surveillance for one hour or even one day might not implicate a reasonable expectation of privacy, but the difficulty in making decisions along the time spectrum should not cause a court not to act where the time frame enables the government to collect a large amount of personal information sufficient to objectively determine that there is a legitimate expectation of privacy in that context. *Id.* at 412 (majority opinion).

²⁷⁹ FRIEDMAN, *supra* note 37, at 27–28, 199–201.

employed by the courts should move the system in the same direction. Comprehensive data regarding stops and frisks, including hit rates, racial disparities, and concurrent crime rates, can yield significant information about best policing practices and the impact on crime rates and the community–police relations of stop-and-frisk practices. This empirical data should inform both policymakers and the courts in their consideration of police interventions.