

## Big Data and Criminal Justice

**Author :** Jeff Welty

**Categories :** [Uncategorized](#)

**Tagged as :** [big data](#), [statistics](#), [studies](#)

**Date :** April 22, 2019

The debate about the criminal justice system increasingly is driven by empirical studies. Phil Dixon [wrote thoughtfully last week](#) about a new analysis of 700,000 drug arrests conducted by UNC faculty members outside the School of Government. [This article](#) by a Georgia law professor is also attracting attention – it claims to be “the most substantial empirical analysis of misdemeanor case processing to date,” based on “multiple court-record datasets, covering several million cases across eight diverse jurisdictions.” Similarly, in the popular media, [this Washington Post article](#) analyzes a huge trove of data to determine the percentage of arrests in each county across the nation that are based on marijuana possession.

I could list many more examples, but the general point is one with which I suspect most readers will agree: that big data is revolutionizing the discussion of criminal justice. This transformation has been unfolding for decades. Drivers include the growth of law and economics and other law and social science approaches, which has fertilized the legal field with social science techniques, and the increasing availability of large datasets, which has made statistical analysis easier. This post offers a few thoughts about the costs and benefits of this new data-focused world.

**North Carolina may be headed further down the road of big data.** Several Republican legislators, including those with ties to law enforcement and the courts, recently filed [H 885](#). The bill would require several departments of state government to “conduct a statewide study to identify the criminal justice data elements currently collected and maintained by jails, courts, and prisons” in order to “identify gaps in data” and find “solutions for improving availability and accessibility of data to inform public policy.” If it passes, more empirical analyses will surely result.

**That’s a good thing.** Data analysis can help us break free of our biases, questionable intuitions, and preconceptions. For example, because women experience a variety of disadvantages in the workplace and in society at large, some might expect that women would also be disadvantaged in the criminal justice system. However, the statistical evidence reveals that women are treated more favorably than men at sentencing and otherwise, as noted [here](#) and [here](#).

Well-designed empirical research has illuminated or may illuminate other issues, like what has driven rising incarceration rates, why crime rates have declined, and whether prosecutors, police officers, judicial officials and other system actors exhibit implicit bias against various racial groups. This information should be welcome even when it is challenging or unflattering. Just as we expect our physicians to prescribe evidence-based treatment for our maladies, it is reasonable to diagnose and treat the criminal justice system based on the empirical facts.

**But we should be skeptical consumers of data.** Doing good empirical research is difficult, time-consuming, and sometimes expensive. So researchers may do smaller studies and hype the results. Does chocolate cure Alzheimer’s? [Probably not](#), even though a small and preliminary study suggested possible benefits. Even the larger studies typically published in leading medical journals often suffer from various flaws and limitations, [according to Harvard and UCLA researchers](#). And many studies that researchers have repeated have yielded different results, leading to the so-called “[replication crisis](#)” in the social sciences.

It can be hard for those of us without extensive backgrounds in research design and statistics to sort the good studies from the bad. That’s especially important because the statistical arms race may lead advocates for various policy

positions to throw together outcome-driven “studies” in order to tout the evidence-based nature of their preferred policies. So at a minimum, we should be cautious consumers of empirical analyses, asking questions like: Who conducted this study? How large is the sample? If it is a survey, could the wording of the questions have influenced the results? Have the results been replicated by other researchers?

**Use in the system vs. use to study the system.** This post is focused on the use of statistical and empirical evidence to study, analyze, and reform the criminal justice system. But it is worth noting that big data is also being used with increasing frequency *within* the criminal justice system. For example, police resources may be allocated based on algorithms regarding where crimes are likely to occur; judicial officials may consult risk assessment tools in setting conditions of pretrial release; and probation departments may use statistically-supported instruments to determine how to supervise offenders. All of these implementations are controversial, with critics generally arguing that using big data [does not increase the accuracy of decision-making](#) and may even [reinforce existing biases](#).

**Conclusion.** Big data is here to stay, and using data skeptically and appropriately is something that all architects of, and participants in, the criminal justice system need to learn to do. As Mark Twain famously remarked, “facts are stubborn things, but statistics are pliable.”