2006

Vitality of Voluntary Guidelines in the Wake of Blakely v. Washington: An Empirical Assessment, The Articles on Guideline Operation Issues

John F. Pfaff
Fordham University School of Law, jpfaff@law.fordham.edu

The Vitality of Voluntary Guidelines in the Wake of *Blakely v. Washington*: An Empirical Assessment

JOHN F. PFAFF*

Assistant Professor of Law, Fordham Law School

The Supreme Court's decisions in *Blakely v. Washington* and *United States v. Booker* have, at least on the surface, profoundly altered sentencing in the United States. In just fifty-seven pages, the *Blakely* Court potentially invalidated the presumptive guidelines or determinate sentencing laws used in thirteen states, which collectively incarcerate over a quarter of all the inmates in the country.[1] Though supreme courts in these states continue to wrestle with *Blakely*'s implications, most have held their systems unconstitutional, and these states likely will have to amend their guidelines.[2]

One possible remedy that has started to receive attention is to make now-impermissibly binding guidelines voluntary. But such guidelines are often perceived as having little effect. Michael Tonry, for example, argued in 1996 that with the possible exception of those used by Delaware, voluntary guidelines had had no perceptible impact on judicial behavior.[3] And following *Booker*, U.S. Attorney General Alberto Gonzales warned that making the federal guidelines voluntary would likely lead to greater sentencing disparity.[4]

The relative efficacy of presumptive guidelines over voluntary, however, appears to have never been subjected to rigorous empirical testing.[5] I recently sought to address this critical gap in the sentencing literature by testing the hypothesis that voluntary guidelines can influence judicial behavior as effectively as presumptive.[6]

In general, my findings suggest that while voluntary guidelines are not *as* effective in curbing disparity as presumptive guidelines, they are nonetheless effective. The variation in sentence lengths imposed on violent offenders, for example, appears to drop by 10 to 24 percent under voluntary guidelines, compared to a drop of 6 to 49 percent under presumptive. For the sentences imposed on those convicted of property crimes, voluntary guidelines seem to reduce variation by as much as 17 percent, and presumptive guidelines by as little as 6 percent and by as much as 45 percent. And for drug offenders, voluntary guideline adoption is correlated with declines in variation of 12 to 18 percent, and that of presumptive guidelines with declines of 17 to 54 percent.

Similarly, both voluntary and presumptive guidelines appear to reduce the extent to which race and sex explain sentence lengths. Again presumptive guidelines seem generally to outperform voluntary, but voluntary guidelines still have salutary effects. For some crimes, for example, voluntary guidelines are at least half as effective as presumptive guidelines (and in at least one instance three-fourths so). In general, however, voluntary guidelines appear to affect the use of impermissible factors less than variation.

Blakely's impact thus may not be that pronounced: states may be able to retain many of the benefits of presumptive guidelines simply by making them voluntary. This article explores these results in more depth. Section I discusses the basic empirical model and its general findings.[7] Of course, no empirical project is without its compromises and limits, and Section II considers several of these. Finally, Section III lays out some important open questions still requiring examination, especially that of *why* judges actually follow voluntary guidelines.

I. The Relative Efficacy of Voluntary Guidelines

In order to empirically test whether voluntary guidelines can viably replace presumptive, I need to first define a metric of "success." Admittedly, states adopt sentencing guidelines, both voluntary and presumptive, to accomplish numerous (and sometimes conflicting) goals. Many states, for example, hope that their guidelines will restrain incarceration growth, but a few (perhaps most famously the federal government, but also Pennsylvania and possibly Michigan) seek to restrain lenient judges instead.[8] However, almost all—if not all—adopting states hope their guidelines will accomplish at least two goals. First, they want their guidelines to reduce sentencing disparities; they want to ensure that similar offenders are sentenced similarly. And second, they want judges to pay less attention to the race and sex of the defendant. Given their universality, I focus on these goals, testing two hypotheses: that (1) voluntary guidelines reduce variation in sentence length as well as presumptive, and (2) voluntary guidelines reduce judicial attention to the race and sex of the defendant as well as presumptive.

A. The Data and the Model

To test these hypotheses, I used data from the National Corrections Reporting Program (NCRP). Started in 1983,

Federal Sentencing Reporter, Vol. 19, No. 3, pp. 202–207, ISSN 1053-9867 electronic ISSN 1533-8363
 DOI: 10.1525/fsr.2007.19.3.202.

the NCRP is a voluntary reporting program in which participating states annually provide detailed information on each prisoner admitted to their prisons. These data include demographic details such as the race and sex of the inmate, as well as data on the crimes for which the prisoner was convicted and the sentences received for them. The NCRP is not a perfect data set—most critically, as discussed in more depth below, it does not include information on the prior criminal history of the prisoner—but it is one of the few on sentencing that permits cross-state comparisons.[9]

I looked at the NCRP between the years 1989 and 2000, during which time fourteen states consistently submitted data. During these years, two of the fourteen adopted voluntary guidelines (Missouri in 1997 and Virginia in 1995[10]), and two adopted presumptive (Michigan in 1999[11] and Ohio in 1996). One of these states used voluntary guidelines throughout the whole period (Maryland), two used presumptive throughout (Tennessee[12] and Washington), and one used a determinate sentencing law (California). The remaining six did not adopt or use guidelines throughout the sample period (Alabama, Illinois,[13] Kentucky, Oklahoma, Nevada, and South Carolina).

To measure the effect of guideline adoption on sentence variation, I developed an individual-level measure of variation I call the "mean difference": it calculates the extent to which the sentence imposed on a prisoner differed from his crime's average sentence in his state in the year in which he was sentenced. In other words, if the average sentence imposed on arsonists in Virginia in 1994 was 10 years, then a defendant who received a 7-year sentence would have a "mean difference" score of 3, that is, of 10 – 7; this measure is symmetric, so an arsonist sentenced in Virginia in 1994 to 13 years would also have a mean difference score of 3 (not –3). I then regressed this score on the race and sex of the defendant and on whether the sentencing state had voluntary or presumptive guidelines in that year.[14] I ran three separate regressions, one each for violent, property, and drug crimes; due to problems in the drug data, the results for those crimes are much less reliable than those for the other two.[15]

The model examining guideline adoption's effect on the use of the defendant's race and sex is somewhat more complex. For this, I regressed the single longest sentence imposed not only on variables indicating whether the state employed guidelines in that year, but also on variables that examined how the adoption of guidelines changed the importance of race and sex.[16]

The biggest empirical problem I faced was endogeneity. Ohio, for example, did not adopt its guidelines randomly in 1996: something induced the state to act then. But what if that "something" was correlated with, say, sentence variation? Perhaps Ohio adopted guidelines because variation was getting worse, or perhaps it adopted guidelines during a period of declining variation. If, as appears to be the case, variation declined in many states prior to adoption, then my regressions may overstate the apparently beneficial effects of adoption. To control for this, I included time trends for each state; with these, the regressions measure the change in variation (or in the use of race or sex) *net* of any preexisting trend. The tables below include the results for two types of corrections, whose differences are not important for the purposes of this Article; since there is no a priori way to prefer one over the other, I provide the results for both types of corrections as well as for the uncorrected results.[17]

B. The Effect of Guideline Adoption

With that basic setup, we can turn to the specific results.

1. The Effect of Guidelines on Variation. Table 1 provides the findings for how well voluntary and presumptive guidelines reduce variation in sentencing.[18]

In words, the adoption of voluntary guidelines is correlated with a 10.282-month decline in mean difference for violent crimes in the regression that does not correct for endogeneity; in that with the linear corrections for endogeneity, with a 7.435-month decline; and so on. Two patterns emerge in Table 1. First, for most specifications voluntary and presumptive guidelines have similar effects, especially for violent and property crimes. Second, correcting for endogeneity reduces the absolute effects of both types of guidelines, but it *raises* the *relative* effect of voluntary guidelines.[19] In some cases, voluntary guidelines actually appear *more* effective in absolute terms than presumptive.

Voluntary guidelines thus seem capable of reducing variation in sentencing, in some cases almost as well, if not as well, as presumptive guidelines. It is true that in some specifications voluntary guidelines appear to have

Table 1. The Effect of Guideline Adoption on the Mean Difference in Sentence Length (Absolute Effect)

		I: Uncorrected Results	II: Correction 1	III: Correction 2
Violent crimes	Voluntary GL	–10.282***	–7.435*	–6.622
	Presumptive GL	–13.271***	–5.682**	–3.341**
Property crimes	Voluntary GL	–3.611	–1.810	–1.408
	Presumptive GL	–8.918**	–1.487	–1.346
Drug crimes	Voluntary GL	–5.632***	–2.631	–3.574***
	Presumptive GL	–14.025***	–9.898***	–5.706

Note: Coefficients marked with *** are significant at the 99 percent level; with **, at the 95 percent level; with *, at the 90 percent level. The coefficients are in months.

little effect, but generally neither do presumptive guidelines in the same specifications. To the extent that policy makers want to replicate the world that exists with presumptive guidelines, the results in Table 1 suggest that voluntary guidelines may be viable options.

2. The Effect of Guidelines on the Use of Impermissible Factors. I also examined whether voluntary and presumptive guidelines reduce judicial reliance on two "impermissible" sentencing factors: race and sex. The results here are less reliable than those in the previous section, due to limitations in the data. The NCRP does not provide any information on a defendant's prior criminal history, and extensive evidence indicates that criminal history is correlated with race.[20] Thus even a race-neutral judge will likely sentence black defendants to longer terms on average. Moreover, there is no effective way to proxy for criminal history, and without such a proxy the resulting coefficients are likely biased, though there is no way to predict the bias's direction.[21] A similar problem exists with regard to sex.

However, this is not as big a problem as it might initially appear. The focus here is not on the role *per se* of race and sex at sentencing, but rather on how that role *changes* with guideline adoption, and on how that change varies between states using voluntary and presumptive guidelines. As long as prior criminal history's importance does not change over time, its omission does not bias the measure of adoption's importance within a state, and it may not complicate comparisons across states.[22] Of course, the assumption that the history's importance is constant is surely too strong—guidelines often explicitly incorporate prior criminal history as a key factor in the sentencing process, which should alter its significance. And given the magnitudes of the effects discussed here, it appears that these tests overstate the true impact of adopting guidelines (i.e., it is unlikely that guidelines reduce the use of race or sex to the extent suggested below). It is not possible, though, to make predictions about the effect of the biases on the relative efficacy of voluntary and presumptive guidelines. These results, then, provide only a first cut at measuring the importance of presumptiveness. They are not definitive, and I hope to develop more reliable tests in the future using state-gathered data.

With that important caveat, Table 2 provides the basic findings. The first cell of Column I implies that voluntary guidelines reduce the importance of race at sentencing for violent crimes by 72 percent. In other words, if prior to guideline adoption blacks received sentences 10 months longer than whites did (controlling for all other possible explanations of sentence length), then after adoption their sentences are only 2.8 months longer. A value over 100 percent means the direction of the "bias" reverses: –106 percent in the second cell of Column I suggests that the adoption of presumptive guidelines causes black sentences, which prior to adoption were longer than white sentences, to become 6 percent *shorter*.

Except for violent crimes, the results for race are consistently statistically insignificant and numerically small, both for the underlying role of race and for the effect of adopting either type of guideline. For violent crimes, presumptive guidelines have a stronger effect than voluntary, both numerically and statistically. Presumptive guidelines also tend to have stronger effects, both numerically and statistically, on the use of sex across all three classes of crimes. While the results for presumptive guidelines are significant at at least the 90 percent level (and at the 95 percent level in all but two instances), those for voluntary guidelines are rarely significant at even the 85 percent or 80 percent level. Thus while voluntary guidelines may reduce the role of race or sex at sentencing, the reliability of these results is less than that for presumptive.

But putting aside the (important) issue of statistical significance, these results point to the same conclusions seen in the variation regressions, at least for violent crimes. While presumptive guidelines are more effective than voluntary, the latter still often manage to contribute. For property and drug crimes, however, the effects are weaker (and in two specifications appear to make the problem worse).

Table 2. The Effect of Guideline Adoption on the Use of Race and Sex at Sentencing

			I: Uncorrected Results	II: Correction 1	III: Correction 2
Violent crimes	Race	Voluntary GL	–72%	–37%	–34%
		Presumptive GL	–106%	–136%	–112%
	Sex	Voluntary GL	–11%	–36%	–14%
		Presumptive GL	–64%	–60%	–65%
Property crimes	Race	Voluntary GL	Insignificant	Insignificant	Insignificant
		Presumptive GL	Insignificant	Insignificant	Insignificant
	Sex	Voluntary GL	–97%	+54%	–16%
		Presumptive GL	–73%	–126%	–106%
Drug crimes	Race	Voluntary GL	Insignificant	Insignificant	Insignificant
		Presumptive GL	Insignificant	Insignificant	Insignificant
	Sex	Voluntary GL	+10%	–3%	–36%
		Presumptive GL	–54%	–82%	–84%

Note: Results are labeled "Insignificant" if both the underlying disparity and the effect of adoption are numerically and statistically small.

II. Qualifications and Concerns with the Results

The model above, like all econometric models, is not perfect. Some of the key limitations—such as the need to control for endogeneity and the lack of data on prior criminal history—I have already discussed. But at least two other important qualifications deserve mention.

A. Endogeneity Redux: Selecting the Type of Guideline

I used time trends above to help control for endogeneity in the timing of guideline adoption. But these trends cannot control for why a state would choose voluntary over presumptive guidelines, and this type of endogeneity may be important as well. After all, one reason why Virginia's legislature may have felt comfortable implementing voluntary guidelines is that it appoints and retains state trial judges, giving it more power than most to encourage compliance.

There is some (weak) evidence of endogeneity in the choice of guidelines. While judges in Virginia are appointed by the legislature, judges in Missouri (the other state adopting voluntary guidelines in my sample) are either elected in partisan elections or appointed by merit commissions. And results from regressions separating Virginia from Missouri suggest that voluntary guidelines generally have stronger effects in Virginia. Of course, that Virginia appears to respond more favorably to voluntary guidelines than Missouri need not imply that its method of judicial selection is the critical differentiating factor: Virginia and Missouri differ in more ways than just judicial selection policies. But the divergent outcomes do indicate that this is an important factor to consider in future research.

B. Guidelines and Prosecutorial Charging Decisions

A second limitation to the data used here is that they provide information only on the offense for which the defendant is ultimately convicted, not on the offense for which he is arrested or with which he is initially charged. The data thus cannot measure the extent to which prosecutors change their charging behavior in response to guidelines. Guidelines, for example, may induce judges to pay less attention to race within a class of crimes while simultaneously encouraging prosecutors to pay *more* attention to race when determining who gets charged with what specific offense. If this is true, then my results may overestimate the benefits of guidelines.

It is possible, however, to overstate this problem. First, Terrance Miethe studied charging behavior in Minnesota before and after it adopted guidelines and found little evidence of changing prosecutorial behavior of this sort. He points out that several formal and informal constraints prevent prosecutors from exercising their discretion too freely. Differing evidentiary requirements for various offenses and "courthouse subcultures," for example, can make it hard to substitute across crimes.[23]

Second, the coarseness of the NCRP data used here helps. The offenses listed in the NCRP are very broad: aggravated assault and kidnapping, not third-degree aggravated assault and second-degree kidnapping. And it is harder for prosecutors to shift across crimes than within the levels of a specific offense. Furthermore, at least for violent and property crimes, the data used here show little change in the distribution of offenses before and after states adopt guidelines; as discussed above, the changes are much more dramatic for drugs. Nonetheless, the significance of prosecutorial discretion in this context is also an important issue to consider further.

III. Future Issues to Consider

Following *Blakely* and *Booker*, state and federal legislators face the challenging task of reforming their sentencing regimes. The results here provide only the first steps toward a rigorous empirical understanding of how the various still-viable sentencing options operate in practice. As a result, many open questions that merit attention remain.

For example, voluntary guidelines are not the only option that states have available to them; saying that voluntary guidelines can be effective is not the same as saying they are the *most* effective alternative. As discussed extensively elsewhere,[24] states have at least two feasible reforms other than voluntary guidelines: using sentencing juries or using uncapped guidelines such as those already used in Michigan and Pennsylvania.[25] They also could return to an indeterminate system of sentencing. Nancy King and other commentators have started to empirically evaluate sentencing juries. There is perhaps less work on the efficacy of uncapped guidelines, but that arises in part because such effects are currently hard to measure: Michigan has had its system only since 2000, and Pennsylvania's Supreme Court has undermined judicial enforcement of its guidelines.[26] Whether voluntary guidelines are the best alternative, then, depends on future results concerning the other options.

Moreover, *that* voluntary guidelines appear to influence judicial behavior does not tell us *why* they do so. And the results above are consistent with at least three different theories, each with different policy implications. Do judges follow guidelines because they want to sentence consistently but fail to do so, either due to forgetfulness or unconscious bias? Then guidelines work simply by providing signposts, and it is not surprising that presumptive and voluntary guidelines may have similar effects. Or do judges follow guidelines only because the legislature can make credible threats to impose more-binding guidelines on them? In the wake of *Blakely* and *Booker* these threats may be weaker, but they remain: legislatures, for example, can expand mandatory minimums or impose sentencing juries on judges. Or do judges adhere to voluntary guidelines because their jobs or promotions depend on it, as may be the case in Virginia? In this case, effective voluntary guidelines may require substantial judicial reform. Determining why voluntary guidelines appear to work is thus a critical open issue.

IV. Some Federal Implications

In the coming term, the Supreme Court will hear two cases, *Rita v. United States* and *Claiborne v. United States*, which will determine precisely how voluntary the federal guidelines in fact are. In *Rita*, the Court faces the question of whether sentences within the guideline ranges are presumptively reasonable; in *Claiborne*, the mirroring issue of whether those outside the ranges are presumptively unreasonable. The results here shed light on some of the implications of the pair of cases.

In short, if the Court finds the guidelines wholly voluntary—if it holds that within-range sentences are not presumptively reasonable nor outside-range sentences presumptively unreasonable—then variation in sentence length should increase, and race and sex will possibly play more important roles in sentencing. But at the same time, declaring the guidelines wholly voluntary will not render them wholly irrelevant. Sentencing should still be more consistent and less mindful of race and sex than in the absence of any guidelines.

It is important to note, though, that my findings are based on state data, and extrapolating state results to the federal system should be undertaken only with some wariness. Federal judges face markedly different incentives, since they are not as responsive to political pressures as state judges. Moreover, the federal guidelines may be uniquely more punitive than the judges who apply them: federal judges may be more inclined than their state counterparts to deviate from nonbinding guidelines. It is thus difficult to predict exactly how effective wholly voluntary federal guidelines would be, but it is likely that they would continue to exert (some) meaningful influence.

***

As states and the federal government continue to reform their sentencing regimes in the wake of *Blakely* and *Booker*, one option they possess is to unbind their guidelines. This is certainly not the only possibility available—sentencing juries and uncapped guidelines being the most obvious alternatives—but my results suggest that it is one at least worthy of consideration.

Notes

* I would like to thank Dan Richman for his helpful comments on this Article.

1 The thirteen states are Alaska, Arizona, California, Colorado, Indiana, Minnesota, New Jersey, New Mexico, North Carolina, Ohio, Oregon, Tennessee, and Washington. According to data from the Bureau of Justice Statistics, *Prison and Jail Inmates at Midyear 2003*, NCJ 203947 (2004), in 2002 these thirteen states held over 28 percent of all prisoners and were responsible for over 38 percent of all new admissions; with the federal prison system included, these numbers rise to 40 percent and 44 percent, respectively.

2 So far eight states have held their guidelines unconstitutional: Arizona, Colorado, Indiana, Minnesota, New Jersey, North Carolina, Ohio, and Oregon. Furthermore, Alaska's legislature amended its guidelines to ensure compliance with *Blakely*. The supreme courts in three other states (California, New Mexico, and Tennessee) have sought to distinguish their presumptive guideline systems from those discussed in *Blakely* and *Booker*, but the Supreme Court recently held California's sentencing law unconstitutional. *See* Cunningham v. California, No. 05-6551 (2007).

3 MICHAEL TONRY, SENTENCING MATTERS 27-28 (1996).

4 Alberto Gonzales, Speech before the American Bar Ass'n House of Delegates in Chicago, Illinois, on August 8, 2005 (available online at http://www.usdoj.gov/ag/speeches/2005/080805agamericanbarassoc.htm).

5 This impression was strengthened by a claim made recently in this journal. Kim S. Hunt and Michael Connelly, *Advisory Guidelines in the Post-*Blakely *Era*, 17 FED. SENT. REP. 233, 236 (2005), noted that "[n]o controlled experiments or rigorous quasi experiments have been done isolating the effects of presumptive guidelines from other variables that may have affected systematic outcomes."

6 The results discussed in this paper come from John F. Pfaff, *The Continued Vitality of Structured Sentencing after* Blakely: *The Effectiveness of Voluntary Guidelines*, 54 UCLA L. REV. 235 (2006).

7 I do not consider any technical econometric issues here. Those interested in a detailed discussion of methodological issues should see *Id.*

8 The punitive nature of the federal guidelines is well documented. Kevin R. Reitz, *The Disassembly and Reassembly of U.S. Sentencing Practices*, in MICHAEL TONRY AND RICHARD S. FRASE, EDS., SENTENCING AND SANCTIONS IN WESTERN COUNTRIES (2001), provides a brief description of Pennsylvania's guidelines, noting that the legislature rejected an earlier version for being too lenient. That Michigan has similar guidelines to Pennsylvania's, which set floors but no ceilings, at least weakly suggests that its legislature was motivated by similar intents. And my empirical results provide some evidence that average sentences imposed in Michigan rose immediately after the state adopted its new guidelines in 1999. *See* Pfaff, *Continued Vitality*, *supra* note 6 at 293.

9 Ilyana Kuziemko & Steven Levitt, *An Empirical Analysis of Imprisoning Drug Offenders*, 88 J. PUB. ECON. 2043 (2004), provides a good description of the NCRP.

10 Virginia first imposed state-level voluntary guidelines in 1991, but it overhauled the guidelines when it adopted its truth-in-sentencing law in 1995. In particular, it severely narrowed the ranges of sentences judges were supposed to consider. Ultimately, 1995 appears to be the better year to use. This is not a minor issue: if 1991 is used as the year of adoption instead of 1995, the results for voluntary guidelines are significantly weaker. Pfaff, *Continued Vitality*, *supra* note 6.

11 Technically, Michigan used judicially created presumptive guidelines since 1984. A Michigan judge noted, however, that these guidelines covered such a small fraction of the possible crimes with which a defendant could be charged that they had little or no effect. *See* Paul L. Maloney, *The Michigan Sentencing Guidelines*, 16 T. M. COOLEY L. REV. 13, 16 (1999). In 1999, the Michigan legislature replaced the porous judicial guidelines with a comprehensive presumptive system. In the end, little turns on whether Michigan's guidelines are classified as presumptive starting in 1984 or 1999.

12 Tennessee adopted its guidelines in 1989; as discussed below, this means I do not classify it as adopting its guidelines until one year later, in 1990. However, with only one pre-adoption year, it makes sense for technical reasons not to include it as one of the "adopting in the period" states, such as Michigan or Ohio.

13 Illinois officially employs a determinate sentencing regime. However, the law establishes wide ranges of sentences, and aggravating and mitigating factors are not used by judges to change the range of available sentences, but rather appear only to help them figure out where in the set range they should sentence. It is widely accepted that Illinois' system survives *Blakely*, and as a result it makes sense to classify

Illinois as a non-guideline state despite its nominal possession of such a system.

[14] Technically, I lagged adoption by one year: Ohio, for example, adopted its guidelines in 1996, so I did not classify it as a presumptive guideline state until 1997. Many defendants sentenced in 1996 committed their relevant crimes in 1995 and thus did not face the guidelines; by 1997 few such defendants should be present. Also, for those readers familiar with panel data, I included state, year, and "offense" fixed effects (in case judges treated particular offenses in systematically different ways).

[15] Many states changed their drug prosecution practices at the same time they adopted their guidelines, making cross-year comparisons difficult.

[16] I regressed sentence length on the defendant's race and an interaction term multiplying the race indicator with the voluntary or presumptive guideline indicator. I did the same for sex.

[17] "Correction 1" incorporates a "linear correction" in the variation regression, interacting the state fixed effects with a linear year term. "Correction 2" includes this linear interaction term as well as one that interacts the state fixed effects with the square of the year term. I used analogous, but slightly more complex, terms in the impermissible factors regressions.

[18] The results in this table come regressions that consider all fourteen states. In Pfaff, *Continued Vitality*, *supra* note 6 at 263-4, Tbl. 3, I also provide results from regressions that include just the four "adopting" states. Both the absolute and relative effects of voluntary guidelines are weaker in those more-limited regressions.

[19] For example, when looking at violent crimes, voluntary guidelines' absolute effectiveness falls by nearly 36 percent, from –10.282 to –6.622, but their relative effectiveness rises from 77 percent as effective (–10.282 to –13.271) to almost twice as effective (198 percent: –6.622 to –3.341).

[20] Robert J. Sampson & Janet L. Lauritsen, *Racial and Ethnic Disparities in Crime and Criminal Justice in the United States*, *in* MICHAEL TONRY, ED., ETHNICITY, CRIME, AND INTEGRATION 311, 346-47 (1997), point out that black defendants on average have longer criminal histories than their white counterparts.

[21] As I discuss in Pfaff, *Continued Vitality*, *supra* note 6 at 273, there appear to be no state-level data sets on criminal histories that would provide even a rough proxy.

[22] In other words, the change from a correctly estimated 10-month disparity to a correctly estimated 6-month disparity is the same as a change from a biased estimate of a 15-month disparity to a biased estimate of an 11-month disparity, as long as the bias pre- and post-adoption is the same.

[23] Terrance D. Miethe, *Charging and Plea Bargaining Practices under Determinate Sentencing: An Investigation of the Hydraulic Displacement of Discretion*, 78 J. CRIM. L. & CRIMINOLOGY 155, 157 (1987).

[24] *See, e.g.*, Stephanos Bibas, Blakely's *Federal Aftermath*, 16 FED. SENT. REP. 333 (2004); Nancy J. King & Susan R. Klein, *Beyond* Blakely, 16 FED. SENT. REP. 316 (2004).

[25] As I discuss in more depth in Pfaff, *Continued Vitality*, *supra* note 6 at 290-94, Michigan (if only because its guidelines are so new) contributes little to the relatively strong performance of presumptive guidelines seen in Tables 1 and 2. The results in those tables, then, do not necessarily indicate (nor disprove) that Michigan-style guidelines are more effective than voluntary guidelines at reducing variation or the use of race and sex.

[26] Kevin R. Reitz, *Sentencing Guideline Systems and Sentencing Appeals: A Comparison of Federal and State Experiences*, 91 NW. U. L. REV. 1441, 1471 (1997).