Given the persistent racial disproportionality in inmate populations in the U.S. (and in many other countries), research on race and sentencing has received, and should continue to receive, a very high priority. Eric Baumer has given us a valuable overview of the apparent goals, accomplishments, and limitations of extant research on this topic, and the new directions which that research might take to better accomplish its goals. In particular, he wisely points to the relatively narrow set of research questions posed and methodologies used, the persistent limitations of those methods, and the need to expand research both in terms of topics and methods. In terms of topics, we surely need much more research on critical case processing steps that come before — and after — the sentencing decisions that have most often been examined. And to supplement the modal application of multiple regression models we need to apply a variety of other methodologies — each, assuredly, subject to its own inherent limitations and problems. In the classic phrase coined by Hans Zeisel, an early pioneer of Law and Social Science research, we need “triangulation of proof.”

The goals of this essay are to give further emphasis and support to some of Baumer’s observations and to suggest additional topics and methods for future race research. [Page references herein are to the manuscript circulated to panelists on June 30, 2010.] Part I elaborates on Baumer’s critique of the limits of prior research methodologies. Part II addresses and expands on Baumer’s suggestions for additional research topics beyond the modal focus on sentencing outcomes. These two subjects (methods, topics) are not independent, of course. For example, research on race effects in critical processing stages that precede sentencing is valuable not only in itself, but also as a way to avoid some of the limitation of research which examines only the sentencing stage.

I. The Goals and Methods of Prior Research

I would state the limits of prior research even more strongly than Baumer does. Some of the points raised below elaborate on issues Baumer discusses, but I also identify some additional problems. Some of the problems produce under-estimates of race effects; other problems cause over-estimates or imprecise understanding of those effects.

A. Goals of Research and Conceptualization of Race Effects

The first two apparent goals Baumer identifies (p. 4) – “detecting racial disparities” and “detecting racial discrimination” – have not been sufficiently articulated and operationalized in prior research. Baumer himself is not always clear in his use of these terms. He does not define “disparities” although at one point (pp. 8-10) he uses the term to refer to disproportionality relative to populations of each race (ratios of minority to non-minority per capita incarceration
rates). Baumer argues, and I agree, that these disproportionalities are a natural starting point for research on race and sentencing, beginning with efforts to “decompose” (explain) the causes of these disproportionalities. In doing so, however, it is important to recognize important differences in the kinds of causal factors at work. Baumer at first (p. 10) distinguishes between racial differences in offending and “racial bias in how the law is applied” or, more neutrally, “post-offense factors and decisions.” But as he later recognizes (pp. 13-15, 28), within the “post-offense” category it is important to distinguish between bias or discrimination due to racial animus and/or stereotyping, and disparate impact – the effects of seemingly legitimate, race-neutral factors (e.g., prior criminal record) which favor whites and disfavor members of racial and ethnic minority groups. Racial disparities such as those shown by per capita incarceration rates can be the product of all three of these types of factors – racial differences in offending; the disparate impact of seemingly race-neutral law enforcement, case-processing, and sentencing policies (some of which magnify the effects of racial differences in offending); and actual bias or discrimination.

Baumer makes the very important point that bias embedded in disparate impact factors (such as pretrial detention, leniency given for a guilty plea, or type of defense counsel) has often been masked by treating those factors merely as controlled variables (p. 21). But the same point can also be applied to the first category: bias may be embedded in controlled variables that measure offense severity and prior conviction record. As discussed in Part II, we need much more research on those embedded biases – and also on disparate impact per se, which many view as a form of systemic bias. Such research poses its own methodological problems, of course; some decisions (e.g., assessments of recidivism risk and “amenability” to treatment or supervision) may reflect both bias and disparate impact, and it may be difficult to untangle their separate effects.

B. Multivariate Model Specification

Apart from ignoring imbedded case-specific and systemic bias in controlled variables, studies of case-specific bias in sentencing (or indeed, at any stage of processing) suffer from a chronic omitted-variable problem which tends to exaggerate the significance of race effects. (This is one of the most important reasons why sentencing research must be expanded to include a wider variety of methodologies, some of which have fewer or at least different omitted-variable problems.) Baumer recognizes this problem (pp. 20-21), but his list of omitted variables is under-inclusive. Sentencing studies are almost always based on archival data which does not include all variables, or details on variables, which affect the outcome variable. Many of those omitted variables (e.g. aggravating or mitigating details of current and/or previously-convicted offenses; family and employment situation; gang involvement) are often correlated with race; when that is true, omitting them from the model causes them to be measured indirectly through the race variable, thus falsely attributing the effects of the omitted variable(s) to race. This is a problem even when existing archival datasets are supplemented by hand-coding of available documents, interviews of participants, and so forth; much information and detail that bears on decisions and outcomes is simply not recorded or recalled in a consistent and reliable form.
C. Controls for Nature and Severity of the Offense

A third problem, related to the first two, is that virtually all archival data is limited to conviction offense, rather than the “real offense” for which the defendant could have been found guilty (or which the sentencing judge takes as "true," even if a conviction could not have been obtained); the conviction offense may reflect race effects in earlier arrest, prosecution, plea bargaining, and trial dispositions. This problem goes beyond Baumer’s point about the importance of studying disparities at earlier stages; analysis that is limited to conviction offense may be comparing apples to oranges if cases are screened differently by race. For example if, for some race or ethnic group, charges that survive to the stage of conviction are more often bargained down, then prosecutors, pre-sentence-report writers, and judges may be influenced by what they see as the more serious “real offense,” and therefore sentence those cases more harshly than the (lesser) conviction offense would seem to warrant – and more harshly than other offenders whose “real offense” is the actual conviction offense. More frequent charge-downs for minorities might be caused by: problems of witness non-cooperation; different charging styles in jurisdictions with high minority populations; and/or prosecutorial leniency toward crimes involving minority victims (most of which are committed by minority offenders, see further discussion below). Of course, charging differences can also work in the opposite direction, at least for some crimes in some jurisdictions – non-minority offenders may more often be allowed to plead to a lesser crime. If these “deals” also involve an agreed sentence or sentence-cap consistent with the lesser crime, that would lower sentence severity for those offenders. But such charge-downs can also result in a “real-offense-sentence” effect which understates racial disparity – non-minority offenders with bargained-down conviction offenses but real-offense sentences are being compared to minority offenders whose “real offense” is the actual conviction offense.

All of this underscores the need more case “flow” data across multiple stages (pp. 10-11). At a minimum we need sentencing studies which compare conviction-offense models to alternative models incorporating estimated-real-offense variables, for example, based on arrest offense or initial charge. Of course, such estimates are inherently crude, since arrest and initial charges are often over-stated and not taken seriously by system actors involved in sentencing. And archival data will often not include arrest charge, or even initial charge, although sometimes it does. One of the only favorable things to be said about the much-criticized “relevant conduct” provision of federal sentencing law is that it leads to recording – and, unfortunately, also basing recommended sentences on – many real-offense details. In a few state-level studies, researchers have been able to hand-code real-offense data from additional archival sources such as arrest reports, initial charging documents, and pre-sentence reports (see, e.g., Frase 1993, discussing the “alleged offense” variable constructed and analyzed in early research conducted by the Minnesota Sentencing Guidelines Commission). In short – even if one believes (as I do) that sentencing guidelines and other sentencing rules should be based on conviction offense, sentencing research should whenever possible include real-offense estimates.

D. Race of the Victim

Baumer points out (pp. 16, 34) that prior research has focused on defendant race, even
though available data suggests that the race of the victim has its own serious race effects. Again, the race-of-victim effect can be due to bias, disparate impact (e.g., of witness non-cooperation), or (most likely) a combination of both; although difficult, it is important to try to untangle these two kinds of reasons why crimes against minority victims may be less severely punished.

Potential bias and disparate impacts involving minority victims are themselves important problems, but they also pose serious problems for studies of the effects of defendant race – without separate measures of victim race and its effects, disparity in the treatment of minority victims may conceal the extent of disparity in the treatment of minority offenders. Given the intra-racial nature of many crimes (people tend to commit crimes close to where they live, and tend to live near other people of their own race), bias or disparate impact in the treatment of minority victims tends to benefit minority offenders. It seems likely that when race-of-victim is separately measured and controlled, race-of-offender disparities will be found to be even greater.

E. Other Forms of Compensating Bias or Disparate Impact

The off-setting effect of victim race and defendant race, noted above, illustrates a broader problem. For example, geographic variations in sentencing severity may interact with offender race if most minority offenders live in jurisdictions with higher or lower overall sentencing severity. Unless jurisdiction-effects are separately measured via hierarchical modeling, race effects will not be accurately measured. In a state where most Blacks live in lower-sentence-severity urban districts, the lowered state-wide estimates of Black sentence severity may conceal bias or other disparities in sentencing of Blacks in suburban and rural districts. Conversely, in states where most Blacks live in higher-sentence-severity non-urban districts, elevated state-wide estimates of Black sentence severity may reflect a “rural” rather than a “race” effect.

Another compensating-biases problem, which arises in most comparisons of racial disproportionality at earlier and later stages of case processing, is discussed in section G below.

F. Limitations of Per Capita Disproportionality Measures as a Starting Point or Diagnostic

As noted in section A above, ratios of minority to non-minority per capita incarceration rates are a frequently-used overall index of racial disparity in different systems and over time. I agree, that “decomposing” these aggregate patterns “can serve as a guide to the sources” of disparity and show us “where our research efforts might best be directed” (p. 9). But in doing so, researchers should be aware of several problems with the ways per capita measures have been computed and used; if not corrected, these problems may limit the value of these measures both as a tool to focus research (e.g., on major cross-sectional or temporal differences in disparity) and as an outcome variable to be explained. As discussed more fully elsewhere (Frase 2009), problems with per capita measures include: 1) the inclusion of juveniles in each race’s population base; 2) failure to adjust for major racial differences in population growth by time-lagging population bases to earlier years when inmates committed their crimes; and 3) problems of incompatible data sources (in particular: hispanics may or may not be included in race categories; inmate “stocks” are not directly comparable to earlier case-processing “flows”).
G. Limitations of Arrest-versus-Prison Comparisons

As a first step in explaining persistently high ratios of minority to non-minority per capita incarceration rates, Baumer summarizes and replicates studies by Blumstein and others that compare racial disparities in arrests with those found in prison populations (pp. 11-12). Baumer admits that these comparisons are crude but even that may be an understatement, since they are subject to several of the problems noted above:

1) Omitted variables. Arrestees-versus-inmates comparisons omit some of the most critical sentencing factors that courts routinely consider, including: the specific offense (within the broad categories used, such as “drugs” and “other/unspecified”); the presence of aggravating or mitigating offense details (victim injury, weapon use, amount of loss or contraband, etc); and the offender’s prior conviction record. Most of these omitted variables are probably strongly correlated with race as well as sentencing outcome. Indeed, given the sentencing importance of the omitted variables it is not too surprising to find that “nearly half of the contemporary racial disparity we see in incarceration rates is unaccounted for by arrest/offense patterns” (p. 34). Perhaps there are ways to refine this approach, at least with respect to the critical prior record factor. For example, one could try to devise formulas that use available national data on prior record differences by race, and the effects these have on sentencing to “scale up” estimates of the inmate populations one would expect from a given number of arrests.

2) Conviction offense versus real offense. Offense-specific inmate population data is not directly comparable to arrest data. Besides the problem of incompatible offense classifications in data compiled by different government agencies, there may be racial differences in charging, plea bargaining, and trial adjudication which distort these comparisons. For example, an inmate might have been arrested for burglary, larceny, or auto theft, but convicted and imprisoned on the “lesser” charge of possessing stolen property.

3) Compensating biases in arrest and sentencing. For some cases, offenses, or groups of offenders, comparisons of disproportionality at the arrest and sentencing stages may be distorted by compensating biases. If, for example, police arrest four Blacks for every reported robbery by a Black suspect (the first bias), most of those suspects will have their cases dismissed, thus temporarily lowering racial disproportionality. But if the remaining Black suspects are then treated more severely than White suspects (more often retaining robbery charges for Blacks, and/or more often imposing a prison term or a longer term), disproportionality will rise again but may not be particularly elevated relative to the (inflated) disproportionality of arrests; estimates of case-processing disparities will thus be understated.

4) Additional problems. Some of the incompatible-data problems in decomposing per capita disproportionality (section F above) also apply to arrest-versus-prison comparisons. National arrest data includes Hispanics in the race categories, whereas national prison data for recent years does not. And arrest measures are flow data, whereas inmate populations are stocks.
II. New Directions for Empirical Research

I will leave to others with greater technical expertise the task of refining and expanding on Baumer’s suggestions for new research methods (though, as noted at the outset, I strongly support the concept of multiple, complementary methods – triangulation of proof). What I want to do in this section is identify some of the topics which merit further research, to supplement and illuminate the findings on the modal topic, sentencing decisions. As Baumer notes, racial disproportionality in prison and jail populations reflects many system inputs and decisions. For convenience, here is partial list (Fraxe 2009): Inmate disproportionality may result, directly or indirectly, from racial differences in: 1) the frequency, severity, or location of criminal behavior, 2) reporting of crime, 3) police decisions to investigate and arrest, 4) pretrial detention decisions, 5) victim cooperation with investigation and prosecution, 6) prosecutorial screening and initial charging decisions, 7) post-filing charge revisions and plea bargaining, 8) the ability of defendants to mount an effective defense or propose alternatives to pretrial detention, custodial sentencing, or revocation of release, 9) criminal and sentencing laws and guidelines, and 10) post-sentencing policies and practices such as probation revocation and the award of good-conduct credits. Each of these merits further research to identify elements of bias and/or disparate impact, and the effect they have on subsequent processing and disproportionate incarceration.

What I want to emphasize, however, is a point which applies to almost all of these separate inputs and stages of processing, namely, the need to focus more research on, and try to minimize the disparate impacts of, seemingly legitimate and race-neutral factors – the variables which, as Baumer notes, prior research has too often merely treated as controlled variables. As Baumer suggests (pp. 6-7), many observers especially in minority communities, view such disparate impact as, itself, a form of systemic bias, one which reflects and perpetuates social inequalities, disadvantage, and the crime which results from inequality and disadvantage. As Baumer says, we need more research on the current and prior biases embedded in controlled variables like offense and prior record. But we also need to critically examine the legitimacy, potential over-breadth, short-term crime-control benefits, and adverse long-term social and criminogenic consequences of law enforcement, case-processing, and sentencing factors with strongly disparate impacts. The disturbing and socially-destructive racial disproportionalities of inmate populations are probably due much more to disparate impacts than to bias, so it is just as important to study and try to minimize those impacts.

In particular, I would urge much more attention be given to offender prior records. What biases in prior and current charging contribute to the higher criminal history scores of minority offenders (Fraxe 2009)? How are such scores defined and weighted for sentencing purposes, in guidelines and customary norms, and how valid are the underlying punishment rationales? In particular, what kinds of prior record are reliable predictors of what kinds of future offending risk? What are the adverse consequences for offenders and their families and communities, of the sentence enhancements resulting from criminal history, and do they outweigh the crime-control benefits? If so, how can criminal history enhancements be more narrowly tailored, to minimize unproductive adverse consequences, in particular, the self-perpetuating cycle of community and individual disadvantage, crime, punishment, further disadvantage, and further
crime?

Other disparate-impact factors also merit much closer study, not only for embedded biases but also for weakness or overbreadth in their punishment rationales. Offender risk assessments are particularly likely to involve bias and/or overbreadth that unnecessarily magnifies disparate impact. Likewise, current-offense factors may reflect not only charging and plea bargaining biases, but also overbreadth, causing unnecessary disparate impact, in terms of the way in which offense severity is translated into sentence severity.

Conclusion ....

References [preliminary list]
