# Trademarks, Press Releases, and Policy: Will Rigorous Research Get in the Way?

*Kristin Bechtel*
*Crime and Justice Institute*
*Anthony W. Flores*
*California State University, Bakersfield*
*Alexander M. Holsinger*
*University of Missouri, Kansas City*
*Christopher T. Lowenkamp*
*Administrative Office of the U.S. Courts*

*The institutional goal of science is the extension of certified knowledge.*
*—Merton, 1942, 117*

**RESEARCH SIMPLY CAN'T** catch a break—it does not move quickly, or perhaps it does not get conducted, written up, reviewed and revised, disseminated, and read as fast as policy and practice can take hold. It is no secret that reports regarding new practices or concepts can be written up and more broadly branded, trademarked, marketed, and distributed to policy makers and practitioners when the information is not subjected to replication and peer review. Let's face it, rigorous research simply cannot be summarized in a tweet of 140 characters or less, and few would follow a twitter feed long enough to wait on a replication study before drafting state legislation and introducing reforms to policy and practice. The question remains, though, what are the implications for this when policies and practices are adopted and substantial funding is allocated without an adequate level of empirical support?

Before introducing the current study, it may be beneficial to offer some context for how these concerns have evolved. Perhaps taking a lesson from history would be beneficial to explain why or how the value of rigorous research and replication lost a bit of its luster. Perhaps history could also explain why rigorous research and replication may be running in second place behind the well-marketed and branded reports that attract such a wide, but more importantly, influential audience.

There have been several studies, even some subjected to peer review but without the findings being replicated elsewhere, that have widely influenced policy and practice or were simply catapulted to the elevated status being described as having achieved scientific merit. Labeling or study branding may be to blame for some of this, but it is unclear if labeling is the sole culprit, especially if the study resulted in a fundamental discussion of existing practices within criminal justice. A brief summary of some of these studies and their impact follows.

There have been several persuasive individual studies that have been labeled as "classic" or even "famous" despite a lack of methodological rigor and limited replication of findings. As Kulig, Pratt, and Cullen (2016) describe it, these studies, including the Stanford Prison Experiment, are often held in such high regard that few scholars question or critique the methodology or findings, despite the clear limitations that may be observed. So, in spite of the Stanford Prison Experiment suffering from both methodological and ethical challenges, this study has been branded a classic, but there may be an underlying reason for why it is held in such high regard. Kulig et al. (2016) clearly recognized that the Stanford Prison Experiment was "groundbreaking" because it called attention to the inhumanity of prisons and their impact on incarcerated individuals. The overall findings were timely and responded to shared concerns that imprisonment may be very detrimental. Essentially, this study propelled the discussion forward regarding imprisonment and the conditions in which individuals are incarcerated. Unfortunately, although attempts have been made to replicate the Stanford Prison Experiment, similar findings have not followed (Reicher & Haslam, 2006; Kulig et al., 2016).

Another single study that lacked methodological rigor but garnered much attention and sweeping political support both from conservatives and liberals is Martinson's 1974 "Nothing Works" article. This narrative review of 231 studies examining the effectiveness of rehabilitation programs suggested that rehabilitative models failed to produce any appreciable impact on recidivism; as Lipton (1998) expressed, Martinson's assertive summary was promptly deemed as fact (Sarre, 1999). As a result, there was growing interest in lengthier but determinate prison sentences without the addition of treatment and programming. Multiple studies followed questioning Martinson's infamous pronouncement, and one year following the publication of the "Nothing Works" article, Palmer (1975) concluded that 48 percent of the 82[1] studies reviewed by Martinson indicated that some rehabilitative programs were actually associated with reduced recidivism. Interestingly, by 1979, Martinson had recanted his findings, although this report was not as widely read as the original "Nothing Works" article. Decades of research followed, providing further empirical support for the effectiveness of rehabilitative approaches, but none of these later endeavors received the amount of publicity and broad but blind acceptance that the Nothing Works doctrine received.

While Martinson's conclusion that nothing works was widely accepted without critical review, the field has also borne witness to other correctional interventions being touted with great fanfare but with minimal replication and evaluation. The most recent intervention that seems to be spreading at an alarming rate despite limited research support is Project HOPE (Hawaii Opportunity with Probation Enforcement). Project HOPE uses swift and certain sanctioning practices for individuals placed on community supervision. It is important to note that there is some evidence to suggest the effectiveness of Project HOPE (Hawken & Kleiman, 2009); however, other researchers have pointed out that Project HOPE has not been subjected to considerable replication and evaluation. Further, the fundamental components of Project HOPE, namely deterrence and sanction-based approaches, have been questioned in previous meta-analytic reviews (see, for example, Gendreau, 2000) and, at a minimum, require additional and more rigorous review (Duriez, Cullen, & Manchak, 2014).

Several explanations have been offered as to why Project HOPE became such an overnight sensation, as the language used by its proponents to describe it carries an extraordinary amount of weight, including "There aren't any magic bullets that can end America's battle with crime and addiction. But HOPE comes closer than anything we have seen in a long time" (Gelb, 2011, p. 2 as cited in Duriez, Cullen, & Manchak, 2014). Given the broad and overwhelming praise that Project HOPE has received, it should come as no surprise that similar deterrence-based HOPE strategies have found their way into state criminal justice reforms and legislation.

Responding to Duriez et al.'s (2014) account of the limited research on the effectiveness of Project HOPE, Kleiman, Kilmer, and Fisher (2014) suggested that while replication of Project HOPE is not a standard recommendation for jurisdictions, the consideration should be directed toward adopting and following swift, certain, and fair sanctioning practices within community-based supervision. Given the attempted but perhaps unsuccessful replication of a similar Project HOPE-based program in Delaware, it seems that swift and certain sanctioning is hardly guaranteed to be an effective model for other jurisdictions or to be easily transferrable, with fidelity and similarly impressive results, to other settings (Duriez et al., 2014; O'Connell, Visher, Martin, Parker, & Brent, 2011). Cullen, Manchak, and Duriez's (2014) rejoinder to the Kleiman et al. (2014) response summarized the upshot of the lively discussion as "buyer beware." This is rather poignant, as a lesson learned from well-branded and marketed research is that we must all become better-informed consumers of information. This certainly does not suggest dismissing information outright, but instead calls us to review evidence within the context of its limitations. This approach has been referred to as "organized skepticism," wherein scholars make a conscious effort to operate from logic and empiricism rather than tradition and belief (Merton, 1942 and see Kulig, Pratt, & Cullen, 2016).

The current study focuses on the pretrial field. There has been an increasing interest in studying pretrial risk assessments and supervision practices to identify what the strongest predictors of pretrial failure are and what pretrial practices are most effective in reducing a defendant's risk of experiencing pretrial failure. Pretrial research is still in its infancy, and this area of criminal justice research does not compare with the extensive research conducted in the post-disposition field (Bechtel, Holsinger, Lowenkamp, & Warren, 2015). Similar to what has been noted within correctional literature, multiple pretrial studies have not yet been subjected to rigorous blind peer review and replication. The implications, of course, are that these pretrial practices and risk assessment instruments may be adopted without a clear understanding of these limitations; specifically, these practices may not prove effective if implemented elsewhere, and the risk assessments may not properly predict pretrial failure on a different target population. Widely marketed reports often use labels and branding that have the potential for attracting attention but do little to truly inform the consumer. For example, one such report describes two pretrial risk assessment instruments, the Ohio Risk Assessment System – Pretrial (ORAS-PAT) and the Virginia Pretrial Risk Assessment Instrument (VPRAI), as the "gold standards" for pretrial risk assessment, although neither of these two tools have been subjected to any blind peer review process regarding how well they predict pretrial failure (Lawrence, 2013, p.10).

Recently, there has been growing interest in understanding the impact of pretrial detention. Certainly, this is an appropriate topic to evaluate and is worthy of study since it has the potential to substantially inform practice. One study looking at data from over 150,000 defendants booked into Kentucky jails between July 2009 and June 2010 sought to examine whether or not the length of pretrial detention increased a defendant's likelihood to experience pretrial failure, including failure to appear and new arrest pending case disposition. The study revealed several interesting things. First, longer stays in pretrial detention, in particular 2 to 3 days (as opposed to 1 day or less), resulted in an increase in the likelihood of failure to appear and new arrest pending case disposition. Second, low-risk defendants were most likely to experience a greater likelihood for failure to appear and new arrest pending case disposition when detained for 2 to 3, and 4 to 7 days. Moderate-risk defendants were also found to experience higher rates of new arrest pending case disposition when exposed to pretrial detention stays of 2 to 3 days. The study also examined the impact of the length of pretrial detention on post-disposition recidivism and suggested that a stay of 2 days or longer was associated with post-disposition recidivism when measured at both 1 year and 2 years post disposition. These results appeared to be strongest when examining the impact on low-risk defendants (Lowenkamp, VanNostrand, & Holsinger, 2013). While this study did not undergo blind peer review, it has sparked great interest in the pretrial field among both practitioners and researchers. In an effort to "practice what we preach" by replicating and expanding upon this research, in the current study we seek to effect an organized skepticism in the pretrial literature by evaluating the impact of pretrial detention length on pretrial failure, and specifically whether or not longer stints of pretrial detention result in an increased likelihood for pretrial failure. The relationship between length of time spent in pretrial incarceration and various outcomes

[1] Palmer (1975) examined a subset of Martinson's 231 studies to exclude those that used an outcome measure other than recidivism (e.g., change in attitude, adjustment to the community, educational achievement).

may be more complex than anyone in the field realizes at this point. This alone should serve as a clarion call for more research investigating every aspect of the issue.[2] Using the State Court Processing Statistics (SCPS) data from the Bureau of Justice Statistics, the study that follows is the first replication to test the impact of pretrial detention on pretrial failure.

## Method

### *Data Source & Participants*

The data used for this study come from the State Court Processing Statistics 1990-2009: Felony Defendants in Large Urban Counties (U.S. Department of Justice, Bureau of Justice Statistics, 2013). A detailed description of these data is available in the study codebook. In summary, this dataset contains data on 151,459 felony cases processed in 40 of the 75 most populous counties during even-numbered years from 1990-2006 and 2009. Data collected on these cases include defendant demographics, criminal history, pretrial release and detention, pretrial conduct, adjudication, and sentencing information.

### *Measures*

Demographic measures used in this study include age in years, gender (coded as 1 for female and 0 for male), race (coded as 0 for white, 1 for black, and 2 for other) and Hispanic origin (coded 0 for not of Hispanic origin and 1 for Hispanic origin).

Data related to case processing included the number of days from arrest to release, offense type (violent, property, drug, or public order), release type (financial release, nonfinancial release, emergency release, held on bail, denied bail, release conditions unknown, detained reasons unknown, and case closed). Measures of conduct while on pretrial release were developed based on data included in the dataset. Three outcome measures were created: failure to appear (FTA), arrest for any new criminal conduct (arrest), and arrest for a new violent crime (violent).

Two measures were created based on the measure "days from arrest to release." One measure is a log transformation of "days from arrest to release" and the other is the squared value of the log transformation. These measures were developed for two reasons. First the distribution of "days from arrest to release" was highly skewed and leptokurtic (not normally distributed). To induce normality and thus make the measure useful in multivariate models the variable was transformed using a log transformation. The second variable created is simply the squared value of the transformed variable. This was done to address the possible nonlinear relationship between "days from arrest to release" and one or more of the outcomes of interest.

In an effort to control for differences in defendant characteristics that relate to pretrial outcomes of interest, three risk scales were developed. These three risk scales predict the three outcomes described above: FTA, arrest, and violent. The variables used to create the FTA scale are the number of prior FTAs, criminal justice system status, the number of prior arrests, gender, offense type, and number of current charges. The risk scale predicting arrest contains measures of prior commitments to jail, criminal justice system status, number of prior serious arrests, number of arrests, gender, offense type, and number of current charges. The risk scale predicting violence contains measures of criminal justice system status, number of prior serious arrests, number of prior arrests, offense type, number of charges, prior convictions for violent offenses, and gender. All three scales produced acceptable AUC-ROC values (0.64, 0.68, 0.68 for the FTA, arrest, and violent scales respectively).

## Analysis

Analysis in this study included bivariate and multivariate statistical models examining the relationship between days from arrest to release and the three outcomes of interest. Since this study focuses on the released population only, the sample was reduced by excluding those defendants that were not released pretrial (n = 55,349). The sample was further reduced by excluding those cases with missing data on one of the key variables (n = 24,896), yielding a final sample size of 47,387. For comparison purposes we provide the descriptive statistics for the entire sample as well as the reduced sample (see Table 1).

In addition to the bivariate and multivariate tests run on the sample of 47,387 cases, a series of matched samples was developed for analysis. These matched samples provide a more rigorous test of the relationship between days detained pretrial and the outcomes of interest. The matching process involved matching defendants who were released in a particular number of days (for example all defendants released on day 5) to defendants released in 0 days. The defendants were matched on county, offense type, gender, age, race, Hispanic origin, type of release, and each of the three risk scales. This matching process was repeated for defendants released in 1 day up to 10+ days. This, in effect, created 10 matched samples comparing the outcomes of defendants released on day 0 to defendants released on day 1, day 2, day 3…day 10+. Since these samples were matched on all the relevant controls, only bivariate analyses were run on these samples.

## Results

Table 1 provides descriptive statistics on the defendant's demographic data, case-related information, and the risk scales. These data are presented for both the complete sample and the reduced sample. While many of the differences between the entire sample and the reduced sample are statistically significant, it can be argued that the two samples are similar although not identical. Even so, the differences noted here probably preclude extending the findings with the reduced sample to the sample containing detained offenders and those that were released but excluded due to missing data.

The sample used for most of the subsequent analyses in this manuscript is the released sample with complete data. This sample is, on average, 30 years old and typically male (78 percent). Fifty percent of the defendants were black, and 48 percent were white. The majority of defendants (88 percent) were not of Hispanic origin. The offense of arrest was categorized as a drug offense (36 percent), followed by property offense (32 percent), violent offense (22 percent) and public order (10 percent). The majority of released defendants were released by financial release (52 percent).

Table 2 provides the failure rates for released defendants with missing data (eventually excluded from the sample) and those without missing data. About 20 percent of the sample is identified as having at least one failure to appear. Defendants are arrested for a new crime while on pretrial release 15 percent of the time (18 percent of the time for those with missing data). Finally, 2 percent of the sample is arrested for a violent offense while on pretrial release.

The main purpose of this manuscript is to explore the relationship between days detained prior to pretrial release and pretrial

[2] In a separate analysis conducted by Holsinger (2016), length of time in pretrial detention was observed to be significantly and positively correlated with FTA (every time increment of pretrial detention), but completely unrelated to NCA. Further, length of time incarcerated pretrial was found to be significantly related to post-disposition NCA at the 12-month point, but not the 24-month point in time.

**TABLE 1.**

*Descriptive Statistics on Demographic Characteristics, Current Offense Type, and Release Type for Entire Sample and Released Sample with Complete Data*

| | All Cases | | Released with Complete Data | |
|---|---|---|---|---|
| **Measure** | **Number** | **Mean or Percent** | **Number** | **Mean or Percent** |
| Age | 149,972 | 30.37 | 47,387 | 30.45 (10.72) |
| Days from Arrest to Release | 86,253 | 10.44 (29.52) | 47,387 | 10.34( 29.58) |
| Days from Arrest to Release | 86,253 | 0.41 (2.13) | 47,387 | 0.35 (2.17) |
| Risk Scale 1 (FTA)* | 118,303 | 21.77 (9.06) | 47,387 | 20.49 (8.43) |
| Risk Scale 2 (arrest)* | 124,326 | 18.42 (9.89) | 47,387 | 15.96 (9.09) |
| Risk Scale 3 (violent)* | 125,925 | 2.11 (1.61) | 47,387 | 1.75 (1.36) |
| Gender* | | | | |
| Female | 25,518 | 17 | 10,461 | 22 |
| Male | 125,407 | 83 | 36,926 | 78 |
| Missing | 534 | <1 | — | — |
| Race* | | | | |
| White | 55,848 | 37 | 22,525 | 48 |
| Black | 69,611 | 46 | 23,659 | 50 |
| Native American, Alaskan Native | 535 | <1 | 217 | <1 |
| Asian, Pacific Islander | 2,549 | 2 | 986 | 2 |
| Missing | 22,916 | 15 | — | — |
| Hispanic Origin* | | | | |
| Yes | 32,822 | 22 | 5,635 | 12 |
| No | 97,721 | 65 | 41,752 | 88 |
| Missing | 20,916 | 14 | — | — |
| Offense Type* | | | | |
| Violent | 37,456 | 25 | 10,479 | 22 |
| Property | 47,117 | 31 | 14,970 | 32 |
| Drug | 52,353 | 35 | 17,191 | 36 |
| Public Order | 14,471 | 10 | 4,747 | 10 |
| General Release Category* | | | | |
| Financial Release | 43,225 | 29 | 24,783 | 52 |
| Nonfinancial Release | 42,325 | 28 | 21,066 | 44 |
| Emergency Release | 744 | 0 | 344 | <1 |
| Held on Bail | 44,767 | 30 | — | – |
| Denied Bail | 8,380 | 6 | — | – |
| Release Conditions Unknown | 4,108 | 3 | 1,194 | 3 |
| Detained, Reasons Unknown | 2,202 | 1 | — | – |
| Case Closed | 2,001 | 1 | — | – |
| Missing | 3,707 | 2 | — | – |

* $p \leq 0.001$

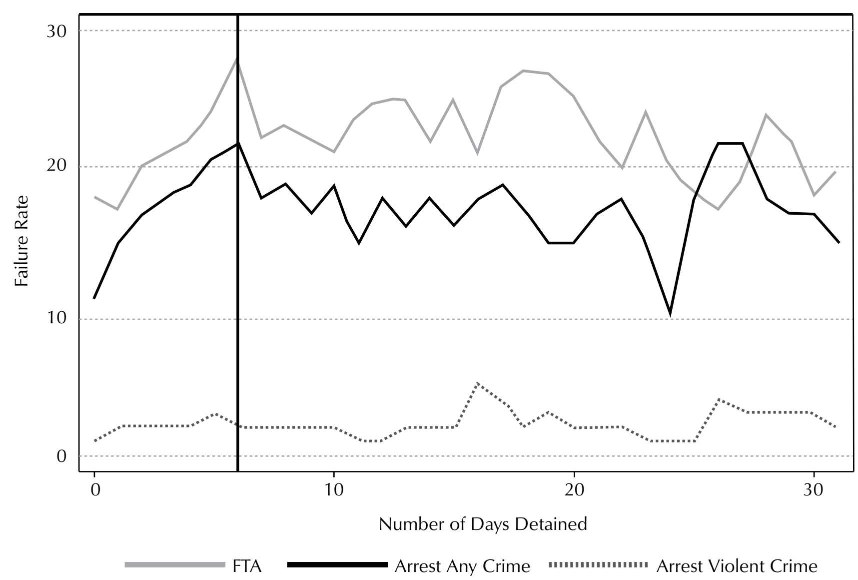
**TABLE 2.**
*Descriptive Statistics on Outcome Measures for Entire Sample and Released Sample with Complete Data*

| | Released with Missing Data | | Released with Complete Data | |
|---|---|---|---|---|
| **Measure** | **Number** | **Mean or Percent** | **Number** | **Mean or Percent** |
| Arrest for Any New Crime* | | | | |
| No | 31,772 | 82 | 40,171 | 85 |
| Yes | 6,899 | 18 | 7,216 | 15 |
| Failure to Appear* | | | | |
| No | 32,090 | 77 | 38,133 | 80 |
| Yes | 9,684 | 23 | 9,254 | 20 |
| Arrest for New Violent Crime | | | | |
| No | 37,604 | 98 | 46,588 | 98 |
| Yes | 756 | 2 | 799 | 2 |

* $p \leq 0.001$

outcomes. Given prior findings regarding this relationship, we suspected that the relationship was nonlinear. Figure 1 plots the failure rates by day of release for each of the three outcomes of interest. As suspected, for both FTA and arrest the rates follow an upward trend up to and including day 6 but then begin to drop and then follow a somewhat random pattern with an overall downward trend. This pattern seems to be absent from the trend for arrest for a violent crime.

**FIGURE 1.**
*Failure Rates By Days Detained (greater than 31 days recoded to 31)*




Given the nonlinearity of the relationship between the number of days detained and FTA and arrest, a squared term was included in the logistic regression models predicting FTA and arrest. The squared term was not included in the model predicting arrest for a violent crime. These logistic regression models are presented in Table 3 and indicate that days detained and the associated squared term are statistically significant only when predicting arrest. As the number of days increases, so too does the likelihood of arrest for any crime while on pretrial release. This positive effect of days detained on arrest seems to diminish as one moves up the scale of days detained. It should be noted that when predicting FTA and violent arrest the days detained prior to release are not significant predictors.

The risk category (based off the risk score) is a significant predictor of each of the three outcomes. Defendants in the moderate category are about 2 to 2.5 times as likely to experience the outcome as those in the low-risk category. High-risk defendants are roughly 4.5 to 5.5 times as likely to experience the outcome when compared to low-risk defendants.

Age and race are significant predictors in two of the three outcomes. Older defendants were less likely to experience the outcomes arrest (any criminal arrest) or violent (arrest for a violent offense). Black defendants were more likely to experience an FTA or violent compared to white defendants and defendants of "other" races. There was no effect of gender[3] or ethnicity (Hispanic origin) in any of the three models.

Release type was a significant predictor of FTA. Those released through nonfinancial release, emergency release, or in situations where release conditions were not known were more likely to fail to appear than those defendants released by financial release. Those released by nonfinancial release and unknown conditions of release were about 1.5 times as likely to fail to appear compared to those released on a financial release. Those released on emergency release were about 2.9 times as likely to FTA as those released by financial release.

Tables 4, 5, and 6 present the failure rates for FTA, arrest, and violent (respectively) using the matched samples. As indicated earlier, the matching process we used generated 10 samples that contained defendants released on a particular day matched to those released on day 0 (arrested and released the same day). The failure rates were then calculated for each

[3] Recall that gender was included in each of the risk scales.

**TABLE 3.**
*Logistic Regression Models Predicting each Outcome*

| Variable | Failure to Appear | Any Criminal Arrest | Violent Arrest |
|---|---|---|---|
| Days Detained Transformed | 1.04 | 1.07* | 1.06 |
| Days Detained Transformed Squared | 1.00 | 0.97* | – |
| Release Type | | | |
| Financial Release | | | |
| Nonfinancial Release | 1.46* | 1.12 | 0.99 |
| Emergency Release | 2.86* | 0.98 | 0.40 |
| Release Conditions Unknown | 1.58* | 1.16 | 1.41 |
| Risk Category | | | |
| Low | | | |
| Moderate | 2.12* | 2.53* | 2.59* |
| High | 4.53* | 5.39* | 5.76* |
| Offense Type | | | |
| Violent | | | |
| Property | 0.90 | 1.05 | 1.04 |
| Drug | 0.95 | 1.06 | 0.95 |
| Public Order | 1.15 | 1.09 | 1.10 |
| Age | 1.00 | 0.97* | 0.96* |
| Race | | | |
| White | | | |
| Black | 1.26* | 1.16 | 1.48* |
| Other | 0.75 | 0.95 | 0.94 |
| Gender | | | |
| Male | | | |
| Female | 0.97 | 0.97 | 0.83 |
| Hispanic Origin | | | |
| No | | | |
| Yes | 1.04 | 1.14 | 1.12 |
| Constant | 0.10* | 0.17* | 0.02* |

**TABLE 4.**
*FTA Rates by Days Detained Matched Cases*

| | 0 Days | | More than 0 Days | | |
|---|---|---|---|---|---|
| Days | Number | Percent | Number | Percent | p |
| 1 | 3814 | 16 | 3814 | 15 | 0.41 |
| 2 | 1374 | 19 | 1374 | 18 | 0.88 |
| 3 | 711 | 18 | 711 | 20 | 0.28 |
| 4 | 548 | 15 | 548 | 21 | 0.01 |
| 5 | 443 | 19 | 443 | 24 | 0.07 |
| 6 | 371 | 22 | 371 | 24 | 0.60 |
| 7 | 336 | 16 | 336 | 19 | 0.42 |
| 8 | 297 | 18 | 297 | 23 | 0.13 |
| 9 | 214 | 17 | 214 | 16 | 0.90 |
| 10+ | 2375 | 19 | 2375 | 21 | 0.09 |

group and are presented in two columns. The column labeled “0 days” contains the failure rates for each of the 10 matched samples of those released in 0 days. For illustration purposes consider the row labeled “1” under the “Days” column. This row indicates that we could match 3,814 defendants released on day 1 to 3,814 defendants released on day 0. The failure rate for those released on day 0 is 16 percent, whereas the failure rate for those released on day 1 is 15 percent. The row labeled “8” under the “Days” column indicates that there were 297 defendants released on day 8 that could be matched to 297 defendants released on day 0. The failure rate for those released on day 0 is 18 percent, whereas the failure rate for those released on day 8 is 23 percent. It should be noted that in only one instance does the difference in FTA rates reach statistical significance (those released on day 4 compared to the matched sample of those released on day 0).

Table 5 contains the arrest failure rates for each of the matched samples. Three of the ten samples generated differences that were significant and favored the group of defendants released on day 0. When compared to defendants released on days 1, 4, and 10 or more, the defendants released on day 0 had a significantly lower rate of arrest for any crime.

Finally, Table 6 contains the rates of arrest for a violent crime for the ten matched samples. In Table 6 only one difference is statistically significant. Defendants released on day 10 or more have a significantly higher arrest rate for a violent offense compared to the matched sample that was released on day 0.

## Discussion

Using empirical evidence to inform, guide, and evaluate policy and practice is the hallmark of providing ethical and professional human service. Unfortunately, the accumulation of knowledge is often a painstakingly slow process that is seemingly never-ending. Areas of policy and practice for which little to no research evidence exists can become quite vulnerable in this regard. In an era of near-instantaneous communication and information sharing, the time required for a research project to go from inception to completion and publication must be trying (to say the least!) for those charged with creating evidence-based policy. In fact, several recent publications within the discipline of criminology/criminal justice have focused on this very issue. The need for informed policy and practice exists in real-time, while the world of

**TABLE 5.**
*Arrest for Any New Crime Rates by Days Detained Matched Cases*

| | 0 Days | | More than 0 Days | | |
|---|---|---|---|---|---|
| Days | Number | Percent | Number | Percent | p |
| 1 | 3814 | 10 | 3814 | 12 | 0.01 |
| 2 | 1374 | 11 | 1374 | 13 | 0.18 |
| 3 | 711 | 12 | 711 | 15 | 0.10 |
| 4 | 548 | 10 | 548 | 15 | 0.01 |
| 5 | 443 | 12 | 443 | 16 | 0.10 |
| 6 | 371 | 13 | 371 | 15 | 0.46 |
| 7 | 336 | 13 | 336 | 15 | 0.43 |
| 8 | 294 | 14 | 294 | 17 | 0.31 |
| 9 | 214 | 9 | 214 | 9 | 1.00 |
| 10+ | 2375 | 11 | 2375 | 13 | 0.01 |

**TABLE 6.**
*Arrest for New Violent Crime Rates by Days Detained Matched Cases*

| | 0 Days | | More than 0 Days | | |
|---|---|---|---|---|---|
| Days | Number | Percent | Number | Percent | p |
| 1 | 3814 | 1.02 | 3814 | 1.47 | 0.08 |
| 2 | 1374 | 1.09 | 1374 | 1.67 | 0.19 |
| 3 | 711 | 1.97 | 711 | 1.13 | 0.20 |
| 4 | 548 | 1.46 | 548 | 1.64 | 0.81 |
| 5 | 443 | 2.26 | 443 | 3.16 | 0.41 |
| 6 | 371 | 1.62 | 371 | 1.62 | 1.00 |
| 7 | 336 | 0.89 | 336 | 1.79 | 0.31 |
| 8 | 294 | 1.00 | 294 | 2.4 | 0.20 |
| 9 | 214 | 0.00 | 214 | 1.4 | 0.08 |
| 10+ | 2375 | 1.05 | 2375 | 1.81 | 0.03 |

empirical research often exists in a vacuum, devoid of "real world" demands.

Project HOPE serves as one example of this quandary. Although an initial study supports the efficacy of Project HOPE's punishment-based strategy for reducing recidivism, program demand has come to supersede calls for additional investigations seeking to replicate the program's initial findings in diverse settings across different client populations (Duriez, Cullen, & Manchak, 2014). So much so that a number of states have adopted HOPE-similar programs more on the basis of hype, branding, and marketing than on the basis of replicated and methodologically rigorous evidence attesting to validity. Another potential example of hastily informed policy surrounds pretrial research. While this is clearly an underdeveloped area of research (Bechtel et al., 2015), recent policy has emerged that relies primarily on the results of one study (Lowenkamp, VanNostrand, & Holsinger, 2013) to support its branded, trademarked, and widely marketed pretrial release policy proposal. We see this as problematic. Accordingly, this research sought to contribute to the existing pretrial literature by replicating the research of Lowenkamp, VanNostrand, and Holsinger using a large, diverse, and fairly representative sample drawn from the 75 largest U.S. counties.[4] Specifically, this study examined the effect of pretrial detention length on several measures of pretrial failure.

The analyses conducted here reveal a number of important findings, particularly as they compare to those of Lowenkamp, VanNostrand, and Holsinger (2013). First, bivariate analysis of failure rates by the number of days detained indicates that there is a sharp increase in both FTA and predisposition arrest (but not violent crime arrest) through the first six days in detention. After that however, the bivariate relationship seems to become random. This pattern corresponds with Lowenkamp et al.'s (2013) findings to some extent in that the first few days of detention seem to impact pretrial outcome (again, based on results from bivariate analysis). Results of this research support the immediacy of the impact that detention has on pretrial outcome, while the results of Lowenkamp et al. (2013) show that the deleterious effects of detention begin to surface after days two or three.

Multivariate analysis further investigating the relationship between detention and pretrial outcome lends credence to the skepticism discussed above. Once we controlled for a number of other variables that are potentially relevant toward the prediction of pretrial outcome, the bivariate and apparent relationships between number of days detained and each of the three outcomes are largely explained away. Given the fairly large sample used for these analyses, and given that sample size drives significance, the importance of this finding should not be overlooked. Analyses indicate that the insignificance of days detained is due, in large part, to the offender risk variable, which demonstrated significant and relatively strong relationships with all three outcome measures (FTA, any criminal arrest, and violent arrest). In this model, financial release, race, and age were also significantly related to some of the outcome measures (but not all three).

In addition to attempting a replication of the Lowenkamp et al. (2013) study, this research also employed a more rigorous analytical approach to exploring the relationship between pretrial detention and outcome through the use of matched samples. Because defendants serving 0 days of pretrial detention were matched to those serving a particular number of days (ranging from 1-10+) in pretrial detention on characteristics theoretically relevant to pretrial misconduct once released, each of the two groups of defendants were rendered essentially "equal." The rigor inherent in this type of analysis is powerful because any difference in defendant outcomes is then more likely to be attributable to the only other thing left to vary, namely time spent in pretrial detention.

Results from the matched samples analyses comparing defendants who served 0 days to defendants who served 1 through 10+ days in detention indicate that the effect of pretrial detention on outcomes disappears in almost every comparison. Although there were a handful of significant relationships evidenced in these matched analyses, recall that 30 different matched comparisons were conducted

[4] Note that the sample used in Lowenkamp, VanNostrand, & Holsinger was drawn from one state—Kentucky.

(10 for each of the three outcomes) and only 5 were significant at the $p < 0.05$ level of significance. While the sample sizes for the matched analyses are smaller than those in the multivariate tests, they are still large enough that meaningful differences would have attained statistical significance. We think that the relatively large sample size employed in these analyses explains at least some of the five significant relationships and conclude that predicting pretrial outcome is likely a very complicated issue that may or may not be affected by days spent in pretrial detention. Furthermore, we absolutely caution against creating, branding, and marketing any policy that is informed by just the Lowenkamp et al. (2013) study, or even that study and this one taken together. Clearly, the inconsistent and in some cases contradictory findings of this and the Lowenkamp et al. (2013) study make the obvious case against deriving policy from one or even a few studies, particularly those that have not undergone the peer review process and/or are lacking in methodological rigor.

There were several limitations present in this research. First, these data include only felony defendants, so the results presented cannot speak to any potential effects of pretrial detention on outcome for defendants with less serious charges. Second, these data were collected from the most populous counties in the U.S., rendering the applicability of these results to smaller and more rural counties questionable. Third, because we were interested in examining pretrial days in detention on pretrial outcome, we were forced to exclude a large number of defendants from our sample who were not released pretrial (and thus could not have experienced FTA or been arrested) as well as a large number of defendants for whom key data were missing. Although analyses comparing these two groups of felony defendants (those with complete versus missing data) did reveal some significant differences between the two groups, we contend that these differences are not substantive (refer back to Tables 1 & 2). Finally, the matching process used in this study was fairly restrictive and led to many cases being eliminated from the matched analysis, as a usable match was not identified. As such, future research attempting to replicate these findings might consider other methods of sample matching, such as propensity score matching, in which propensity score values can be used as matching and/or regression weights that will allow for the use of a greater percentage of cases.

In conclusion, this research represents the second study to examine the effect of days in pretrial detention on pretrial outcome, and the first attempt at fostering an organized skepticism about this topic. We feel this skepticism is especially justified given the policy implications derived from the first study by Lowenkamp et al. (2013). That study, using data solely from the state of Kentucky, found that longer stays in pretrial detention affected pretrial outcome. However, the data used in the present study (collected from a national sample) shows the effect only in bivariate models (save for one multivariate model). Furthermore, the effects of days in pretrial detention on pretrial outcome evidenced here appear to be few (a mere 5 significant effects out of 30 models) as well as inconsistent, especially once the results of the matched models are considered. Unfortunately, these findings fail to replicate the Lowenkamp et al. (2013) results and seem to indicate that this is very possibly a function of increased methodological rigor.

Undoubtedly, a balance must be struck between the need for replication, peer review, and disseminating information broadly, but reliably, to stakeholders, practitioners, researchers, and students alike. There has to be a consensus that both peer-reviewed journals and research reports that do not undergo a peer-reviewed process or have yet to be subjected to replication serve a valuable purpose. Primarily, we must seek to increase the knowledge of the consumer, but also clearly offer what the limitations are for the existing research and what next steps should occur before broad adoption and implementation of new practices and tools follow. The next steps for the current study will be to submit this evaluation for blind peer review. Although this process will certainly require additional time, we reserve the right to market and broadly share the results—perhaps with a 140-character tweet.

## References


Bechtel, K., Holsinger, A., Lowenkamp, C., & Warren, M. (2015). A meta-analytic review of pretrial research: Risk assessment, bond type, and interventions. *Under review at Journal of Crime and Justice.*

Cullen, F. T., Manchak, S. M., & Duriez, S. A. (2014). Before adopting Project HOPE, read the warning label: A rejoinder to Kleiman, Kilmer, and Fisher's comment. *Federal Probation, 78*(2), 75-77.

Duriez, S. A., Cullen, F. T., & Manchak, S. M. (2014). Is Project HOPE creating a false sense of hope? A case study in correctional popularity. *Federal Probation, 78*(2), 57-70.

Gendreau, P., Goggin, C., Cullen, F. T., & Andrews, D. A. (2000). The effects of community sanctions and incarceration on recidivism. *Forum*, *12*(2), 10-13.

Hawken, A., & Kleiman, M. (2009). Managing drug involved probationers with swift and certain Sanctions: Evaluating Hawaii's HOPE. Washington, D.C.: Office of Justice Programs, U.S. Department of Justice.

Holsinger (2016). Unpublished Data From Johnson County, Kansas. University of Missouri, Kansas City.

Kleiman, M. A. R., Kilmer, B., & Fisher, D. T. (2014). Response to Duriez, Cullen, and Manchak: Theory and evidence on the swift-certain-fair approach to enforcing conditions of community supervision. *Federal Probation, 78*(2), 71-74.

Kulig, T. C., Pratt, T. C., & Cullen, F. T. (2016 in press). Revisiting the Stanford Prison Experiment: A case study in organized skepticism. *Journal of Criminal Justice Education.*

Lawrence, S. (2013). Managing jail populations to enhance public safety: Assessing and managing risk in the post re-alignment era. Unpublished manuscript. Paper written for the Executive Session on Public Safety Realignment.

Lipton, D. S. (1998). The effectiveness of correctional treatment revisited thirty years later: Preliminary meta-analytic findings from the CDATE Study. Paper presented to the 12th International Congress on Criminology, Seoul, Korea. August 1998.

Lowenkamp, C. T., VanNostrand, M., & Holsinger, A. M. (2013). The hidden costs of pretrial detention. Houston, TX: Laura and John Arnold Foundation.

Martinson, R. (1974). What Works? - Questions and answers about prison reform. *The Public Interest, 35,* 22-54.

Martinson, R. (1979). New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review, 7,* 243-258.

Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology*, *1*, 115–126.

O'Connell, D., Visher, C., Martin, A., Parker, S., & Brent, J. (2011). Decide your time: Testing deterrence theory's certainty and celerity effect on substance-using probationers. *Journal of Criminal Justice, 39,* 261-267.

Palmer. T. (1975). Martinson revisited. *Journal of Research in Crime and Delinquency*, *12*, 133-152.

Reicher, S., & Haslam, S. A. (2006). Rethinking the psychology of tyranny: The BBC prison study. *British Journal of Social Psychology, 45,* 1-40.

Sarre, R. (1999). Beyond "What Works?": A 25 year jubilee retrospective of Robert Martinson. *Paper presented at the History of Crime, Policing and Punishment Conference.* Australian Institute of Criminology.

U.S. Department of Justice, Bureau of Justice Statistics. (2013). *Felony defendants in large urban counties, 2009 – Statistical tables.* Washington, D.C.