Communicating Risk Information at Criminal Sentencing in Pennsylvania: An Experimental Analysis

R. Barry Ruback
Penn State University
Cynthia A. Kempinen
Pennsylvania Commission on Sentencing
Leigh A. Tinik
Pennsylvania Commission on Sentencing
Lauren K. Knoth
Penn State University

RESEARCH IS RELEVANT to policy if it assesses the effects of different policy options using measures that are important to policy makers (Ruback & Innes, 1988).[1] But policy-relevant research by itself does not necessarily lead to policy change. Unless policy makers understand both the findings of the substantive research and how it can be implemented in the policy context, the research is unlikely to be used. Thus, for implementing policy and for understanding the implementation process, it is important to study how policy-relevant research is best communicated to policy makers.

One policy decision currently facing many states is the type of information that should be available for sentencing, treatment, and release decisions in criminal justice. In particular, at criminal sentencing the information judges are given can include or not include an actuarial instrument predicting the likelihood of recidivism. This study experimentally examined the communication of such recidivism risk information to judges, attorneys, and probation officers in order to determine how best to communicate both risk scores and the meaning of those risk scores.

[1] Points of view expressed here do not necessarily represent those of the Pennsylvania Commission on Sentencing.

Actuarial Predictions of Recidivism

Most actuarial risk instruments classify individuals into risk categories, each of which has an associated probability of recidivism based on the proportion of individuals who recidivated (Scurich & John, 2012). The assumption is that an individual in a particular risk group has a probability of recidivating similar to the overall group. Critics of actuarial risk assessments have suggested that the application of group-level probabilities to assess an individual's likelihood of recidivating is inappropriate and fails to meet any standards of precision or certainty (Hart & Cook, 2013). But there have been responses to these criticisms, and the debate over the accuracy and use of risk assessments is ongoing (Mossman, 2015; Harris, Lowenkamp, & Hilton, 2015).

In general, actuarially based predictions are more accurate than clinical judgments because humans are subject to numerous errors and biases (Kahneman, 2011; Meehl, 1986). Moreover, because humans can be tired or bored or distracted, they may make different decisions at different times about the same problem. This inconsistency further lowers the validity of their predictions. Similar criticisms have been made of structured professional judgment (SPJ) models, with some authors concerned about the subjectivity introduced by clinicians during completion of the assessment (Hilton, Harris, & Rice, 2006). The superiority of actuarial predictions over clinical judgments has been argued for 60 years (Meehl, 1954), with studies comparing the accuracy of actuarial risk assessment instruments, SPJ tools, and unstructured clinical judgment finding that pure actuarial models perform as well as or better than SPJ tools or clinical judgments (Harris & Rice, 2015; Campbell et al., 2007; Hanson & Morton-Bourgon, 2004).

Actuarial risk assessment is presumed to have several advantages in sentencing, including improving decision making, limiting discretion, increasing accountability, and better predicting future risks. Because of these advantages, the trend is for jurisdictions to use actuarial instruments in sentencing. In recent years, states have begun to require that actuarial risk scales be incorporated into criminal justice decisions (Monahan & Skeem, 2013). In particular, seven states are developing or have developed statistical models of recidivism for use at sentencing (Hannah-Moffat, 2013). The assumption behind these laws is that judges will be able to make more accurate predictions of future offending if they are given actuarial models than if they rely only

on their own knowledge and experience.

In practice, however, actuarial models will be better only if judges, attorneys, probation officers, and others concerned with sentencing understand the statistical information given to them. Although there has been research on how best to convey actuarial risk information to mental health practitioners and (regarding weather) to the general public (Monahan & Steadman, 1996), there has been little work on how risk information should be conveyed to practitioners in the criminal justice system (Buchanan, 2013). States vary in how risk information is presented at sentencing (e.g., length of report, specific scores versus summary levels of risk), but there are "no evidence-based practices to guide decisions" (Casey, Warren, & Elek, 2011, p. 54) about which methods are best. More generally, whether risk assessment tools actually affect and improve sentencing needs to be tested (Skeem, 2013).

The study presented here is an experimental investigation of the communication of statistical information about recidivism risk in sentencing. We were interested in knowing whether the statistical information affects beliefs about risk and, if so, whether these effects are consistent across crimes and cases.

Background for the Study

Act 95 of 2010 (42 PA.C.S. §2154.7) mandated that the Pennsylvania Commission on Sentencing adopt an empirically based risk assessment instrument to be used by judges at sentencing that takes into account an offender's risk of re-offense and threat to public safety and that can be used to help determine whether the offender should be considered for alternative sentencing programs.

Over the past four years, the research staff of the Pennsylvania Sentencing Commission (PCS) have developed an actuarial instrument, based on the procedure outlined by Gottfredson and Snyder (2005), for offenders at Levels 3 and 4 of the Guidelines.[2] The staff focused on these levels because of the wide variety of offense seriousness encompassed in these levels and the variety of possible sentences (including incarceration, probation, and alternative sentencing) that are available under the Guidelines. The researchers developed the model using a random sample of half of the Level 3 and 4 offenders sentenced during the three-year period 2004-2006. Predictors included information in the PCS database, as well as prior criminal history information available from the Pennsylvania State Police. The dependent variable was a rearrest for any crime within three years after sentencing (for those on probation) or after release from incarceration[3] as evidenced in the Pennsylvania State Police database.

The final scale, which was developed to predict recidivism, not reduce recidivism (see Monahan & Skeem, 2013, for the distinction), was a weighted measure of eight factors: age, gender, county, total prior arrests, prior property arrests, prior drug arrests, offense gravity score (the PCS measure of offense seriousness), and whether the current crime was a property offense.[4] Information was included in the scale if it was available statewide to probation and court staff at the time of sentencing (such information as prior drug use, criminal attitudes, and psychopathy was not), if it was reliable, and if it was predictive of subsequent arrest, the latter two being standards proposed by Gottfredson and Moriarty (2006). What these conditions meant was that no dynamic factors and no validated scales (e.g., LSI-R) could be used in the risk scale. Scores on the constructed scale could range from 0 to 14. The scale was validated on the remaining half of the PCS data for the 2004-2006 period and revalidated on PCS data from the years 2007-2008.[5]

Although risk assessment instruments have been used by practitioners in criminal justice, especially for prison classification and parole release decisions, there is little research on how risk information can best be presented to nonspecialists, particularly individuals without statistical training, such as lawyers and judges. This lack of research is problematic because how information is presented affects the way it is used in decisions (Sanfey & Hastie, 1998).

Three examples illustrate the effects of presentation. First, estimates involving frequencies (e.g., 1 in 10) lead to greater perceived risk than the equivalent percentage (e.g., 10 percent), probably because frequencies are easier than percentages to visualize (Slovic, Monahan, & MacGregor, 2000). Second, one study found that clinical psychologists who work in forensic settings were less likely to make release decisions when violent behavior was described in vivid rather than pallid terms (Monahan, Heilbrun, Silver, Nabors, Bone, & Slovic, 2002). Third, there are differences in perceived risk depending on whether information is presented as the probability of an event occurring (e.g., violence) rather than the probability of no event (e.g., no violence). Thus, as compared to statistics framed in a negative fashion (e.g., 74 percent likely to be nonviolent), statistics framed in a positive fashion (e.g., 26 percent likely to be violent) lead to more commitment decisions (Scurich & John, 2011), an effect that occurs because people have a strong aversion to loss (Kahneman, 2011).

The present study addressed four issues. First, does the risk information affect judgments? Second, if so, does the way it is presented make a difference in the size of the effect? We expected the risk information to affect participants' judgments, but we did not have hypotheses about the effects of type of presentation. Third, does the risk information have the same effect across all types of crime, or does it vary by the type of crime?

Fourth, do decision makers in criminal justice have a preference regarding the presentation of risk information? Do they want just the score, information about the items on which the offender had a score, or information about all of the items, regardless of whether or not the item was a risk? And, regarding the meaning of risk score in terms of predicted risk of recidivism, do they prefer that the information be presented in a table or a graph? Based on a study by Scurich, Monahan, and John (2012), we expected participants to prefer more rather than less information, but based on the absence of empirical data, we did not make a hypothesis regarding a preference for type of presentation.

These questions were tested in an experimental framework using case information from six actual cases. Two examples of each

[2] The Guidelines have five sentencing levels, with 1 representing the least serious offenders and 5 representing the most serious offenders. Since this study, the Commission has been working on the development of a risk assessment instrument for all five sentencing levels.

[3] For those sentenced to state prison we used the release date from the Pennsylvania Department of Corrections. For those sentenced to county jail, we estimated the release date using the minimum sentenced imposed.

[4] Subsequent to this study, the Commission decided to eliminate county from the scale, although it is included in the statistical models as a control.

[5] Details about the validation of the risk scale are available in "Interim Report 7: Validation of Risk Scale" available online: http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-7-validation-of-risk-scale/view

of three crimes (Burglary, Theft, Drugs) were presented to judges, district attorneys, public defenders, probation officers, and other criminal justice personnel from four counties in Pennsylvania.

Method

Before implementing the risk model the staff had developed, the Pennsylvania Sentencing Commission wanted to test different ways of presenting risk information, in order to determine which method is best understood by those individuals who will incorporate the risk information into the sentencing decision (judges and probation officers), as well as the attorneys (public defenders and assistant district attorneys) who are responsible for making legal arguments about the appropriateness of a criminal sentence. In this study, participants saw one of the six presentation styles of the risk information and were asked to make judgments about each of six cases using this risk information. The results were subsequently presented to focus groups in the four counties in which the study was conducted: Allegheny, Blair, Philadelphia, and Westmoreland. Discussions with a subsample of respondents at these subsequent focus groups were used to better understand the patterns identified in analysis of the survey. The study was approved by the University IRB.

Sample

The target frame consisted of 63 criminal court judges, 449 district attorneys and assistant district attorneys, 248 public defenders, 230 probation officers, and 10 others who worked in one of four counties in Pennsylvania: Allegheny (Pittsburgh), Blair (Altoona), Philadelphia, and Westmoreland (Greensburg). We used a stratified random assignment procedure in order to ensure that across occupations and counties there were approximately an equal number of participants assigned to each of the conditions. That is, within the 20 cells (5 occupations × 4 counties), participants were randomly assigned to the 12 different conditions that are described below.

Emails were sent to the 1000 individuals identified by agency representatives in the four counties. Of these, 38 were returned because the individuals were no longer at the agency or because the address was rejected. We received usable responses from 200 individuals, 21 percent of the 962 individuals who received an email.

The final sample of 200 individuals comprised 79 from Allegheny County (26 percent response rate), 19 from Blair County (53 percent response rate), 75 from Philadelphia County (15 percent response rate), and 27 from Westmoreland County (25 percent response rate).[6] There were 57 district attorneys (13 percent response rate), 24 judges (38 percent response rate), 73 probation officers (32 percent response rate), 30 public defenders (13 percent response rate), and 16 individuals in other positions. Of the 200 individuals, 34 (17 percent) had attended an earlier session at which the risk scale had been presented and discussed.

Procedure

The initial email about the survey was sent out on July 8, 2013. Subsequent reminders were sent out on July 24 and August 9. Data collection was closed on September 1, 2013. Participants received an email from someone in their office that they would be sent a survey in which they would be presented with six cases, including risk information about the offenders. The email from the Research Director of the Pennsylvania Commission on Sentencing is available in Appendix A. Each participant received an email directing him or her to a site in Survey Monkey containing one of the 12 versions of the survey. The case information, based on actual presentence investigation reports, consisted of the types of information typically used by judges: (a) demographic information about the offender (age, sex, race, date of birth); (b) information from the Sentencing Guidelines (offense, Offense Gravity Score, Prior Record Score, guideline recommendation); (c) prior record (juvenile, adult, detainers or charges pending); (d) social history of the offender (marital history, education, employment history, mental health, drug and alcohol history). The information from the presentence investigation reports was condensed into one-page single-spaced summaries for each of the six cases.

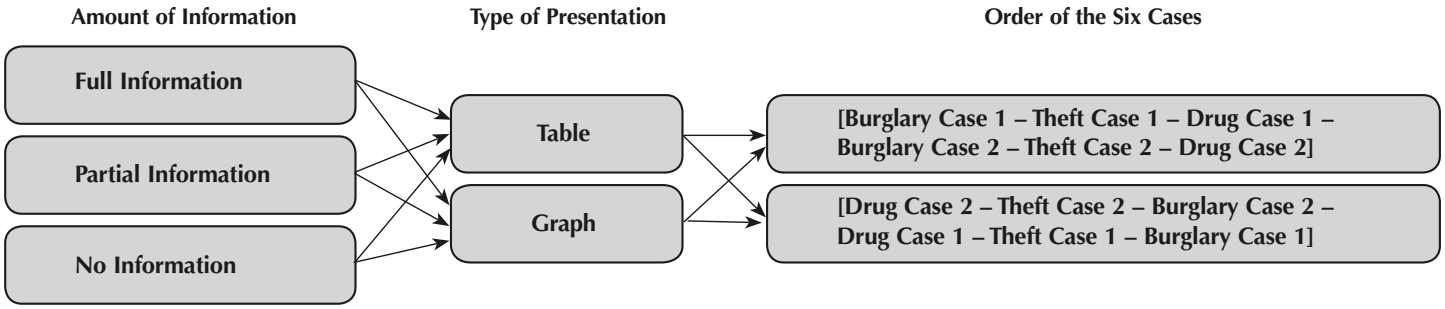
Participants received one of six presentation methods in a 3 × 2 (Amount of Risk Scale Information × Presentation of Recidivism Risk) between-subjects x 3 x 2 (Type of Crime × Cases) within-subjects (repeated measures) design. This mixed-design (between and within subjects) was analyzed using analysis of variance models. There were two between-subjects variables that related to the risk information: amount of information presented about the risk scale (three levels) and presentation of recidivism risk (two levels). The offender's risk score on the overall scale was presented in one of three ways (see Appendix B): (a) the score alone without any further information about the eight factors or the offender's points for each of the eight factors (Risk Score Only); (b) the total score and the number of points for each of the risk factors on which the offender received points (Partial Scale Information); and (c) the total score, the number of possible points for each of the eight categories, and the number of actual points received for each of the eight categories (Full Scale Information). We included the Risk Score Only condition because one of the criticisms of risk assessment scales is that sometimes the people who use them receive only a score, without understanding how that score was arrived at (Hannah-Moffat, 2013).

The risk of recidivism for the offender's risk score was presented in one of two ways, a bar graph or a table (see Appendix C for greyscale version of survey). Each format (graph and table) presented the likelihood of being arrested within three years of release for each of the risk scores (0-14). For the bar graph, the specific offender's likelihood of being arrested within three years of release was highlighted in yellow, and the likelihood of offenders with other risk scores being arrested was shown in blue. For the table, the offender's likelihood of being arrested within three years of release was shown by a number (percentage arrested) and was highlighted in yellow while the recidivism likelihood for offenders with other scores was presented but not highlighted.

After reading summary information about a real case, participants were asked to indicate their judgment of the likelihood that the offender would be arrested within three years of release using a ten-point scale. The scale ranged from 0 to 100 percent and was divided into 10 percent increments (0 - <10 percent, 10 - <20 percent, 20-<30 percent, 30-<40 percent, 40-<50 percent, 50-<60 percent, 60-<70 percent, 70-<80 percent, 80- <90 percent, 90 – 100 percent). For the analyses, this 10-point scale was scored, respectively, as 5 percent, 15 percent, 25 percent, 35 percent, 45 percent, 55 percent, 65 percent, 75 percent, 85 percent, and 95 percent.

[6] An additional 46 individuals started but did not complete the survey (20 individuals dropped out during the first case, an additional 14 dropped out during the second case, an additional 6 dropped out during the third case, and a final 6 dropped out during the fourth case). The responses of all 46 of these individuals were excluded from subsequent analyses.

FIGURE 1.
Design of the Experiment: Three Amounts of Information, Two Types of Presentation, Two Orders of Six Cases



Note. Participants in the study received one of three amounts of information (full, partial, or none) in one of two formats (table or graph). The six cases they reviewed were presented in one of two orders.

After making this risk judgment, participants then saw the risk scale information (Risk Score Only, Partial Scale Information, or Full Scale Information) and the presentation of the recidivism risk (Bar Graph or Table). Following exposure to the risk and contextual information, respondents were again asked to indicate the likelihood that the offender would be arrested within three years. They then indicated the type of sentence they would impose (e.g., prison, jail, probation) and the length of the sentence they would impose. Participants made these same judgments about risk and sentence for each of the six cases.[7]

The six cases were presented in one of two counterbalanced orders: (a) Burglary-1, Theft-1, Drug-1, Burglary-2, Theft-2, Drug-2 or (b) Drug-2, Theft-2, Burglary-2, Drug-1, Theft-1, Burglary-1. In sum, what varied among participants was (a) the presentation of the risk scale (each participant saw one of three ways), (b) the presentation of the recidivism risk (each participant saw one of two ways), and (c) the order in which the six cases were presented (each participant saw one of two ways). Thus, there were 12 different surveys (3 presentations of risk scale x 2 presentations of recidivism information x 2 orders of the six cases). A diagram of the study design is presented in Figure 1. Participants received one of these 12 surveys and made judgments about six cases.

In addition to representing different crimes, the cases represented different actuarial risk levels, as shown by the scores on the 14-point risk scale we created and the associated risk of recidivism within three years (see Table 1). Participants read and made judgments on all six cases.

With regard to the order of presentation of the six cases, across the analyses there were 6 significant effects (of 48 tests involving order).[8] Although this number is above chance, there was no systematic pattern of effects across the three repeated measures analyses of variance that were conducted (one analysis for each set of two cases within a crime type). Thus, we collapsed the other variables across order and do not discuss order further. However, that there were any significant effects indicates that we were correct in counterbalancing order across participants, since punishment judgments can be affected by order (e.g., Pepitone & DiNubile, 1976).

The final set of questions asked participants to make judgments about the six different ways of presenting risk information. Respondents were presented with each of the six methods of presenting risk scale and recidivism information. For each method of presentation, respondents were asked how satisfied they were with the level of detail included in that particular presentation of risk information and how easy it was to understand and interpret each presentation of risk information. For measures of satisfaction, respondents were given a 5-point scale ranging from very dissatisfied to very satisfied. For measures of understanding, respondents were given a 5-point scale ranging from very difficult to very easy. Finally, respondents were asked to rank each of the six methods of presentation in order from most favorite to least favorite. The survey took about 35 minutes to complete.

Results

The results are presented in terms of the four issues that were the focus of the investigation: (a) whether the risk information affects practitioners' judgments; (b) whether the way risk information is presented affects those judgments; (c) whether the effects of actuarial risk information are consistent across crimes and cases; and (d) whether criminal justice practitioners have a preference regarding the presentation of risk information.

Effect of the Risk Information

The effect of the risk information was assessed in two different ways. First, we examined, across all cases and all respondents, the difference between respondents' initial judgment of risk (after the case information) and the final judgment of risk (after the risk information had been presented). Second, we examined the effects of risk information by type of crime and by presentation of the risk information.

Pre/Post difference. Each of the 200 respondents was asked to make pre/post risk judgments about six cases, for a total of 1200 difference scores. Responses were excluded for cases in which the respondent did not provide both a pre and post estimate (36 instances),

[7] The length of the presentence investigations ranged from 14 to 31 pages, about 2-6 pages of which were the state's guidelines forms for the case. Because there is no uniform method for reporting a presentence investigation, there are dramatic differences across jurisdictions. In the one-page summaries, we included those important static factors that were consistently recorded across all counties. Because presentence investigations are conducted in only about one-quarter of criminal cases in Pennsylvania, detailed social histories and validated needs and risk scales are not available statewide and were not used in either the PCS risk scale or in this study. In that these six cases were based on cases with presentence investigation, they may not be representative of cases in general.

[8] For each of the three crimes, there are 4 between-subjects effects involving order, 4 within-subjects effects involving case and order, 4 within-subjects effects involving risk judgments and order, and 4 within-subjects effects involving case, risk judgments, and order. Across the three crimes, there are 48 effects involving order.

TABLE 1.
Descriptive Information, Risk Estimates, and Proposed Sentence for the Six Cases

Case Type	Actuarial Risk Level	Recidivism Rate for Risk Level	Pre Risk Estimate[a]	Post Risk Estimate[a]	t-test[b]	% Prison Sentence Imposed	*M* Length of Prison Sentence Imposed (months)[c]	Actual Prison Sentence (months)	
								Minimum	Maximum
Burglary Case 1	10	69%	72.0%	73.5%	1.73	78%	24.6	48	120
Burglary Case 2	7	47%	63.7%	61.2%	3.51***	53%	15.4	12	36
Theft Case 1	4	26%	34.9%	31.4%	4.89***	28%	11.5	9	24
Theft Case 2	10	69%	75.8%	76.7%	1.27	76%	23.1	15	36
Drug Delivery Case 1	5	33%	47.6%	44.6%	3.39***	34%	12.8	24	48
Drug Delivery Case 2	4	26%	63.7%	50.1%	11.66***	29%	12.4	12	24

Note. Actuarial Risk Level and the Recidivism Rate for Risk Level came from analyses conducted by research staff at the Pennsylvania Commission on Sentencing. The Actual Prison Sentence came from records of the Pennsylvania Commission on Sentencing.

[a]The Pre-Risk and Post-Risk estimates, based on the 10-point scale completed by respondents (0 - <10%, 10 - <20%, 20-<30%, 30-<40%, 40-<50%, 50-<60%, 60-<70%, 70-<80%, 80- <90%, 90 – 100%), were scored, respectively, as 5%, 15%, 25%, 35%, 45%, 55%, 65%, 75%, 85%, and 95%.

[b]Paired sample t-tests were used to test for significant changes in the pre and post risk estimates.

[c]Mean Length of Prison Sentence Imposed in months was converted from the 8-point scale completed by respondents: 1 = 0 months; 2 = < 6 months; 3 = 6-12 months; 4 = 12-18 months; 5 = 18-24 months; 6 = 24-30 months; 7 = 30-36 months; 8 = > 36 months.

*** p < .001

resulting in a final sample of 1164 scores. Only 13 percent of the respondents made no changes in any of the six cases after seeing the risk information. The distribution of respondents by number of changes across the six cases was as follows: 1–18 percent, 2–17 percent, 3–15 percent, 4–17 percent, 5–13 percent, 6–6 percent. On average, respondents changed their risk judgments on 2.61 cases (SE = .18). Of the 1164 possible pre/post judgments, there was a change in 521 (45 percent). Change was most likely for the two drug crimes and somewhat less likely for the two burglaries and the two thefts (see Table 2).

Over all participants, crimes, and cases, there was a significant mean pre/post difference in respondents' judgments, $M = .34$, 95 percent CI [.26, .43], indicating that overall the risk information decreased respondents' judgments of risk. However, these changes were not consistent across respondents or crimes. As can be seen in Table 1, there were significant differences between the pre and post risk judgments on four of the six crimes. For all four of these crimes, the post-information mean was smaller and closer to the actuarially determined rate of recidivism than was the pre-information mean, indicating that in general respondents' judgments were influenced in the direction indicated by the actuarial information. Notably, there was no significant difference between the pre-information mean and the post-information mean for offenders with the highest risk score.

TABLE 2.
Descriptive Information: Changes in Pre and Post Information Risk Estimates

	Total Responses	Estimates Changed Pre/Post		Estimates Unchanged Pre/Post	
		N	%	N	%
Burglary Case 1	195	77	39.49	118	60.51
Burglary Case 2	193	82	42.49	111	57.51
Theft Case 1	194	80	41.24	114	58.76
Theft Case 2	193	63	32.64	130	67.36
Drug Case 1	193	96	49.74	97	50.26
Drug Case 2	196	123	62.76	73	37.24
Total	**1164**	**521**	**44.76**	**643**	**55.24**

Effects of presentation type and type of crime on ratings of risk. Aside from knowing whether the actuarial risk information affected ratings of risk overall, we tested whether the effect of the risk information differed depending on how it was presented in terms of the amount of risk information (none, partial, full) and the way the recidivism information was presented (table or graph). Analyses were conducted within each of the three types of crimes using a doubly repeated measures (2 risk judgments for each of 2 cases) analysis of variance. The between-subjects factors were amount of risk information (none, partial, full) and the type of presentation of the meaning of the risk score for recidivism (table or graph). No significant effects were found for the amount of information provided about the risk scale, and this variable is consequently not discussed further.

For Burglary, there were no significant between-subjects effects. Regarding within-subjects effects, there was a significant effect for the two burglary cases, $F(1, 194) = 52.47$, $p = .000$, $\eta_p^2 = .21$, indicating that there was a significant difference in perceived risk in the two cases ($M_1 = 7.25$, SE = .13 vs. $M_2 = 6.23$, SE = .13). The effects for the two burglary cases

were conditioned by two significant interactions. First there was a significant Burglary Cases x Risk Judgments interaction, $F(1, 194) = 17.66$, $p = .000$, $\eta_p^2 = .08$ (see Table 1), such that for the first case there was a slight increase from pre to post ($M_{pre} = 7.17$, SE = .14 and $M_{post} = 7.34$, SD = .13), whereas for the second case there was a slight decrease ($M_{pre} = 6.37$, SE = .15 and $M_{post} = 6.09$, SE = .13). Second, there was a significant interaction of Burglary Case x Risk Presentation Type, $F(1, 194) = 4.29$, $p < .04$, $\eta_p^2 = .02$. For the first case, there was greater judged risk when recidivism information was presented using a graph rather than a table ($M_{graph} = 7.42$, SE = .18 and $M_{table} = 7.09$, SE = .19). In contrast, for the second case, there was more judged risk when recidivism information was presented using a table rather than a graph ($M_{graph} = 6.10$, SE = .19 and $M_{table} = 6.36$, SE = .19).

For Theft, there were no significant between-subjects effects. Regarding within-subjects effects, there was a significant effect for the two theft cases, $F(1, 190) = 634.81$, $p = .000$, $\eta_p^2 = .77$, indicating that there was a significant difference in perceived risk in the two cases ($M_1 = 3.33$, SE = .13 vs. $M_2 = 7.61$, SE = .13). There was a significant effect for risk scores, $F(1, 190) = 9.22$, $p = .003$, $\eta_p^2 = .05$, such that the pre score was higher than the post score ($M_{pre} = 5.54$, SE = .11 vs. $M_{post} = 5.40$, SE = .09). Both of the main effects were conditioned by a significant Theft Cases x Risk Judgments interaction, $F(1, 190) = 18.00$, $p = .000$, $\eta_p^2 = .09$ (see Table 1), such that for the first case there was a slight decrease from pre to post ($M_{pre} = 3.51$, SE = .15 to $M_{post} = 3.15$, SE = .13), whereas for the second case there was a slight increase ($M_{pre} = 7.57$, SE = .14 to $M_{post} = 7.65$, SE = .12).

For the Drug crimes, there were no significant between-subjects effects. Regarding within-subjects effects, there was a significant effect for the two drug cases, $F(1, 192) = 42.75$, $p = .000$, $\eta_p^2 = .18$, indicating that there was a significant difference in perceived risk in the two cases ($M_1 = 4.58$, SE = .15 vs. $M_2 = 5.70$, SE = .15). There was also a significant effect for the two risk judgments, $F(1, 192) = 97.50$, $p = .000$, $\eta_p^2 = .34$, such that the pre score was higher than the post score ($M_{pre} = 5.57$, SE = .13 vs. $M_{post} = 4.72$, SE = .13). Both of the main effects were conditioned by a significant Drug Cases x Risk Judgments interaction, $F(1, 192) = 61.50$, $p = .000$, $\eta_p^2 = .24$ (see Table 1), such that for the first case there was a decrease from pre to post ($M_{pre} = 4.75$, SE = .17 to $M_{post} = 4.42$, SE = .14), whereas for the second case there was a large decrease ($M_{pre} = 6.39$, SE = .16 and $M_{post} = 5.02$, SE = .16).

In general, then, there were differences between cases within crimes and differences between pre and post risk judgments. But the amount of risk information did not affect any of the post risk judgments, and the presentation of the recidivism information affected only the post judgments for the two burglary crimes.

Relationship to actual sentences. One of the fears of providing actuarial risk information at sentencing is that there will be an increase in punishment severity (Hannah-Moffat, 2013). To test that notion, we examined the percentage of individuals who said they would incarcerate the individual. In actuality, all six individuals had been incarcerated in state prison. As can be seen in Table 1, among respondents the incarceration rates for the six cases ranged from 28 percent to 78 percent, and two of the mean incarceration sentences were for less time than was actually imposed. Thus, these data suggest that actuarial risk information does not necessarily increase punishment severity, an initial conclusion that warrants further research.

Preference for How Risk Information is Presented

At the end of the survey, respondents were shown all six combinations of risk information and the meaning of the risk information used in this study. They were then asked to rank the six combinations in terms of their preference for how the information should be presented at sentencing. As shown in Table 3, respondents showed two clear preferences: (a) a preference for more information about the risk scale: full information over partial information over no information, and (b) a preference for the graph over a table, within each one of those information levels. Ratings of understanding were related only to the level of information about the risk scale: full information over partial information over no information. Ratings of satisfaction followed the same pattern regarding level of information, although within the full information and partial information categories, respondents said they were more satisfied with the table than with the graph.

Discussion

This study was designed to test the impact of actuarial risk information on decision makers' judgments of risk and to examine whether these effects were consistent across crime types and across cases within crime types.

TABLE 3.
Ratings of Six Different Ways of Presenting the Risk Information and the Meaning of the Risk Information

Amount of Risk Information	Presentation of Recidivism Information	Mean Rank		Satisfaction		Understanding	
		Mean	SE	Mean	SE	Mean	SE
Full	Graph	4.88^{f}	(.104)	3.46^{d}	(.067)	3.78^{c}	(.053)
Full	Table	4.46^{e}	(.101)	3.54^{e}	(.070)	3.71^{c}	(.062)
Partial	Graph	3.69^{d}	(.094)	2.98^{b}	(.075)	3.44^{b}	(.068)
Partial	Table	3.26^{c}	(.096)	3.13^{c}	(.074)	3.47^{b}	(.064)
None	Graph	2.68^{b}	(.104)	2.48^{a}	(.076)	2.96^{a}	(.081)
None	Table	2.35^{a}	(.117)	2.55^{a}	(.080)	3.10^{a}	(.080)
	F	94.88***	$\eta_p^2 = .348$	69.33***	$\eta_p^2 = .270$	37.45***	$\eta_p^2 = .169$
	N	179		188		185	

Note. For rankings, higher numbers indicate greater preference. For ratings of satisfaction and understanding, higher numbers indicate, respectively, higher satisfaction and greater understanding. Within a column, means with different superscripts are significantly different according to a post-hoc Newman-Keuls test ($p < .05$).

*** $p < .001$

Effect of Risk Score Information

Even though the risk score information significantly affected mean risk judgments overall and in four of the six cases, only 45 percent of all possible decisions were affected by the risk information. Moreover, about a third of the sample changed no judgments or only one risk judgment, and almost half changed two or fewer of the six judgments. Given the fear that actuarial risk information would overwhelm all other information in risk judgments, it is somewhat surprising that the risk score information did not have stronger effects on the post-risk-presentation ratings.

There are four possible reasons why the risk information may not have had a stronger effect. First, the participants knew the study was a simulation and they may not have taken the study seriously. As with all simulations, this possibility cannot be discounted. Second, it is possible that the participants did not understand the risk information and therefore were not influenced by it. This possibility is unlikely, however, in that the average movement was in the direction toward that indicated by the actuarial information. Third, the participants likely considered themselves to be experts, and, as such, they would be likely to discount other information. This explanation would be an example of resistance to using actuarial information (Elstein, 1976). In this study, the respondents may have thought that they were considering cases that were exceptions to the general information presented in the actuarial recidivism scale (i.e., what Meehl calls the "broken leg" problem). Fourth, in the presentence report the participants already had all of the information that was used in the actuarial scale presented to them (i.e., age, gender, prior record, offense severity, county), and they may have believed that the scale thus added no new information.

One of the concerns about the use of risk scales, which have the appearance of scientific validity, is that it would be too determinative of the final outcome (Hannah-Moffat, 2013). Our results suggest that is not the case. For the cases we used, although overall the risk information tended to reduce respondents' judgments of risk, the resulting reductions were small in magnitude (an average change of 4.2 percent across the six cases, ranging from .9 percent change to 13.6 percent change). This differential effect across cases suggests that respondents were using the information appropriately for individual cases, rather than being overwhelmed by the actuarial information.

The inconsistent effects of the risk information across the three crimes and the two cases within each crime type also suggest that it is incorrect to say that actuarial risk information has a single effect. Rather, pending further research, it appears that decision makers consider it differently for different crimes and different cases.

Preference for and Effect of Presentation of Information

One of Hannah-Moffat's (2013) fears was that judges would receive only summary scores of the risk scale and would therefore not understand the underlying basis for the score. The results of this study suggest, consistent with Hannah-Moffat's concern, that judges, attorneys, and other criminal justice personnel prefer full information about the risk scale. Moreover, the possible concern about full information (i.e., that it would confuse people) was not borne out, in that there was no difference in the post-risk judgments with respect to the amount of risk scale information presented (none, partial, full), although low statistical power is a possible explanation for the absence of difference.

There was only one significant effect of the type of presentation (table or graph) of recidivism information (on the two burglary crimes, but not on the two theft crimes or the two drug crimes). Our conclusion would be that although respondents had a clear preference for full information about the risk scale and a preference for the graph over the table, in general these factors have little effect on judgments of risk.

Subsequent Focus Groups

About six months after the survey, we presented the results of the survey to focus groups of 15-20 individuals in each of the four counties and asked participants to comment on the findings. At all four focus groups, attendees agreed with the finding that the full risk information should be presented, rather than only a partial amount of information or only the summary risk score. Similarly, the general sense of participants was a slight preference for a graph over a table.

The real question to focus group participants was how the risk scale should be used at sentencing. Three options were presented: (a) as simply another piece of information to be used at sentencing, (b) as information incorporated into the Sentencing Guidelines, or (c) as explicitly mitigating or aggravating information. For the most part, participants suggested that the scale should be just another piece of information that judges consider at sentencing.

In part, this preference for a limited use of the actuarial risk scale was because of a concern, voiced at all four focus groups, about biases that may be built in to the recidivism scale (e.g., arrest bias against certain groups). Although judges, district attorneys, and probation officers are almost certainly aware of these possible biases inherent in the data (see Hannah-Moffat, 2013), the focus group discussions suggest that it might be worthwhile to remind decision makers about the limitations of actuarial scales.

Policy-Relevant Research

This study was initiated by the Commission in order to help them determine how best to communicate risk information to judges, attorneys, and probation officers. Our research suggests that these decision makers want complete information, that they generally understand the concepts and findings, and that they are not overwhelmed by the improved accuracy of actuarial over clinical predictions. Our finding that the actuarial risk information tended to lower respondents' judgments of risk may be unique to the cases we used in the study and must be replicated with other cases and other crimes. In addition, the next step is to test how the presentation of actuarial risk information is used in actual decisions.

References

Buchanan, A. (2013). Violence risk assessment in clinical settings: Being sure about being sure. *Behavioral Sciences & the Law*, *31*(1), 74-80.

Campbell, M. A., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism for adult offenders*. Ottawa, ON: Public Safety Canada.

Casey, P. M., Warren, R. K., & Elek, J. K. (2011). Using offender risk and needs assessment information at sentencing: Guidance for courts from a national working group. Williamsburg, VA: National Center for State Courts. http://www.ncsc.org/~/media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Sentencing%20Probation/RNA%20Guide%20Final.ashx

Elstein, A. S. (1976). Clinical judgment: Psychological research and medical practice. *Science*, *194*, 696-700. doi: 10.1126/science.982034

Gottfredson, D. M., & Snyder, H. N. (2005). The mathematics of risk classification: Chang-

ing data into valid instruments for juvenile courts. NCJ 209158. Washington, DC: Office of Juvenile Justice and Delinquency Prevention. https://www.ncjrs.gov/pdffiles1/ojjdp/209158.pdf

Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime and Delinquency*, *51*, 178-200. doi:10.1177/0011128705281748

Hannah-Moffat, K. (2013). Actuarial sentencing: An "unsettled" proposition. *Justice Quarterly*, *30*, 270-296. doi:10.1080/07418825.2012.682603

Hanson, R. K., & Morton-Bourgon, K. (2004). Predictors of sexual recidivism: An updated meta-analysis 2004-02. *Ottawa, Canada: Public Safety and Emergency Preparedness Canada*.

Harris, G. T., Lowenkamp, C. T., & Hilton, N. Z. (2015). Evidence for risk estimate precision: implications for individual risk communication. *Behavioral Sciences & the Law*, *33*, 111-127.

Harris, G. T., & Rice, M. E. (2015). Progress in violence risk assessment and communication: Hypothesis versus evidence. *Behavioral Sciences & the Law*, *33*(1), 128-145.

Hart, S. D., & Cooke, D. J. (2013). Another look at the (im-)precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences and the Law*, *31*, 81-102. doi: 10.1002/bsl.2049

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*, *34*(3), 400-409.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Meehl, P. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, *50*, 370-375. doi:10.1207/s15327752jpa5003_6

Monahan, J., Heilbrun, K., Silver, E., Nabors, E., Bone, J., & Slovic, P. (2002). Communicating violence risk: Frequency formats, vivid outcomes, and forensic settings. *International Journal of Forensic Mental Health*, *1*, 121-126. doi:10.1080/14999013.2002.10471167

Monahan, J., & Skeem, J. L. (2013). Risk redux: The resurgence of risk assessment in criminal sanctioning. Presentation at the annual meeting of the National Association of Sentencing Commissions, Minneapolis.

Monahan, J., & Steadman, H. J. (1996). Violent storms and violent people: How meteorology can inform risk communication in mental health law. *American Psychologist*, *51*, 931-938. doi:10.1037/0003-066X.51.9.931

Mossman, D. (2015). From group data to useful probabilities: The relevance of actuarial risk assessment in individual instances. *Journal of the American Academy of Psychiatry and the Law Online*, *43*(1), 93-102.

Pepitone, A., & DiNubile, M. (1976). Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology*, *33*, 448-459. doi:10.1037/0022-3514.33.4.448

Ruback, R. B., & Innes, C. A. (1988). The relevance and irrelevance of psychological research: The example of prison crowding. *American Psychologist*, *43*, 683-693.

Sanfey, A., & Hastie, R. (1998). Does evidence presentation format affect judgment? An experimental evaluation of displays of data for judgments. *Psychological Science*, *9*, 99-103. doi:10.1111/1467-9280.00018

Scurich, N., & John, R. S. (2011). Prescriptive approaches to communicating the risk of violence in actuarial risk assessment. *Psychology, Public Policy, and Law*, *18*, 50-78. doi:10.1037/a0024592

Scurich, N., & John, R. S. (2012). A Bayesian approach to the group versus individual prediction controversy in actuarial risk assessment. *Law and Human Behavior*, *36*, 237-246. doi:10.1037/h0093973

Scurich, N., Monahan, J., & John, R. S. (2012). Innumeracy and unpacking: Bridging the nomothetic/idiographic divide in violence risk assessment. *Law and Human Behavior*, *36*, 548-554. doi:10.1037/h0093994

Skeem, J. (2013). Risk technology in sentencing: Testing the promises and perils (Commentary on Hannah-Moffat, 2011). *Justice Quarterly*, *30*, 297-303. doi:10.1080/07418825.2012.687513

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24*, 271-296. doi:10.1023/A:1005595519944

Appendix A: Survey recruitment email issued by the Research Director of the Pennsylvania Commission on Sentencing to potential respondents.

Act 95 of 2010 mandated the Pennsylvania Commission on Sentencing to develop a risk assessment instrument to assist the court at sentencing. To address this new mandate, the Commission undertook a Risk Assessment study to determine what factors best predict which offenders will be rearrested for a new crime. This study involved over 18,000 offenders from Levels 3 and 4 of the sentencing guidelines who were sentenced during 2004-2006. Eight factors were found to be the best predictors of rearrest: gender, age, county, number of prior adult arrests, prior property arrest, prior drug arrest, property offender, and offense gravity score. These eight factors were used to develop a Risk Assessment Scale to identify offenders at low risk of rearrest. The Scale resulted in risk scores ranging from 0 to 14, with low risk being defined by the Commission as a score of 0 to 4.

The Pennsylvania Commission on Sentencing is conducting this survey to determine how best to present risk information. This survey will take about 30 minutes to complete. You will be presented with 6 case scenarios and corresponding risk information and asked to answer 4 questions per scenario. After completing the 6 scenarios, you will be asked to compare different presentations of risk information. Your participation is voluntary. You can stop at any time and you do not have to answer any questions you do not want to answer.

Your participation is confidential. The survey does not ask for any information that would identify you or allow us to link you to your responses. In the event of any publication or presentation resulting from the research, information will be presented only in large categories of people. Please contact *** with any questions about this survey. We thank you in advance for your participation.

Appendix B. Amount of Information About the Risk Scale Presented with the Risk Score

OPTION 1
Risk Score only

This offender has a risk score of 10.

The Commission has determined risk scores 0-4 to be low risk.

OPTION 2
Risk Score with partial information

This offender has a risk score of 10.

The Commission has determined risk scores 0-4 to be low risk. Below is the calculation of the offender's risk score based on the 8 identified risk factors. Displayed is the number of actual points received by the offender. The offender's total risk score is the sum of points received across all 8 risk factors. The sum ranges from 0-14.

	Risk *Actual Points*
Gender	
Male	1
Age	
30-49	1
County	
Semi-urban	1
Number Prior Adult Arrests	
13+	4
Prior Property Arrest	
Yes	1
Prior Drug Arrest	
Yes	1
Current Property Conviction	
Yes	1
Offense Gravity Score [OGS]	
4+	0
Total Risk Score	10

OPTION 3
Risk Score with full information

This offender has a risk score of 10.

The Commission has determined risk scores 0-4 to be low risk. Below is the calculation of the offender's risk score based on the 8 identified risk factors. Displayed is the number of possible points for each risk factor and the number of actual points received by the offender. The offender's total risk score is the sum of points received across all 8 risk factors. The sum ranges from 0-14.

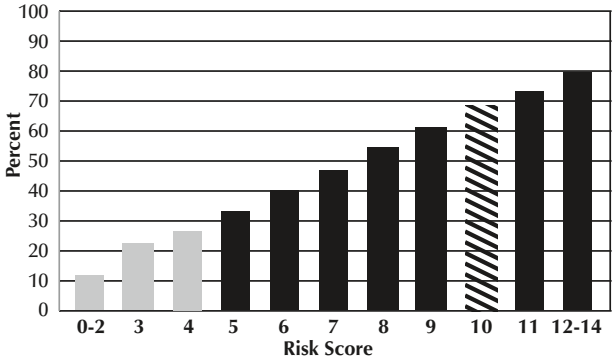
Risk Scale	*Possible Points*	*Actual Points*
Gender		
Male	1	1
Female	0	
Age		
Less than 24	3	
24-29	2	
30-49	1	1
50+	0	
County		
Rural	0	
Semi-urban	1	1
Urban	2	
Number Prior Adult Arrests		
0	0	
1	1	
2-4	2	
5-12	3	
13+	4	4
Prior Property Arrest		
No	0	
Yes	1	1
Prior Drug Arrest		
No	0	
Yes	1	1
Current Property Conviction		
No	0	
Yes	1	1
Offense Gravity Score [OGS]		
1-3	1	
4+	0	0
Total Risk Score		**10**

Appendix C. Types of Presentation of Recidivism Rates for Different Risk Scores

OPTION 1
Graph

The graph below depicts the offender's likelihood of being arrested within 3 years of release from incarceration or imposition of probation/county IP (striped bar) compared to other offenders with different risk scores. The low risk scores are highlighted in grey.

Percentage of Offenders Arrested within 3 Years of Release from Incarceration or Imposition of Probation/County IP by Risk Score



OPTION 2
Table

The table below displays the offender's likelihood of being arrested within 3 years of release from incarceration or imposition of probation/county IP (highlighted in dark grey) compared to other offenders with different risk scores. The low risk scores are highlighted in light grey.

Percentage of Offenders Arrested within 3 Years of Release from Incarceration or Imposition of Probation/ County IP by Risk Score

Risk Score	Percent Arrested
0-2	12
3	23
4	26
5	33
6	40
7	47
8	55
9	61
10	69
11	73
12-14	80