

# Supplementary Appendix for Imai, K., Z. Jiang, D. J. Greiner, R. Halen, and S. Shin. “Experimental Evaluation of Algorithm-Assisted Human Decision-Making: Application to Pretrial Public Safety Assessment.”

## S1 Distribution of Judge’s Decisions Given the PSA for Subgroups S1.1 Female Arrestees

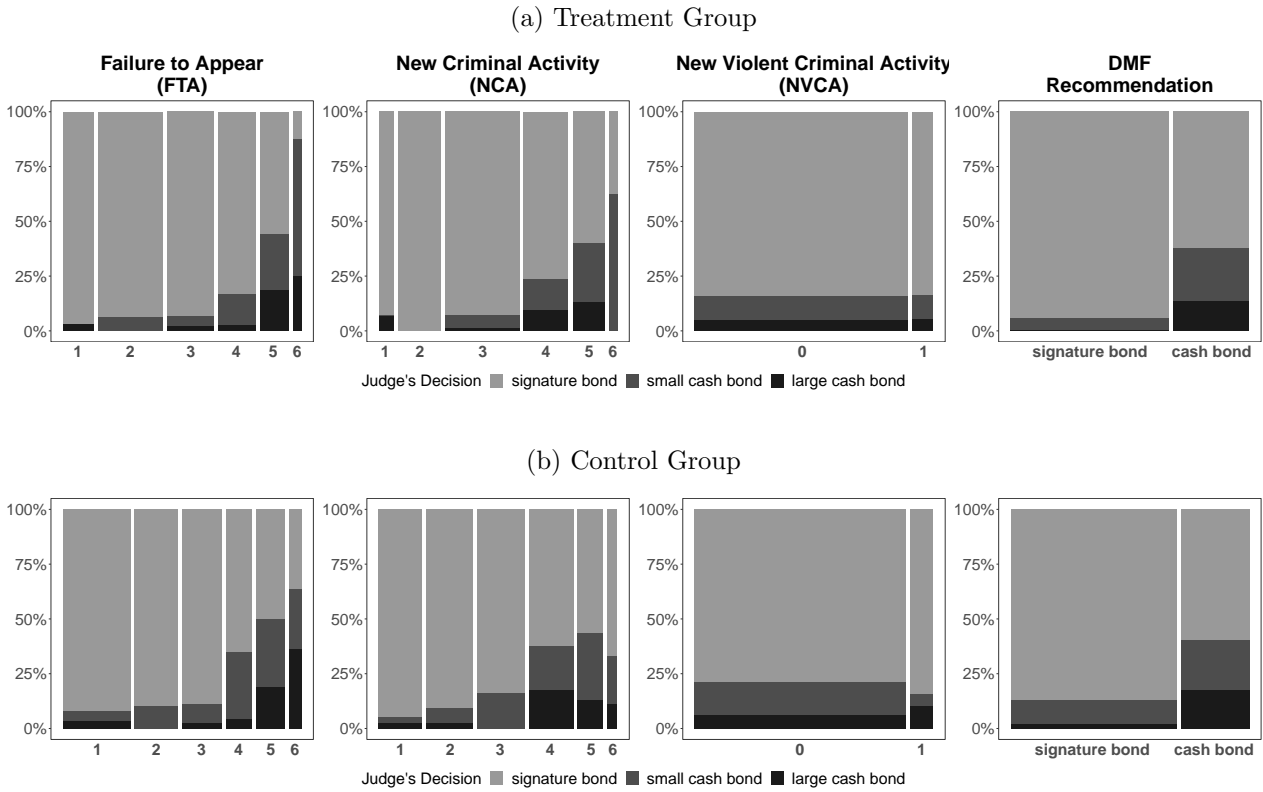


Figure S1: The Distribution of Judge’s Decisions Given the Pretrial Public Safety Assessment (PSA) among the Cases in the Treatment (Top Panel) and Control (Bottom Panel) Groups Among Female Arrestees.

## S1.2 Non-white Male Arrestees

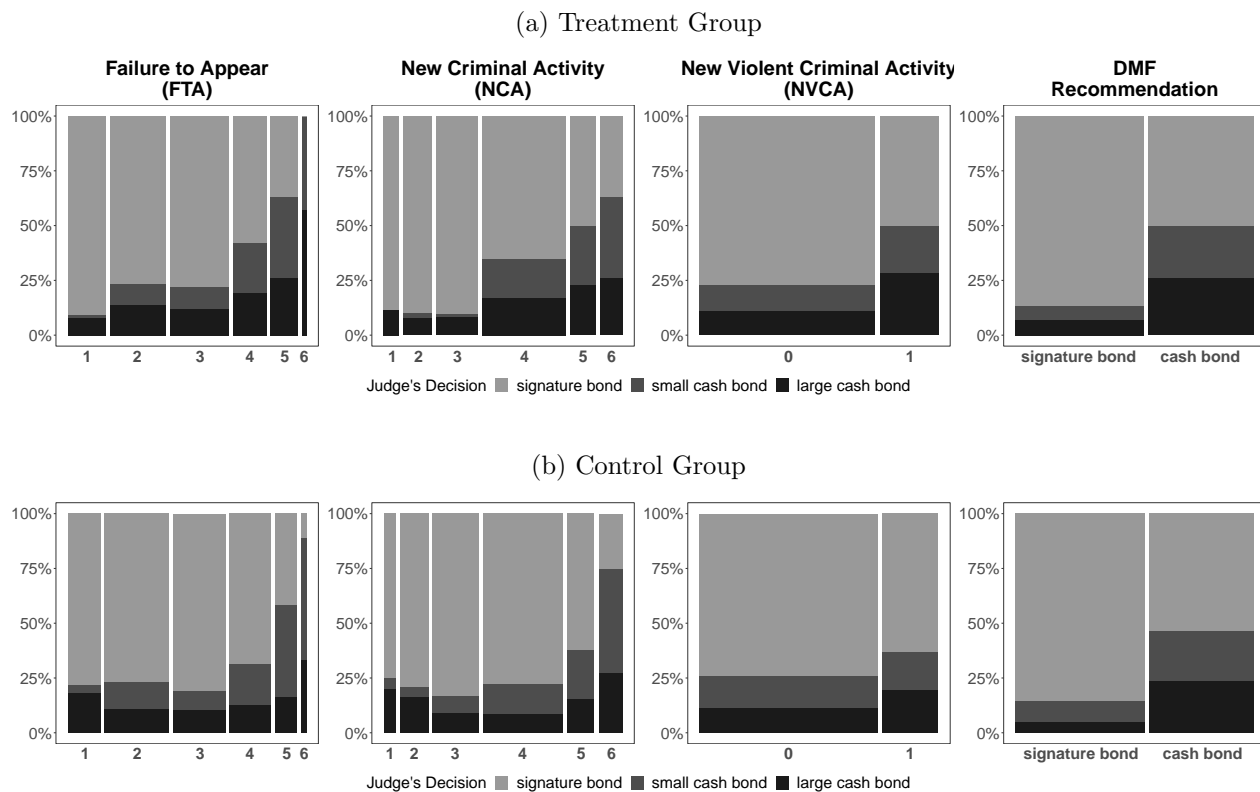


Figure S2: The Distribution of Judge's Decisions given the Pretrial Public Safety Assessment (PSA) among the Cases in the Treatment (Top Panel) and Control (Bottom Panel) Groups Among Non-white Male Arrestees.

### S1.3 White Male Arrestees

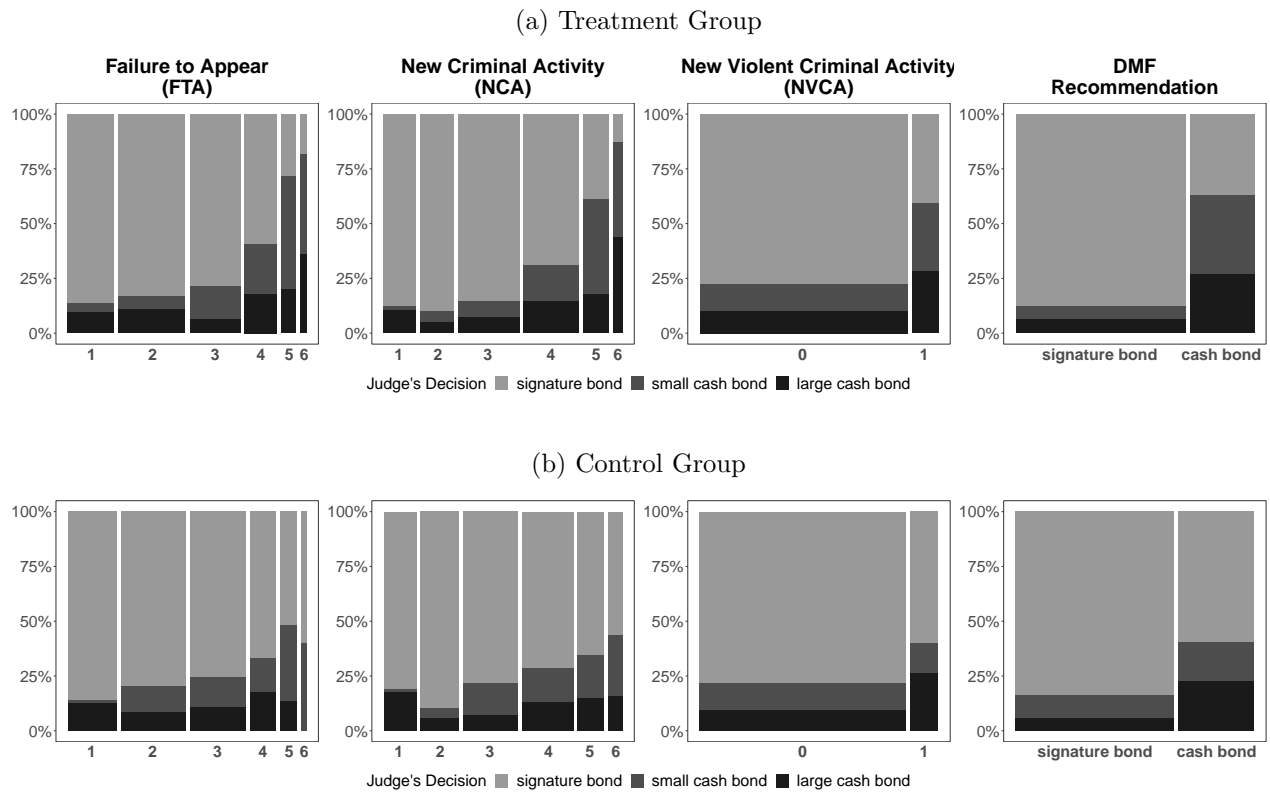


Figure S3: The Distribution of Judge's Decisions Given the Pretrial Public Safety Assessment (PSA) among the Cases in the Treatment (Top Panel) and Control (Bottom Panel) Groups Among White Male Arrestees.

## S2 Subgroup Analysis for Age Groups

In this appendix, we conduct the subgroup analysis for different age groups.

### S2.1 Age Distribution, Descriptive Statistics, and Average Causal Effects

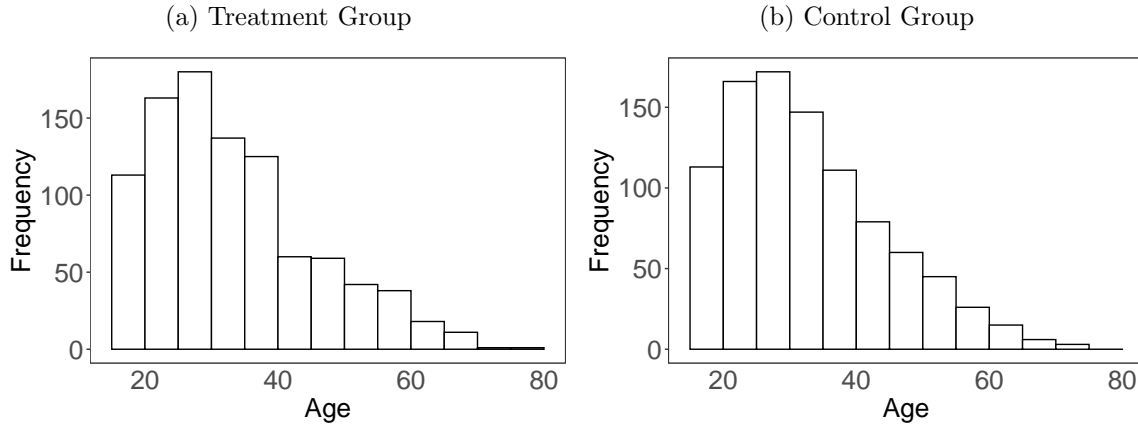


Figure S4: The Distribution of Age in the Treatment (Left Panel) and Control (Right Panel) Groups Among Arrestees.

	<i>no</i> PSA			PSA			Total (%)
	Signature bond	Cash bond $\leq$ \$1000	Cash bond $>$ \$1000	Signature bond	Cash bond $\leq$ \$1000	Cash bond $>$ \$1000	
22 or below	135 (7.1)	24 (1.3)	22 (1.2)	136 (7.2)	24 (1.3)	16 (0.8)	357 (18.9)
23 – 28	158 (8.4)	25 (1.3)	23 (1.2)	148 (7.8)	29 (1.5)	28 (1.5)	411 (21.7)
29 – 35	157 (8.3)	40 (2.1)	14 (0.7)	151 (8.0)	33 (1.7)	28 (1.5)	423 (22.3)
36 – 45	142 (7.5)	22 (1.2)	26 (1.4)	133 (7.0)	30 (1.6)	22 (1.2)	375 (19.9)
46 or above	113 (6.0)	21 (1.1)	21 (1.1)	137 (7.2)	14 (0.7)	19 (1.0)	325 (17.1)

Table 2: The Joint Distribution of Treatment Assignment, Decisions, and Age. The table shows the number of cases in each category with the corresponding percentage in parentheses.

Figure S4 presents the distribution of age for the treatment and control groups. As expected, the two distributions are similar. We observe that the age distribution is right skewed with many more young arrestees. Table 2 presents the descriptive statistics for different age groups examined here. We divide the arrestees into five subgroups with different ranges of age (aged 22 or below, between 23 to 28, between 29 to 35, between 36 to 45, 46 or above). Within each age group, the signature bond appears to be the dominant decision.

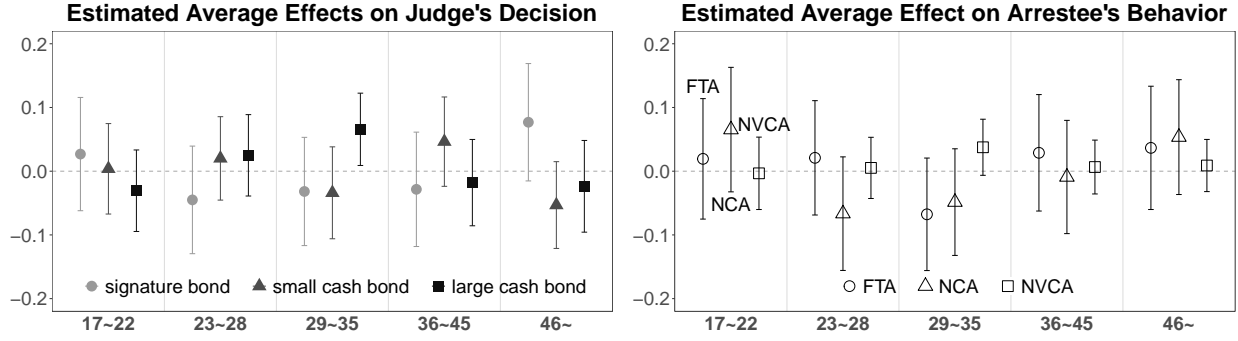


Figure S5: Estimated Average Causal Effects of PSA Provision on Judge’s Decisions and Outcome Variables for First Arrest Cases (FTA, NCA, and NVCA). The results are based on the difference-in-means estimator. The vertical bars represent the 95% confidence intervals. In the left figure, we report the estimated average causal effect of PSA provision on the decision to charge a signature bond (circles), a small cash bail (\$1,000 dollars or less; triangles), and a large cash bail (greater than \$1,000; squares). In the right figure, we report the estimated average causal effect of PSA provision on the three different outcome variables: FTA (open circles), NCA (open triangles), and NVCA (open squares).

Figure S5 presents the estimated Intention-to-Treat (ITT) effects of PSA provision on judge’s decisions (left panel) and arrestee’s behaviors (right panel). We find that PSA provision has little effect on the judge’s decisions with the exception of the 29 – 35 years old group and the oldest group. For the 29 – 35 years old group, the PSA appears to lead to a harsher decision while for the 46 or older group the effect is the opposite. As for the effects on arrestee’s behavior, our analysis suggests that PSA provision may increase NVCA among the 29 – 35 years old group though the estimate is only marginally significant.

## S2.2 Principal Stratum Proportion and Average Principal Strata Effects

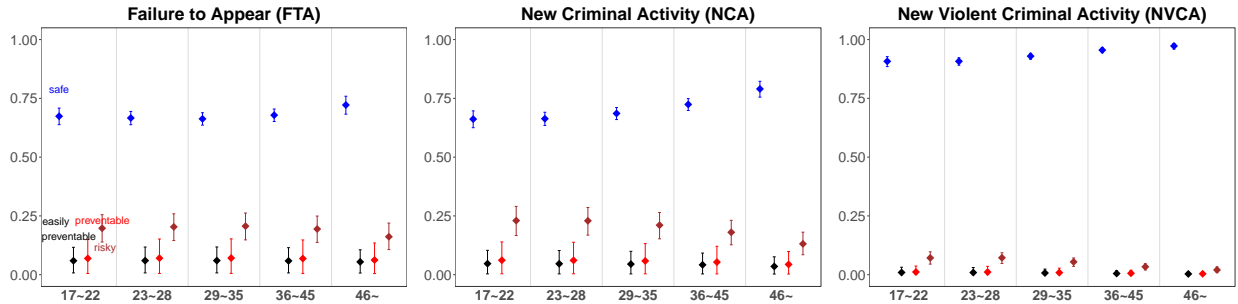


Figure S6: Estimated Proportion of Each Principal Stratum. Each plot represents the result using one of the three outcome variables (FTA, NCA, and NVCA), where the blue, black, red, and brown diamonds represent the estimates for safe, easily preventable, preventable, risky cases, respectively. The solid vertical lines represent the 95% Bayesian credible intervals.

Figure S6 presents the estimated proportion of each principal stratum for different age groups. We observe that the principal stratum size is similar across age groups with the safe cases being the most dominant. The proportion of safe cases appears to be greater for older age groups though the rate of increase is modest. The interpretation of Figure S7 is given in the last paragraph of Section 4.2.

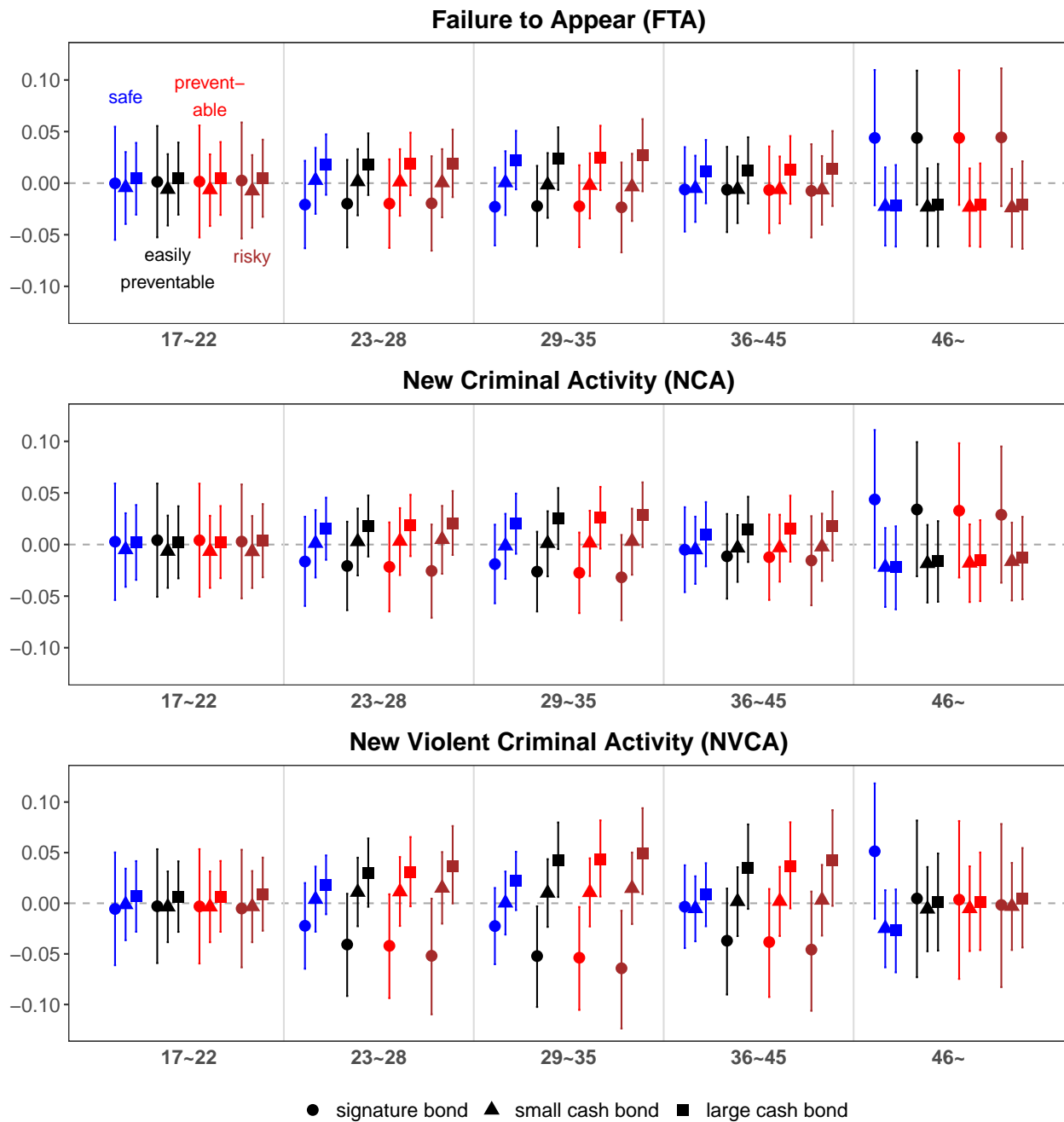


Figure S7: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision. Each panel presents the age group-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the Bayesian 95% credible interval.

## S3 Testing the Potential Existence of Spillover Effects

### S3.1 Conditional Randomization Test

We examine the possible existence of spillover effects. In particular, we use a conditional randomization test to examine whether or not PSA provision of prior cases affects the judge’s decision in later cases (e.g., Aronow, 2012; Athey, Eckles and Imbens, 2018; Candès et al., 2018). The basic idea is to test whether the decision,  $D_i$ , is conditionally independent of the treatment assignment of the other cases whose court hearing date is prior to that of case  $i$ , given its own treatment assignment status  $Z_i$ . The judge made decision for 1,891 cases on 274 different dates. Unfortunately, we do not have information about the ordering of decisions within each hearing date. Let  $O_i \in \{1, 2, \dots, 274\}$  denote the order of the hearing date of case  $i$ . Let  $\tilde{Z}_i = |\{i' \in \{1, 2, \dots, n\} : O_{i'} = O_i - 1\}|$  denote the proportion of treated cases whose hearing date order is immediately before that of case  $i$ . Then, the null hypothesis is given by  $H_0 : \tilde{Z}_i \perp\!\!\!\perp D_i \mid Z_i$ . We conduct a conditional randomization test as follows:

1. Create a new treatment assignment  $Z'_i$  as follows:

- (a) For each  $i$ , if  $O_i$  is even then  $Z'_i = Z_i$
- (b) For each  $i$ , if  $O_i$  is odd then randomly sample  $Z'_i \sim \text{Bernoulli}(1/2)$

Then compute  $\tilde{Z}'_i$  based on  $Z'_i$ , i.e.,  $\tilde{Z}'_i = |\{i' \in \{1, 2, \dots, n\} : O_{i'} = O_i - 1\}|$ .

2. Regress  $D_i$  on  $(1, Z_i, \tilde{Z}'_i)$  only using the subset of observations whose  $O_i$  is even. Let our test statistic  $T$  be the squared term of estimate of coefficient of  $\tilde{Z}'_i$ .
3. Repeat the above  $S$  times and compute (one-sided) p-value:  $\frac{1}{S} \sum_{s=1}^S \mathbf{1}\{T^{(s)} \geq T_{\text{obs}}\}$  where  $T^{(s)}$  is the test statistic for the  $s$ th iteration and  $T_{\text{obs}}$  is the observed test statistic.

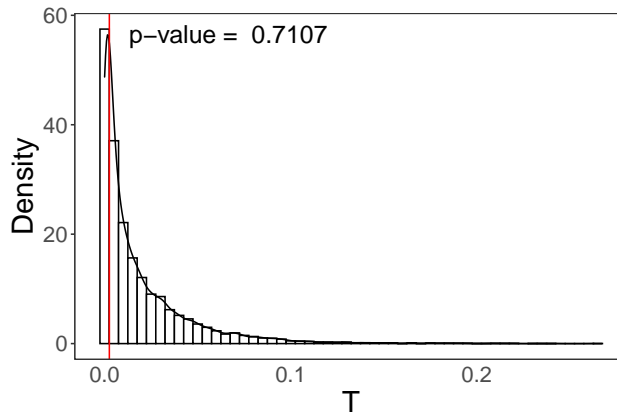


Figure S8: The Distributions of Test Statistics. The red vertical lines indicate the observed test statistics.

Figure S8 presents the resulting distribution of our test statistics. The  $p$ -value is 0.71 for the test statistics  $T$ , and thus we fail to reject the null hypothesis. That is, we find no statistically significant evidence that the judge’s decision is influenced by PSA provision of the prior cases. This is consistent with the assumption of no inference among the cases, which is made throughout our analysis.

### S3.2 Power: A Simulation Analysis

We examine the power of the statistical test used above via a simulation study. Our simulation procedure is as follows:

1. Regress  $D_i$  on  $(1, Z_i, \tilde{Z}_i)$  using the ordinal logistic regression model based on the observed data. Let  $\omega$  denote the coefficient for  $\tilde{Z}_i$ .
2. Choose a value of  $\omega$ , and set the other model parameters to their estimated values. Using this mode, generate  $D_i$  with the same sample size and observed treatment variable.
3. Conduct the conditional randomization test as described in Section S3.1. Repeat this for 1,000 times and calculate the proportion of rejecting the null hypothesis at the 0.05 level.
4. Repeat the above procedure for each value of  $\omega \in \{-1.5, -1, -0.5, 0, 0.5, 1, 1.5\}$ .

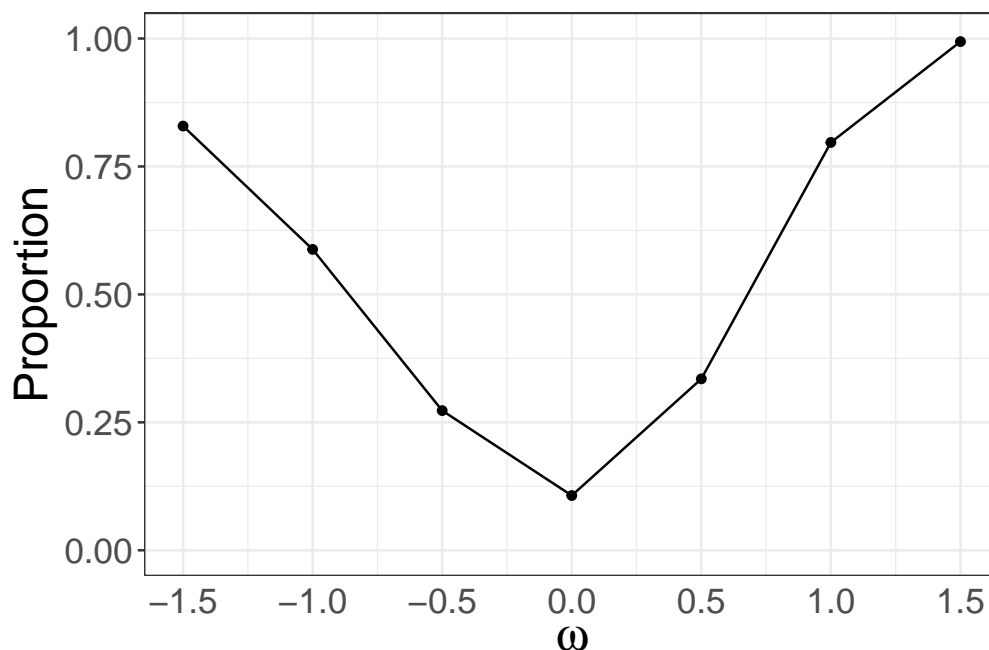


Figure S9: The Proportion of Rejecting the Null Hypothesis at the 0.05 Level.

Figure S9 presents the results of our simulation study for calculating the power of the test. Here, if the proportion of treated cases whose hearing date order is immediately before is 1, the odds of judges making harsher decision is  $\exp \omega$  times that of the arrestees whose proportion of treated cases whose hearing date order is immediately before is 0. According to the simulation, the power of the test reaches about 0.8 when  $\omega = 1$  or  $\exp \omega = 2.72$ . Thus, it is possible that with the given sample size, only the relatively large effect can be detected. This suggests that we must interpret the result of this test presented in Section S3.1 with caution.



## S4 Proofs of the Theorems

### S4.1 Lemmas

To prove the theorems, we need some lemmas.

LEMMA S1 *Consider two random variables  $X$  and  $Y$ . Suppose that they have finite moments and the support of  $Y$  contains that of  $X$ . Let  $f_1(x)$  and  $f_2(y)$  be their density functions. Then, any function  $g(\cdot)$ ,*

$$\mathbb{E}\{g(X)\} = \mathbb{E}\left\{\frac{f_1(Y)}{f_2(Y)}g(Y)\right\}.$$

Proof is straightforward and hence omitted.

LEMMA S2 *For a binary decision, Assumption 4 implies  $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp D_i \mid \mathbf{X}_i, Z_i = z$  under Assumption 3. For an ordinal decision, Assumption 4 implies  $R_i \perp\!\!\!\perp D_i \mid \mathbf{X}_i, Z_i = z$  under Assumption 6.*

*Proof of Lemma S2.* For a binary decision, we have

$$\begin{aligned} \Pr\{Y_i(1) = 1, Y_i(0) = 1 \mid D_i, \mathbf{X}_i, Z_i = z\} &= \Pr\{Y_i(1) = 1 \mid D_i, \mathbf{X}_i, Z_i = z\} \\ &= \Pr\{Y_i(1) = 1 \mid \mathbf{X}_i, Z_i = z\} \\ &= \Pr\{Y_i(1) = 1, Y_i(0) = 1 \mid \mathbf{X}_i, Z_i = z\}, \end{aligned}$$

where the first and third equality follow from Assumption 3 and the second equality follows from Assumption 4. Similarly, we have

$$\begin{aligned} \Pr\{Y_i(1) = 0, Y_i(0) = 0 \mid D_i, \mathbf{X}_i, Z_i = z\} &= \Pr\{Y_i(0) = 0 \mid D_i, \mathbf{X}_i, Z_i = z\} \\ &= \Pr\{Y_i(0) = 0 \mid \mathbf{X}_i, Z_i = z\} \\ &= \Pr\{Y_i(1) = 0, Y_i(0) = 0 \mid \mathbf{X}_i, Z_i = z\}, \end{aligned}$$

where the first and third equality follow from Assumption 3 and the second equality follows from Assumption 4. As a result,  $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp D_i \mid \mathbf{X}_i, Z_i = z$  because  $\{Y_i(1), Y_i(0)\}$  takes only three values.

For a discrete decision  $D_i$  taking values in  $\{0, \dots, k\}$ , we have

$$\begin{aligned} \Pr(R_i = r \mid D_i, \mathbf{X}_i, Z_i = z) &= \Pr(R_i \geq r \mid D_i, \mathbf{X}_i, Z_i = z) - \Pr(R_i \geq r + 1 \mid D_i, \mathbf{X}_i, Z_i = z) \\ &= \Pr(Y_i(r - 1) = 1 \mid D_i, \mathbf{X}_i, Z_i = z) - \Pr(Y_i(r) = 1 \mid D_i, \mathbf{X}_i, Z_i = z) \\ &= \Pr(Y_i(r - 1) = 1 \mid \mathbf{X}_i, Z_i = z) - \Pr(Y_i(r) = 1 \mid \mathbf{X}_i, Z_i = z) \\ &= \Pr(R_i \geq r \mid \mathbf{X}_i, Z_i = z) - \Pr(R_i \geq r + 1 \mid \mathbf{X}_i, Z_i = z) \\ &= \Pr(R_i = r \mid D_i, \mathbf{X}_i, Z_i = z) \end{aligned}$$

for  $r = 1, \dots, k$ , where the second and the fourth equality follow from the definition of  $R_i$  and the third equality follows from Assumption 4. Similarly, we have

$$\begin{aligned} \Pr(R_i = 0 \mid D_i, \mathbf{X}_i, Z_i = z) &= \Pr(Y_i(0) = 0 \mid D_i, \mathbf{X}_i, Z_i = z) \\ &= \Pr(Y_i(0) = 0 \mid D_i, Z_i = z) \\ &= \Pr(R_i = 0 \mid D_i, Z_i = z), \\ \Pr(R_i = k + 1 \mid D_i, \mathbf{X}_i, Z_i = z) &= \Pr(Y_i(k) = 1 \mid D_i, \mathbf{X}_i, Z_i = z) \\ &= \Pr(Y_i(k) = 1 \mid D_i, Z_i = z) \\ &= \Pr(R_i = k + 1 \mid D_i, Z_i = z). \end{aligned}$$

As a result,  $R_i \perp\!\!\!\perp D_i \mid \mathbf{X}_i, Z_i = z$ . □

## S4.2 Proof of Theorem 1

First, Assumption 3 implies,

$$\begin{aligned}\Pr\{Y_i(0) = 0, Y_i(1) = 0\} &= \Pr\{Y_i(0) = 0\}, \quad \Pr\{Y_i(0) = 1, Y_i(1) = 1\} = \Pr\{Y_i(1) = 1\}, \\ \Pr\{Y_i(0) = 1, Y_i(1) = 0\} &= 1 - \Pr\{Y_i(0) = 0\} - \Pr\{Y_i(1) = 1\}.\end{aligned}$$

Second, we have

$$\begin{aligned}&\Pr\{D_i(z) = 1, Y_i(0) = 0, Y_i(1) = 0\} \\ &= \Pr\{Y_i(0) = 0, Y_i(1) = 0\} - \Pr\{D_i(z) = 0, Y_i(0) = 0, Y_i(1) = 0\} \\ &= \Pr\{Y_i(0) = 0\} - \Pr\{D_i(z) = 0, Y_i(0) = 0\} \\ &= \Pr\{Y_i(0) = 0\} - \Pr\{D_i(z) = 0, Y_i(D_i(z)) = 0 \mid Z_i = z\} \\ &= \Pr\{Y_i(0) = 0\} - \Pr(D_i = 0, Y_i = 0 \mid Z_i = z),\end{aligned}$$

where the second equality follows from Assumption 3 and the third equality follows from Assumption 1. Similarly, we can obtain

$$\begin{aligned}\Pr\{D_i(z) = 1, Y_i(0) = 1, Y_i(1) = 1\} &= \Pr\{D_i(z) = 1, Y_i(1) = 1\} \\ &= \Pr\{D_i(z) = 1, Y_i(D_i(z)) = 1 \mid Z_i = z\} \\ &= \Pr(D_i = 1, Y_i = 1 \mid Z_i = z).\end{aligned}$$

Therefore,

$$\begin{aligned}&\Pr\{D_i(z) = 1, Y_i(0) = 1, Y_i(1) = 0\} \\ &= \Pr\{D_i(z) = 1\} - \Pr\{D_i(z) = 1, Y_i(0) = 0, Y_i(1) = 0\} - \Pr\{D_i(z) = 1, Y_i(0) = 1, Y_i(1) = 1\} \\ &= \Pr\{D_i = 1 \mid Z_i = z\} - \Pr\{Y_i(0) = 0\} + \Pr(D_i = 0, Y_i = 0 \mid Z_i = z) - \Pr(D_i = 1, Y_i = 1 \mid Z_i = z) \\ &= \Pr(Y_i = 0 \mid Z_i = z) - \Pr\{Y_i(0) = 0\}.\end{aligned}$$

Finally, we have,

$$\begin{aligned}\text{APCEp} &= \frac{\Pr\{D_i(1) = 1, Y_i(0) = 1, Y_i(1) = 0\} - \Pr\{D_i(0) = 1, Y_i(0) = 1, Y_i(1) = 0\}}{\Pr\{Y_i(0) = 1, Y_i(1) = 0\}} \\ &= \frac{\Pr(Y_i = 1 \mid Z_i = 0) - \Pr(Y_i = 1 \mid Z_i = 1)}{\Pr\{Y_i(0) = 1\} - \Pr\{Y_i(1) = 1\}},\end{aligned}$$

$$\begin{aligned}\text{APCEr} &= \frac{\Pr\{D_i(1) = 1, Y_i(0) = 1, Y_i(1) = 1\} - \Pr\{D_i(0) = 1, Y_i(0) = 1, Y_i(1) = 1\}}{\Pr\{Y_i(0) = 1, Y_i(1) = 1\}} \\ &= \frac{\Pr(D_i = 1, Y_i = 1 \mid Z_i = 1) - \Pr(D_i = 1, Y_i = 1 \mid Z_i = 0)}{\Pr\{Y_i(1) = 1\}},\end{aligned}$$

and

$$\begin{aligned}\text{APCEs} &= \frac{\Pr\{D_i(1) = 1, Y_i(0) = 0, Y_i(1) = 0\} - \Pr\{D_i(0) = 1, Y_i(0) = 0, Y_i(1) = 0\}}{\Pr\{Y_i(0) = 0\}} \\ &= \frac{\Pr(D_i = 0, Y_i = 0 \mid Z_i = 0) - \Pr(D_i = 0, Y_i = 0 \mid Z_i = 1)}{\Pr\{Y_i(0) = 0\}}.\end{aligned}$$

□

### S4.3 Proof of Theorem 2

Assumption 4 and Lemma S2 imply,

$$\begin{aligned}\mathbb{E}\{D_i(z) \mid Y_i(1) = y_1, Y_i(0) = y_0\} &= \mathbb{E}[\mathbb{E}\{D_i(z) \mid \mathbf{X}_i, Y_i(1) = y_1, Y_i(0) = y_0\} \mid Y_i(1) = y_1, Y_i(0) = y_0] \\ &= \mathbb{E}[\mathbb{E}\{D_i(z) \mid \mathbf{X}_i\} \mid Y_i(1) = y_1, Y_i(0) = y_0].\end{aligned}$$

Based on Lemma S1,

$$\begin{aligned}& \mathbb{E}[\mathbb{E}\{D_i(z) \mid \mathbf{X}_i\} \mid Y_i(1) = y_1, Y_i(0) = y_0] \\ &= \mathbb{E}\left[\frac{\Pr\{\mathbf{X}_i \mid Y_i(1) = y_1, Y_i(0) = y_0\}}{\Pr(\mathbf{X}_i)} \mathbb{E}\{D_i(z) \mid \mathbf{X}_i\}\right] \\ &= \mathbb{E}\left(\mathbb{E}\left[\frac{\Pr\{\mathbf{X}_i \mid Y_i(1) = y_1, Y_i(0) = y_0\}}{\Pr(\mathbf{X}_i)} D_i(z) \mid \mathbf{X}_i\right]\right) \\ &= \mathbb{E}\left(\mathbb{E}\left[\frac{\Pr\{Y_i(1) = y_1, Y_i(0) = y_0 \mid \mathbf{X}_i\}}{\Pr\{Y_i(1) = y_1, Y_i(0) = y_0\}} D_i(z) \mid \mathbf{X}_i\right]\right) \\ &= \mathbb{E}\left[\frac{\Pr\{Y_i(1) = y_1, Y_i(0) = y_0 \mid \mathbf{X}_i\}}{\Pr\{Y_i(1) = y_1, Y_i(0) = y_0\}} D_i(z)\right] \\ &= \mathbb{E}\left[\frac{\Pr\{Y_i(1) = y_1, Y_i(0) = y_0 \mid \mathbf{X}_i\}}{\Pr\{Y_i(1) = y_1, Y_i(0) = y_0\}} D_i \mid Z_i = z\right],\end{aligned}\tag{S1}$$

where the last equality follows from Assumption 1. We can then obtain the expressions for APCEp, APCEr, and APCEs by choosing different values of  $y_1$  and  $y_0$  in (S1).  $\square$

### S4.4 Proof of Theorem 3

Assumption 1 implies,

$$\Pr\{D_i(z) = d, Y_i(d) = y\} = \Pr\{D_i(z) = d, Y_i(D_i(z)) = y \mid Z_i = z\} = \Pr(D_i = d, Y_i = y \mid Z_i = z).$$

Therefore,

$$\begin{aligned}\Pr\{D_i(z) = 1 \mid Y_i(0) = y\} &= \frac{\Pr\{D_i(z) = 1, Y_i(0) = y\}}{\Pr\{Y_i(0) = y\}} \\ &= \frac{\Pr\{Y_i(0) = y\} - \Pr\{D_i(z) = 0, Y_i(0) = y\}}{\Pr\{Y_i(0) = y\}} \\ &= \frac{\Pr\{Y_i(0) = y\} - \Pr(D_i = 0, Y_i = y \mid Z_i = z)}{\Pr\{Y_i(0) = y\}}.\end{aligned}$$

As a result, we have

$$\begin{aligned}\text{APCEp} &= \frac{\Pr(D_i = 0, Y_i = 1 \mid Z_i = 0) - \Pr(D_i = 0, Y_i = 1 \mid Z_i = 1)}{\Pr\{Y_i(0) = 1\}}, \\ \text{APCEs} &= \frac{\Pr(D_i = 0, Y_i = 0 \mid Z_i = 0) - \Pr(D_i = 0, Y_i = 0 \mid Z_i = 1)}{\Pr\{Y_i(0) = 0\}}.\end{aligned}$$

$\square$

### S4.5 Proof of Theorem 5

Using the law of total expectation, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{D_i(z) \geq r\} \mid R_i = r] &= \mathbb{E}(\mathbb{E}[\mathbf{1}\{D_i(z) \geq r\} \mid \mathbf{X}_i, R_i = r] \mid R_i = r) \\
&= \mathbb{E}(\mathbb{E}[\mathbf{1}\{D_i(z) \geq r\} \mid \mathbf{X}_i] \mid R_i = r) \\
&= \mathbb{E}\left(\frac{\Pr(\mathbf{X}_i \mid R_i = r)}{\Pr(\mathbf{X}_i)} \mathbb{E}[\mathbf{1}\{D_i(z) \geq r\} \mid \mathbf{X}_i]\right) \\
&= \mathbb{E}\left(\frac{\Pr(R_i = r \mid \mathbf{X}_i)}{\Pr(R_i = r)} \mathbb{E}[\mathbf{1}\{D_i(z) \geq r\} \mid \mathbf{X}_i]\right) \\
&= \mathbb{E}\left[\frac{\Pr(R_i = r \mid \mathbf{X}_i)}{\Pr(R_i = r)} \mathbf{1}\{D_i(z) \geq r\}\right] \\
&= \mathbb{E}\left[\frac{\Pr(R_i = r \mid \mathbf{X}_i)}{\Pr(R_i = r)} \mathbf{1}\{D_i \geq r\} \mid Z_i = z\right],
\end{aligned}$$

where the second equality follows from Assumption 4 and Lemma S2, and the last equality follows from Assumption 1. Thus,

$$\text{APCEp}(r) = \mathbb{E}\{w_r(\mathbf{X}_i)\mathbf{1}(D_i \geq r) \mid Z_i = 1\} - \mathbb{E}\{w_r(\mathbf{X}_i)\mathbf{1}(D_i \geq r) \mid Z_i = 0\}.$$

We can prove the expression for APCEs similarly.  $\square$

## S5 Details of the Bayesian Estimation

We only consider the algorithm for sensitivity analysis with ordinal decision since the computation of the original analysis is straightforward by setting the sensitivity parameters to zero. Consider the model given in Equations (7) and (8). We can write Equation (7) in terms of the observed data as,

$$D_i^* = \beta_Z Z_i + \mathbf{X}_i^\top \beta_X + Z_i \mathbf{X}_i^\top \beta_{ZX} + \epsilon_{i1}, \quad (\text{S2})$$

where

$$D_i = \begin{cases} 0 & D_i^* \leq \theta_{Z_i,1} \\ 1 & \theta_{Z_i,1} < D_i^* \leq \theta_{Z_i,2} \\ \vdots & \vdots \\ k-1 & \theta_{Z_i,k-1} < D_i^* \leq \theta_{Z_i,k} \\ k & \theta_{Z_i,k} < D_i^* \end{cases}.$$

We then consider Equation (8). For  $r = 0, \dots, k$ , because  $R_i \geq r + 1$  is equivalent to  $Y_i(r) = 1$ , we have

$$\Pr\{Y(r) = 1\} = \Pr(R_i^* > \delta_r) = \Pr(\mathbf{X}_i^\top \alpha_X + \epsilon_{i2} > \delta_r) = \Pr(-\delta_r + \mathbf{X}_i^\top \alpha_X + \epsilon_{i2} > 0).$$

Therefore, we can introduce a latent variable  $Y^*(r)$ , and write

$$Y_i^*(r) = -\delta_r + \mathbf{X}_i^\top \alpha_X + \epsilon_{i2}, \quad (\text{S3})$$

where  $Y_i(r) = 1$  if  $Y_i^*(r) > 0$  and  $Y_i(r) = 0$  if  $Y_i^*(r) \leq 0$ . We can further write Equation (S3) in terms of the observed data as

$$Y_i^* = -\sum_{r=0}^k \delta_r \mathbf{1}(D_i = r) + \mathbf{X}_i^\top \alpha_X + \epsilon_{i2}, \quad (\text{S4})$$

where  $Y_i = 1$  if  $Y_i^* > 0$  and  $Y_i = 0$  if  $Y_i^* \leq 0$ .

Combining Equations (S2) and (S4), we have

$$D_i^* = \beta_Z Z_i + \mathbf{X}_i^\top \beta_X + Z_i \mathbf{X}_i^\top \beta_{ZX} + \epsilon_{i1}, \quad (\text{S5})$$

$$Y_i^* = - \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) + \mathbf{X}_i^\top \alpha_X + \epsilon_{i2}, \quad (\text{S6})$$

where

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

and

$$D_i = \begin{cases} 0 & D^* \leq \theta_{Z_i,1} \\ 1 & \theta_{Z_i,1} < D^* \leq \theta_{Z_i,2} \\ \vdots & \vdots \\ k-1 & \theta_{Z_i,k-1} < D^* \leq \theta_{Z_i,k} \\ k & \theta_{Z_i,k} < D^* \end{cases}, \quad Y_i = \begin{cases} 0 & Y_i^* \leq 0 \\ 1 & Y_i^* > 0 \end{cases}$$

with  $\delta_d \leq \delta_{d'}$  for  $d \leq d'$ .

We choose multivariate normal priors for the regression coefficients,  $(\beta_Z, \beta_X^\top, \beta_{ZX}^\top) \sim \mathbf{N}_{2p+1}(\mathbf{0}, \Sigma_D)$  and  $\alpha_X \sim \mathbf{N}_p(\mathbf{0}, \Sigma_R)$ . We choose the priors for  $\theta$  and  $\delta$  in the following manner. We first choose a normal prior for  $\theta_{z1}$  and  $\delta_0$ ,  $\theta_{z1} \sim N(0, \sigma_0^2)$  and  $\delta_0 \sim N(0, \sigma_0^2)$  for  $z = 0, 1$ . We then choose truncated normal priors for other parameters,  $\theta_{zj} \sim N(0, \sigma_0^2) \mathbf{1}(\theta_{zj} \geq \theta_{z,j-1})$  for  $j = 2, \dots, k$  and  $\delta_l \sim N(0, \sigma_0^2) \mathbf{1}(\delta_l \geq \delta_{l-1})$  for  $l = 1, \dots, k$ . In this way, we guarantee that  $\theta$ 's and  $\delta$ 's are increasing. In our empirical analysis, we choose  $\Sigma_D = 0.01 \cdot \mathbf{I}_{2p+1}$ ,  $\Sigma_R = 0.01 \cdot \mathbf{I}_p$ , and  $\sigma_0 = 10$ .

Treating  $Y_i^*$  and  $D_i^*$  as missing data, we can write the complete-data likelihood as

$$\begin{aligned} & L(\theta, \beta, \delta, \alpha) \\ &= \prod_{i=1}^n L_i(\theta, \beta, \delta, \alpha) \\ &\propto \prod_{i=1}^n \exp \left( -\frac{1}{2(1-\rho^2)} \left[ (D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX})^2 + \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\}^2 \right. \right. \\ &\quad \left. \left. - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i) \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\} \right] \right). \end{aligned}$$

**Imputation Step.** We first impute the missing data given the observed data and parameters. Using R package *tmvtnorm*, we can jointly sample  $Y_i^*$  and  $D_i^*$ . Given  $(D_i, Y_i, Z_i, \mathbf{X}_i^\top, \theta, \beta, \alpha, \delta)$ ,  $(D_i^*, Y_i^*)$  follows a truncated bivariate normal distribution whose means are given by  $\mathbf{X}_i^\top \beta_X + \beta_Z Z_i + Z_i \mathbf{X}_i^\top \beta_{ZX}$  and  $-\sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) + \mathbf{X}_i^\top \alpha_X$ , and whose covariance matrix has unit variances and correlation  $\rho$  where  $D^*$  is truncated within interval  $[\theta_{zd}, \theta_{z,d+1}]$  if  $Z_i = z$  and  $D_i = d$  (we define  $\theta_0 = -\infty$  and  $\theta_{k+1} = \infty$ ) and  $Y_i^*$  is truncated within  $(-\infty, 0)$  if  $Y_i = 0$  and  $[1, \infty)$  if  $Y_i = 1$ .

**Posterior Sampling Step.** The posterior distribution is proportional to

$$\begin{aligned}
& \prod_{i=1}^n \exp \left( -\frac{1}{2(1-\rho^2)} \left[ (D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX})^2 + \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\}^2 \right. \right. \\
& \quad \left. \left. - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\} \right] \right) \\
& \cdot \exp \left\{ -\frac{(\beta_Z, \beta_X^\top, \beta_{ZX}^\top) \Sigma_D^{-1} (\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top}{2} \right\} \cdot \exp \left( -\frac{\alpha_X^\top \Sigma_R^{-1} \alpha_X}{2} \right) \\
& \cdot \exp \left( -\frac{\theta_{11}^2}{2\sigma_0^2} \right) \exp \left( -\frac{\delta_0^2}{2\sigma_0^2} \right) \prod_{j=2}^k \left\{ \exp \left( -\frac{\theta_{1j}^2}{2\sigma_0^2} \right) \mathbf{1}(\theta_{1j} \geq \theta_{1,j-1}) \right\} \prod_{l=1}^k \left\{ \exp \left( -\frac{\delta_l^2}{2\sigma_0^2} \right) \mathbf{1}(\delta_l \geq \delta_{l-1}) \right\} \\
& \cdot \exp \left( -\frac{\theta_{01}^2}{2\sigma_0^2} \right) \prod_{j=2}^k \left\{ \exp \left( -\frac{\theta_{0j}^2}{2\sigma_0^2} \right) \mathbf{1}(\theta_{0j} \geq \theta_{0,j-1}) \right\}.
\end{aligned}$$

We first sample  $(\beta_Z, \beta_X^\top, \beta_{ZX}^\top)$ . From the posterior distribution, we have

$$\begin{aligned}
& f(\beta_Z, \beta_X^\top, \beta_{ZX}^\top | \cdot) \\
& \propto \prod_{i=1}^n \exp \left( -\frac{1}{2(1-\rho^2)} \left[ (D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX})^2 \right. \right. \\
& \quad \left. \left. - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\} \right] \right) \cdot \exp \left\{ -\frac{(\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top \Sigma_D^{-1} (\beta_Z, \beta_X^\top, \beta_{ZX}^\top)}{2} \right\} \\
& \propto \prod_{i=1}^n \exp \left( -\frac{1}{2(1-\rho^2)} \left[ (\beta_Z, \beta_X^\top, \beta_{ZX}^\top) (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top)^\top (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top) (\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top - 2D_i^* (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top) (\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top \right. \right. \\
& \quad \left. \left. + 2\rho \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\} (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top) (\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top \right] \right) \cdot \exp \left\{ -\frac{(\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top \Sigma_D^{-1} (\beta_Z, \beta_X^\top, \beta_{ZX}^\top)}{2} \right\}.
\end{aligned}$$

Therefore, we can sample

$$(\beta_Z, \beta_X^\top, \beta_{ZX}^\top)^\top | \cdot \sim \mathbf{N}_{p+1}(\hat{\mu}_D, \hat{\Sigma}_D),$$

where

$$\begin{aligned}
\hat{\Sigma}_D &= \left\{ \frac{1}{1-\rho^2} \sum_{i=1}^n (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top)^\top (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top) + \Sigma_D^{-1} \right\}^{-1}, \\
\hat{\mu}_D &= \hat{\Sigma}_D \left( \frac{1}{1-\rho^2} \sum_{i=1}^n (Z_i, \mathbf{X}_i^\top, Z_i \mathbf{X}_i^\top)^\top \left[ D_i^* - \rho \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\} \right] \right).
\end{aligned}$$

We then consider sampling  $\alpha_X$ . We have

$$\begin{aligned}
& f(\alpha_X | \cdot) \\
& \propto \prod_{i=1}^n \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\}^2 \right. \right. \\
& \quad \left. \left. - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) - \mathbf{X}_i^\top \alpha_X \right\} \right] \right) \cdot \exp \left( -\frac{\alpha_X^\top \Sigma_R^{-1} \alpha_X}{2} \right) \\
& \propto \prod_{i=1}^n \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \alpha_X^\top \mathbf{X}_i^\top \mathbf{X}_i \alpha_X - 2 \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) \right\} \mathbf{X}_i \alpha_X + 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \mathbf{X}_i \alpha_X \right] \right)
\end{aligned}$$

$$\cdot \exp\left(-\frac{\alpha_X^\top \Sigma_R^{-1} \alpha_X}{2}\right).$$

Therefore, we can sample

$$\alpha_X | \cdot \sim \mathbf{N}_p(\widehat{\mu}_R, \widehat{\Sigma}_R),$$

where

$$\begin{aligned} \widehat{\Sigma}_R &= \left\{ \frac{1}{1-\rho^2} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i + \Sigma_R^{-1} \right\}^{-1}, \\ \widehat{\mu}_R &= \widehat{\Sigma}_R \left( \frac{1}{1-\rho^2} \sum_{i=1}^n \mathbf{X}_i \left[ \left\{ Y_i^* + \sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) \right\} - \rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \right] \right). \end{aligned}$$

To sample  $\delta$ 's, we write  $\sum_{d=0}^k \delta_d \mathbf{1}(D_i = d) = \delta_0 + \sum_{d=1}^k (\delta_d - \delta_{d-1}) \mathbf{1}(D_i \geq d)$  and denote  $\mathbf{W}_i = (1, \mathbf{1}(D_i \geq 1), \dots, \mathbf{1}(D_i \geq k))$  and  $\delta = (\delta_0, \delta_1 - \delta_0, \dots, \delta_k - \delta_{k-1})$ . Thus, we have

$$\begin{aligned} & f(\delta | \cdot) \\ & \propto \prod_{i=1}^n \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \left\{ Y_i^* + \mathbf{W}_i \delta - \mathbf{X}_i^\top \alpha_X \right\}^2 - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \left\{ Y_i^* + \mathbf{W}_i \delta - \mathbf{X}_i^\top \alpha_X \right\} \right]\right) \\ & \quad \cdot \exp\left(-\frac{\delta_0^2}{2\sigma_0^2}\right) \prod_{l=1}^k \left\{ \exp\left(-\frac{\delta_l^2}{2\sigma_0^2}\right) \mathbf{1}(\delta_l - \delta_{l-1} \geq 0) \right\} \\ & \propto \prod_{i=1}^n \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \delta^\top \mathbf{W}_i^\top \mathbf{W}_i \delta + 2(Y_i^* - \mathbf{X}_i^\top \alpha_X) \mathbf{W}_i \delta - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \mathbf{W}_i \delta \right]\right) \\ & \quad \cdot \exp\left(-\frac{\delta_0^2}{2\sigma_0^2}\right) \prod_{l=1}^k \left\{ \exp\left(-\frac{\delta_l^2}{2\sigma_0^2}\right) \mathbf{1}(\delta_l - \delta_{l-1} \geq 0) \right\} \\ & \propto \prod_{i=1}^n \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \delta^\top \mathbf{W}_i^\top \mathbf{W}_i \delta + 2(Y_i^* - \mathbf{X}_i^\top \alpha_X) \mathbf{W}_i \delta - 2\rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) \mathbf{W}_i \delta \right]\right) \\ & \quad \cdot \exp\left(-\frac{\delta^\top C^\top C \delta}{2\sigma_0^2}\right) \prod_{l=1}^k \mathbf{1}(\delta_l - \delta_{l-1} \geq 0), \end{aligned}$$

where  $C$  is a  $(k+1) \times (k+1)$  lower triangular matrix with all non-zero entries equal to 1. Therefore, we can draw from a truncated normal distribution with mean and covariance matrix

$$\begin{aligned} \widehat{\Sigma}_\delta &= \left\{ \frac{1}{1-\rho^2} \sum_{i=1}^n \mathbf{W}_i^\top \mathbf{W}_i + \frac{C^\top C}{\sigma_0^2} \right\}^{-1}, \\ \widehat{\mu}_\delta &= \widehat{\Sigma}_\delta \left[ \frac{1}{1-\rho^2} \sum_{i=1}^n \mathbf{W}_i^\top \left\{ \rho(D_i^* - \mathbf{X}_i^\top \beta_X - \beta_Z Z_i - Z_i \mathbf{X}_i^\top \beta_{ZX}) - (Y_i^* - \mathbf{X}_i^\top \alpha_X) \right\} \right], \end{aligned}$$

where the 2-th to  $(k+1)$ -th element is truncated within interval  $[0, \infty)$ . We can then transform  $\delta$  to obtain  $(\delta_0, \delta_1, \dots, \delta_k)$ .

Finally, we sample

$$\theta_{z1} | \cdot \sim TN(0, \sigma_0^2; \max_{i: Z_i=z, D_i=0} D_i^*, \min_{i: Z_i=z, D_i=1} (D_i^*, \theta_2)).$$

We then sample

$$\theta_{zj} | \cdot \sim TN(0, \sigma_0^2; \max_{i: Z_i=z, D_i=j-1} (D_i^*, \theta_{j-1}), \min_{i: Z_i=z, D_i=j} (D_i^*, \theta_{j+1}))$$

for  $j = 2, \dots, k-1$ , and

$$\theta_{zk} \mid \cdot \sim TN(0, \sigma_0^2; \max_{i: Z_i=z, D_i=k-1} (D_i^*, \theta_{k-1}), \min_{i: Z_i=z, D_i=k} D_i^*).$$

The MCMC gives the posterior distributions of the parameters and therefore we can obtain the posterior distributions of  $\Pr(D_i \mid R_i, \mathbf{X}_i = \mathbf{x}, Z_i = z)$  and  $\Pr(R_i \mid \mathbf{X}_i = \mathbf{x})$ . As a result, for  $r = 1, \dots, k$ , we have

$$\begin{aligned} \text{APCEp}(r) &= \Pr\{D_i(1) \geq r \mid R_i = r\} - \Pr\{D_i(0) \geq r \mid R_i = r\} \\ &= \frac{\mathbb{E}\{\Pr(D_i(1) \geq r, R_i = r \mid \mathbf{X}_i)\}}{\mathbb{E}\{\Pr(R_i = r \mid \mathbf{X}_i)\}} - \frac{\mathbb{E}\{\Pr(D_i(0) \geq r, R_i = r \mid \mathbf{X}_i)\}}{\mathbb{E}\{\Pr(R_i = r \mid \mathbf{X}_i)\}}, \\ \text{APCEs} &= \Pr\{D_i(1) = 0 \mid R_i = 0\} - \Pr\{D_i(0) = 0 \mid R_i = 0\} \\ &= \frac{\mathbb{E}\{\Pr(D_i(1) = 0, R_i = 0 \mid \mathbf{X}_i)\}}{\mathbb{E}\{\Pr(R_i = 0 \mid \mathbf{X}_i)\}} - \frac{\mathbb{E}\{\Pr(D_i(0) = 0, R_i = 0 \mid \mathbf{X}_i)\}}{\mathbb{E}\{\Pr(R_i = 0 \mid \mathbf{X}_i)\}}. \end{aligned}$$

We can calculate the conditional probabilities  $\Pr\{D_i(z), R_i \mid \mathbf{X}_i\}$  and  $\Pr(R_i \mid \mathbf{X}_i)$  based on the posterior sample of the coefficients, and then replace the expectation with the empirical average to obtain the estimates.

## S6 Optimal PSA Provision

In this appendix, we consider the optimal PSA provision rule and conduct an empirical analysis. Let  $\xi$  be a PSA provision rule, i.e.,  $\xi(\mathbf{x}) = 1$  (the PSA is provided) if  $\mathbf{x} \in \mathcal{B}_1$  and  $\xi(\mathbf{x}) = 0$  (the PSA is not provided) if  $\mathbf{x} \in \mathcal{B}_0$ , where  $\mathcal{X} = \mathcal{B}_0 \cup \mathcal{B}_1$  and  $\mathcal{B}_0 \cap \mathcal{B}_1 = \emptyset$ . The judges will make their decisions based on the PSA and other available information included in  $\mathbf{X}_i = \mathbf{x}$ . To consider the influence of the PSA on judges' decision, we define  $\delta_{i1}$  the potential decision rule of case  $i$  if the judge received the PSA and  $\delta_{i0}$  if not. Thus,  $\delta_{iz}(\mathbf{x}) = d$  if  $\mathbf{x} \in \mathcal{X}_{i,zd}$  where  $\mathcal{X}_{i,zd}$  is a partition of the covariate space with  $\mathcal{X} = \bigcup_{d=0}^k \mathcal{X}_{i,zd}$  and  $\mathcal{X}_{i,zd} \cap \mathcal{X}_{i,zd'} = \emptyset$  for  $z = 0, 1$ . Although we allow the judge to make a different decision even if the observed case characteristics  $\mathbf{X}_i$  are identical, we assume that the judges' decisions are identically distributed given the observed case characteristics and PSA provision. That is, we assume  $\Pr\{\delta_{iz}(\mathbf{x}) = d\} = \Pr\{\delta_{i'z}(\mathbf{x}) = d\}$  for fixed  $\mathbf{x}$ ,  $z$  and  $i \neq i'$ , where the probability is taken with respect to the super population of all cases.

Given this setup, we derive the optimal PSA provision rule. We consider the 0–1 utility  $U_i(\xi) = \mathbf{1}\{\delta_{i,\xi(\mathbf{x}_i)}(\mathbf{X}_i) = R_i\}$ . This utility equals one, if the judge makes the most lenient decision to prevent an arrestee from engaging in NCA (NVCA or FTA), and equals zero otherwise. As before, we begin by rewriting the expected utility in the following manner,

$$\begin{aligned} \mathbb{E}\{U_i(\xi)\} &= \mathbb{E}[\mathbf{1}\{R_i = \delta_{i,\xi(\mathbf{x}_i)}(\mathbf{X}_i)\}] \\ &= \sum_{r=0}^k \mathbb{E}[\mathbf{1}\{R_i = r, \delta_{i,\xi(\mathbf{x}_i)}(\mathbf{X}_i) = r\}] \\ &= \sum_{r=0}^k \sum_{z=0}^1 \mathbb{E}[\mathbf{1}\{R_i = r, \delta_{iz}(\mathbf{X}_i) = r, \mathbf{X}_i \in \mathcal{B}_z\}]. \end{aligned}$$

Under the unconfoundedness assumption, we can write,

$$\begin{aligned} \mathbb{E}[\mathbf{1}\{R_i = r, \delta_{iz}(\mathbf{X}_i) = r, \mathbf{X}_i \in \mathcal{B}_z\}] &= \mathbb{E}[\Pr(R_i = r \mid \mathbf{X}_i) \cdot \Pr\{\delta_{iz}(\mathbf{X}_i) = r \mid \mathbf{X}_i\} \cdot \mathbf{1}\{\mathbf{X}_i \in \mathcal{B}_z\}] \\ &= \mathbb{E}[e_r(\mathbf{X}_i) \cdot \Pr\{\delta_{iz}(\mathbf{X}_i) = r\} \cdot \mathbf{1}\{\mathbf{X}_i \in \mathcal{B}_z\}]. \end{aligned}$$



Because in the experiment, the provision of the PSA is randomized, we can estimate  $\Pr\{\delta_{iz}(\mathbf{X}_i) = r\} = \Pr(D_i = r \mid Z_i = z, \mathbf{X}_i)$  from the data. Therefore, we obtain

$$\mathbb{E}\{U_i(\xi)\} = \sum_{z=0,1} \mathbb{E} \left( \left[ \sum_{r=0}^k e_r(\mathbf{X}_i) \cdot \Pr(D_i = r \mid Z_i = z, \mathbf{X}_i) \right] \cdot \mathbf{1}\{\mathbf{X}_i \in \mathcal{B}_z\} \right).$$

Then, the optimal PSA provision rule is,

$$\xi(\mathbf{x}) = \underset{z=0,1}{\operatorname{argmax}} h_z(\mathbf{x}) \quad \text{where} \quad h_z(\mathbf{x}) = \sum_{r=0}^k e_r(\mathbf{x}) \cdot \Pr(D_i = r \mid Z_i = z, \mathbf{X}_i). \quad (\text{S7})$$

Thus, we can use the experimental data to derive the optimal PSA provision rule.

## S7 Frequentist Analysis

In this appendix, we implement frequentist analysis and present the results. We fit the model defined in Equation (S4) with probit regression. Recall that for  $r = 0, \dots, k$ ,  $R_i \geq r + 1$  is equivalent to  $Y_i(r) = 1$ . Hence, we can estimate the conditional probabilities  $e_r(\mathbf{X}_i)$  for each  $r = 0, \dots, k + 1$  based on the estimates of the regression coefficients,

$$\begin{aligned} \hat{e}_r(\mathbf{x}) &= \Phi(-\hat{\delta}_{r-1} + \mathbf{x}^\top \hat{\alpha}_X) - \Phi(-\hat{\delta}_r + \mathbf{x}^\top \hat{\alpha}_X), \quad \text{for } r = 1, \dots, k, \\ \hat{e}_{k+1}(\mathbf{x}) &= \Phi(-\hat{\delta}_k + \mathbf{x}^\top \hat{\alpha}_X), \\ \hat{e}_0(\mathbf{x}) &= 1 - \Phi(-\hat{\delta}_0 + \mathbf{x}^\top \hat{\alpha}_X), \end{aligned}$$

where  $\Phi(\cdot)$  denotes the cumulative distribution function of the standard normal distribution. We estimate  $\widehat{\text{APCE}}_p(r)$  and  $\widehat{\text{APCE}}_s$  using Hajek estimator as follows,

$$\begin{aligned} \widehat{\text{APCE}}_p(r) &= \frac{\sum_i \hat{w}_r(\mathbf{X}_i) \mathbf{1}(D_i \geq r) \mathbf{1}(Z_i = 1)}{\sum_i \hat{w}_r(\mathbf{X}_i) \mathbf{1}(Z_i = 1)} - \frac{\sum_i \hat{w}_r(\mathbf{X}_i) \mathbf{1}(D_i \geq r) \mathbf{1}(Z_i = 0)}{\sum_i \hat{w}_r(\mathbf{X}_i) \mathbf{1}(Z_i = 0)}, \\ \widehat{\text{APCE}}_s &= \frac{\sum_i \hat{w}_0(\mathbf{X}_i) \mathbf{1}(D_i = 0) \mathbf{1}(Z_i = 1)}{\sum_i \hat{w}_0(\mathbf{X}_i) \mathbf{1}(Z_i = 1)} - \frac{\sum_i \hat{w}_0(\mathbf{X}_i) \mathbf{1}(D_i = 0) \mathbf{1}(Z_i = 0)}{\sum_i \hat{w}_0(\mathbf{X}_i) \mathbf{1}(Z_i = 0)}, \end{aligned}$$

where  $\hat{w}_r(\mathbf{x}) = \hat{e}_r(\mathbf{x}) / \{\frac{1}{n} \sum_i \hat{e}_r(\mathbf{X}_i)\}$ . We use bootstrap to compute the 95% confidence interval.

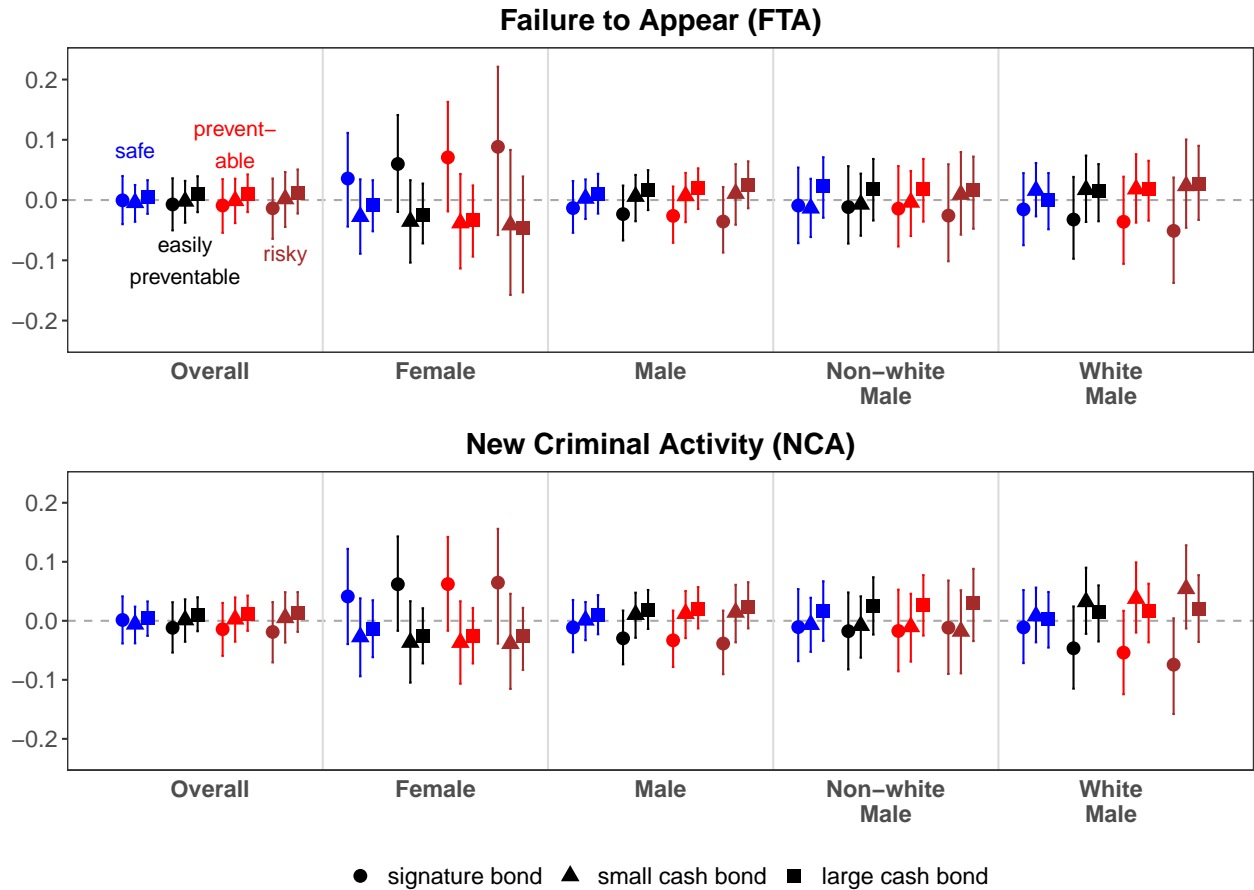


Figure S10: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision based on Frequentist Analysis. Each panel presents the overall and subgroup-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the 95% credible interval.

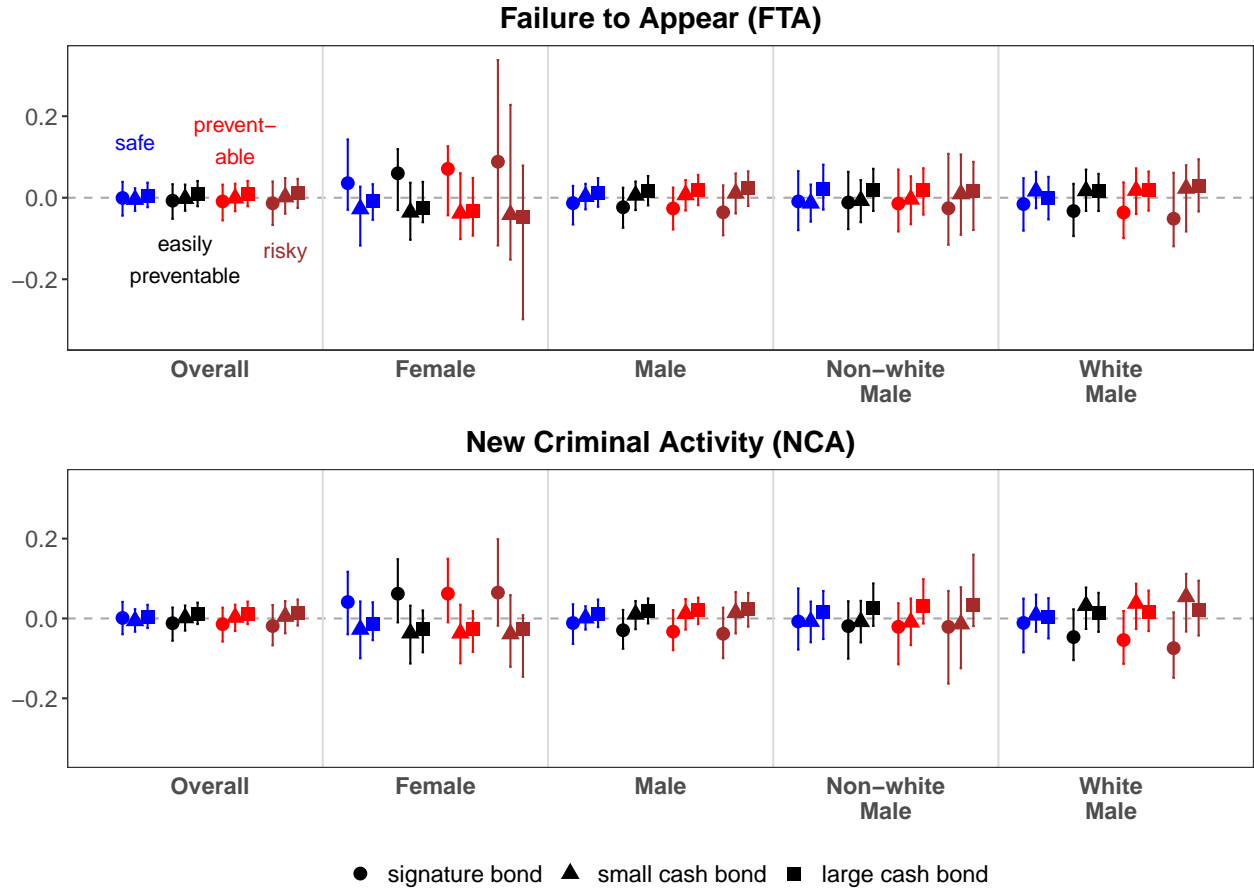


Figure S11: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision based on Frequentist Analysis with Random Effects. Each panel presents the overall and subgroup-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the 95% credible interval.

Figures S10 presents the estimated APCE of PSA provision on the three ordinal decision categories, separately for FTA and NCA within each principal stratum. The results for NVCA are not presented due to the fact that the number of events is too small for an informative subgroup analysis. The results are largely consistent with those of the Bayesian analysis presented in the main text. As a robustness check for the assumption of no interference among the cases, Figure S11 presents the estimated APCE of PSA provision with the model including random effects for the hearing date of the case, and the results are the same. Figure S12 presents the results for each age group similar to the one in Appendix S2.

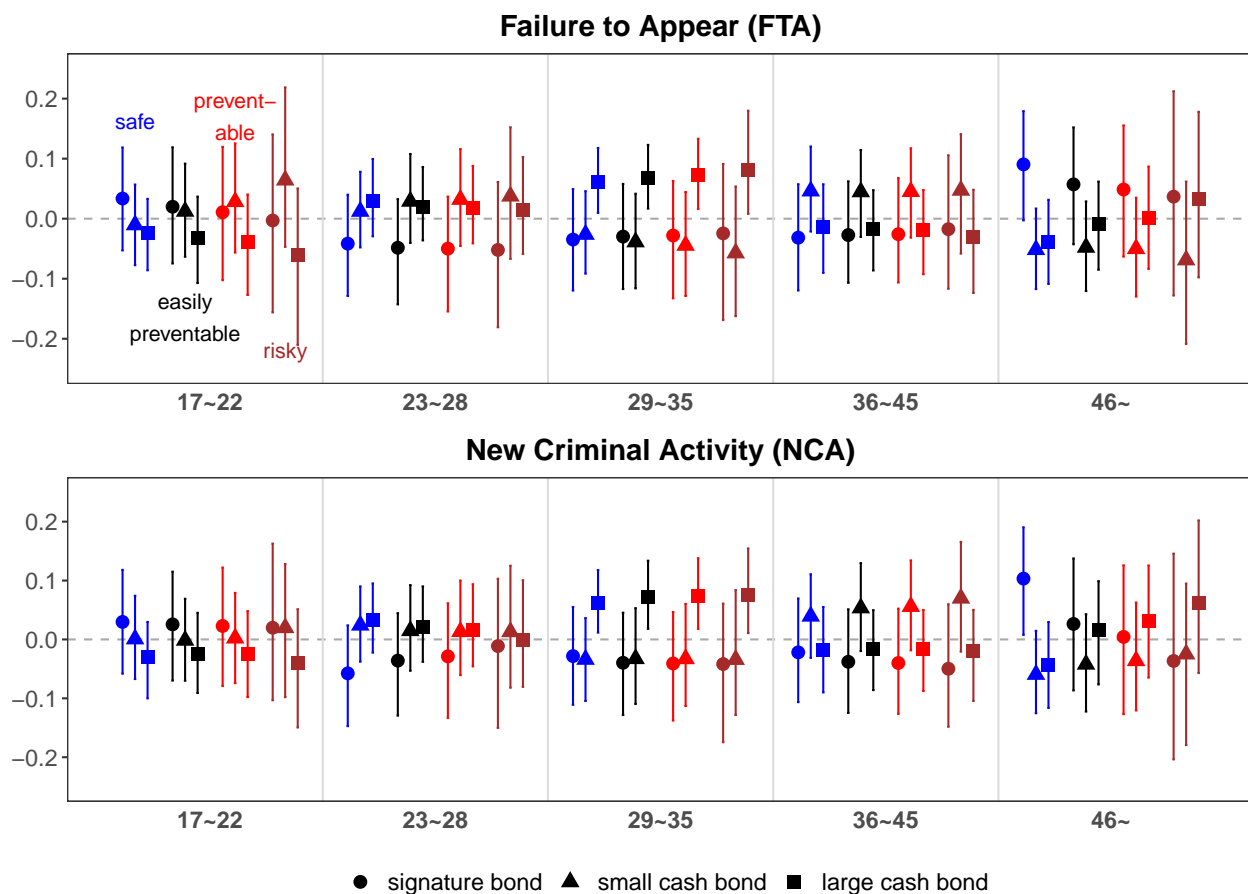


Figure S12: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision based on Frequentist Analysis. Each panel presents the age group-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the 95% credible interval.

## S8 Nonparametric Sensitivity Analysis

We consider a nonparametric sensitivity analysis for the ordinal decision under the monotonicity assumption (Assumption 6). We introduce the following sensitivity parameters,  $\xi_{rdz}(\mathbf{x})$  for  $r, d = 0, \dots, k$  and  $z = 0, 1$ , to characterize the deviation from the unconfoundedness assumption,

$$\xi_{rdz}(\mathbf{x}) = \frac{\Pr\{Y_i(r) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\}}{\Pr\{Y_i(r) = 1 \mid D_i(z) = 0, \mathbf{X}_i = \mathbf{x}\}},$$

which is equal to 1 for all  $(r, d, z)$  and  $\mathbf{x}$  when the unconfoundedness assumption holds.

We can directly relate the parametric sensitivity parameter  $\rho$  to the parameters of the nonparametric sensitivity analysis. Because  $R_i \geq r + 1$  is equivalent to  $Y_i(r) = 1$ , we can obtain the following formula from Equations (7) and (8),

$$\Pr\{Y_i(r) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} = \frac{\Pr(\theta_{zd} < \beta_Z z + \mathbf{x}^\top \beta_X + z \mathbf{x}^\top \beta_{ZX} + \epsilon_{i1} \leq \theta_{z,d+1}, \delta_r < \mathbf{x}^\top \alpha_X + \epsilon_{i2})}{\Pr(\theta_{zd} < \beta_Z z + \mathbf{x}^\top \beta_X + z \mathbf{x}^\top \beta_{ZX} + \epsilon_{i1} \leq \theta_{z,d+1})},$$

where  $\theta_{z0} = -\infty$  and  $\delta_{k+1} = \infty$ . Together with Proposition S1, we can express the sensitivity parameters in the nonparametric sensitivity analysis  $\xi_{rdz}(\mathbf{x})$  in terms of the model parameters given in Equations (7) and (8). Thus, the parametric sensitivity analysis, while much simpler, imposes restrictions on the nonparametric counterpart.

The following proposition gives the identification formulas for  $\Pr\{D_i(z) = d \mid R_i = r\}$  for all  $(r, d, z)$  with any given value of  $\xi_{rdz}(\mathbf{x})$ .

**PROPOSITION S1** *Under Assumptions 1, 2, and 6, if  $\xi_{rdz}(\mathbf{x})$  is known for all  $(r, d, z)$  and  $\mathbf{x}$ , then we have*

$$\Pr\{D_i(z) = d \mid R_i = r\} = \frac{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid \mathbf{X}_i = \mathbf{x}\} - \Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]} - \frac{\mathbb{E}[\Pr\{Y_i(r) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid \mathbf{X}_i = \mathbf{x}\} - \Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]}$$

for  $r = 1, \dots, k$  and all  $(d, z)$ , and

$$\Pr\{D_i(z) = d \mid R_i = k + 1\} = \frac{\mathbb{E}[\Pr\{Y_i(k) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(k) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]},$$

$$\Pr\{D_i(z) = d \mid R_i = 0\} = \frac{\mathbb{E}[\Pr\{Y_i(0) = 0 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(0) = 0 \mid \mathbf{X}_i = \mathbf{x}\}]}$$

for all  $(d, z)$ , where

$$\Pr\{Y_i(r) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} = \frac{\xi_{rdz}(\mathbf{x})}{\xi_{rrz}(\mathbf{x})} \cdot \Pr(Y_i = 1 \mid Z_i = z, D_i = r, \mathbf{X}_i = \mathbf{x}),$$

$$\Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\} = \frac{\sum_{d=0}^k \xi_{rdz}(\mathbf{x}) \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})}{\xi_{rrz}(\mathbf{x})} \cdot \Pr(Y_i = 1 \mid Z_i = z, D_i = r, \mathbf{X}_i = \mathbf{x}).$$

**PROOF:** The randomization of treatment assignment (Assumption 1) implies,

$$\Pr\{Y_i(r) = 1 \mid D_i(z) = r, \mathbf{X}_i = \mathbf{x}\} = \Pr(Y_i = 1 \mid Z_i = z, D_i = r, \mathbf{X}_i = \mathbf{x}).$$

Therefore, with given values of  $\xi_{rdz}(\mathbf{x})$ , we have,

$$\begin{aligned}
\Pr\{Y_i(r) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} &= \frac{\xi_{rdz}(\mathbf{x})}{\xi_{rrz}(\mathbf{x})} \cdot \Pr(Y_i = 1 \mid Z_i = z, D_i = r, \mathbf{X}_i = \mathbf{x}), \\
\Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\} &= \sum_{d=0}^k \Pr\{Y_i(r) = 1 \mid D(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr\{D(z) = d \mid \mathbf{X}_i = \mathbf{x}\} \\
&= \frac{\sum_{d=0}^k \xi_{rdz}(\mathbf{x}) \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})}{\xi_{rrz}(\mathbf{x})} \\
&\quad \cdot \Pr(Y_i = 1 \mid Z_i = z, D_i = r, \mathbf{X}_i = \mathbf{x}).
\end{aligned}$$

From the above two terms, we have

$$\begin{aligned}
&\Pr\{D_i(z) = d \mid R_i = r\} \\
&= \frac{\mathbb{E}[\Pr\{D_i(z) = d, R_i = r \mid \mathbf{X}_i = \mathbf{x}\}]}{\mathbb{E}\{\Pr(R_i = r \mid \mathbf{X}_i = \mathbf{x})\}} \\
&= \frac{\mathbb{E}[\Pr\{D_i(z) = d, Y_i(r-1) = 1 \mid \mathbf{X}_i = \mathbf{x}\} - \Pr\{D_i(z) = d, Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]}{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid \mathbf{X}_i = \mathbf{x}\} - \Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]} \\
&= \frac{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid \mathbf{X}_i = \mathbf{x}\} - \Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]} \\
&= \frac{\mathbb{E}[\Pr\{Y_i(r) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(r-1) = 1 \mid \mathbf{X}_i = \mathbf{x}\} - \Pr\{Y_i(r) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]}
\end{aligned}$$

for  $r = 1, \dots, k$ , where the first equality follows from the law of total expectation, and the second equality follows from Assumption 6.

Similarly, we can obtain

$$\begin{aligned}
&\Pr\{D_i(z) = d \mid R_i = k+1\} \\
&= \frac{\mathbb{E}[\Pr\{D_i(z) = d, R_i = k+1 \mid \mathbf{X}_i = \mathbf{x}\}]}{\mathbb{E}\{\Pr(R_i = k+1 \mid \mathbf{X}_i = \mathbf{x})\}} \\
&= \frac{\mathbb{E}[\Pr\{D_i(z) = d, Y_i(k) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]}{\mathbb{E}[\Pr\{Y_i(k) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]} \\
&= \frac{\mathbb{E}[\Pr\{Y_i(k) = 1 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(k) = 1 \mid \mathbf{X}_i = \mathbf{x}\}]}, \\
&\Pr\{D_i(z) = d \mid R_i = 0\} \\
&= \frac{\mathbb{E}[\Pr\{D_i(z) = d, R_i = 0 \mid \mathbf{X}_i = \mathbf{x}\}]}{\mathbb{E}\{\Pr(R_i = 0 \mid \mathbf{X}_i = \mathbf{x})\}} \\
&= \frac{\mathbb{E}[\Pr\{D_i(z) = d, Y_i(0) = 0 \mid \mathbf{X}_i = \mathbf{x}\}]}{\mathbb{E}[\Pr\{Y_i(0) = 0 \mid \mathbf{X}_i = \mathbf{x}\}]} \\
&= \frac{\mathbb{E}[\Pr\{Y_i(0) = 0 \mid D_i(z) = d, \mathbf{X}_i = \mathbf{x}\} \Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})]}{\mathbb{E}[\Pr\{Y_i(0) = 0 \mid \mathbf{X}_i = \mathbf{x}\}]}
\end{aligned}$$

□

Using this result, we can compute the APCE with any given value of  $\xi_{rdz}(\mathbf{x})$ . Unfortunately, this nonparametric sensitivity analysis requires the specification of too many sensitivity parameters, making it unsuitable for practical use.

## S9 Parametric Sensitivity Analysis Results

In this appendix, we implement sensitivity analysis for unconfoundedness assumption (Assumption 4) and present the results. For nonparametric sensitivity analysis, we estimate  $\Pr(Y_i = 1 \mid Z_i = z, D_i = r, \mathbf{X}_i = \mathbf{x})$  and  $\Pr(D_i = d \mid Z_i = z, \mathbf{X}_i = \mathbf{x})$  using the model defined in Equations (S5) and (S6). Figures S13, S14, and S15 show the results for the parametric sensitivity analysis. The patterns of the estimated APCEs of PSA provision with different sets of sensitivity parameters are generally consistent with those in the case where the unconfoundedness assumption holds.

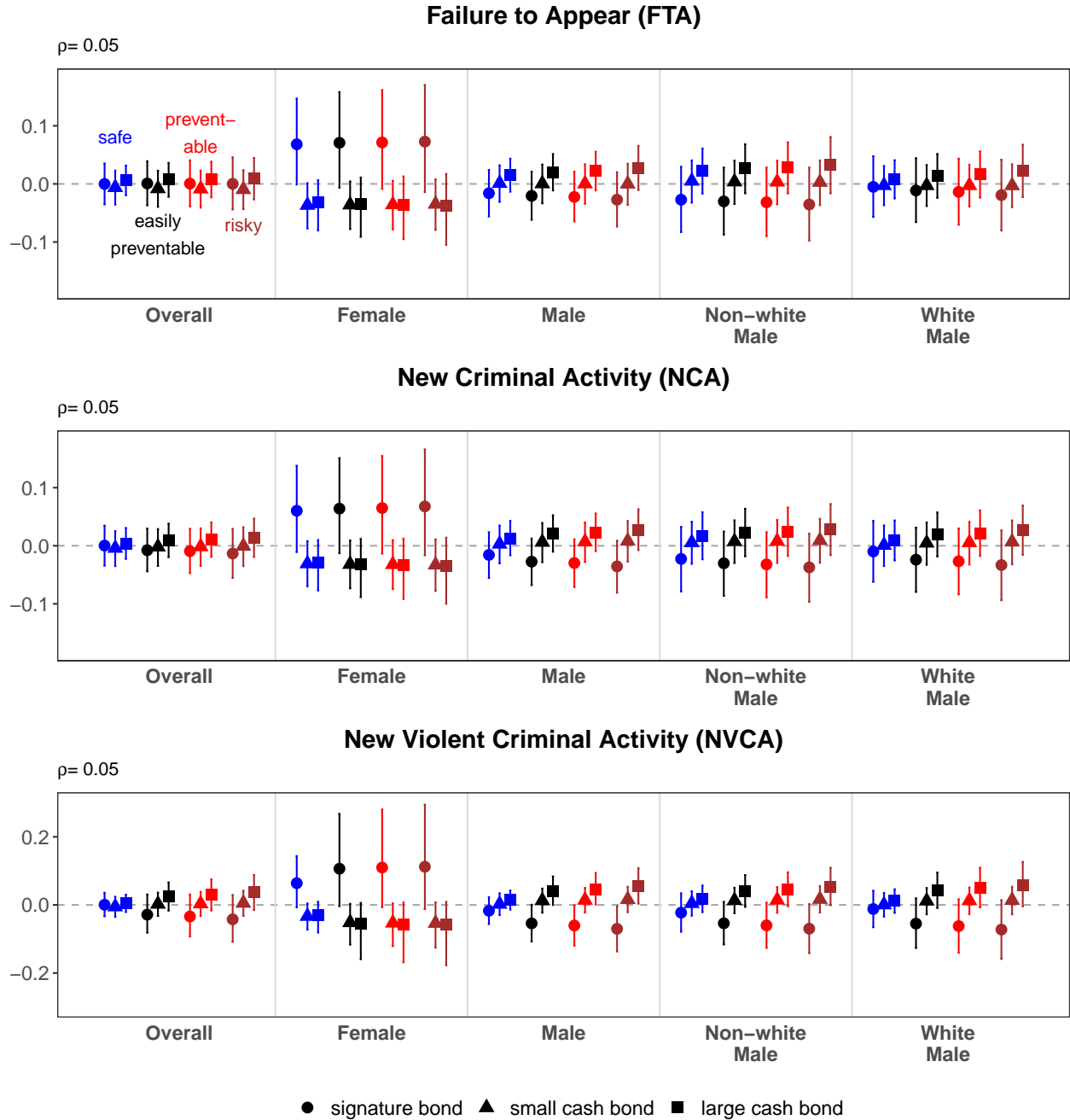


Figure S13: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision with  $\rho = 0.05$ . Each panel presents the overall and subgroup-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the Bayesian 95% credible interval.



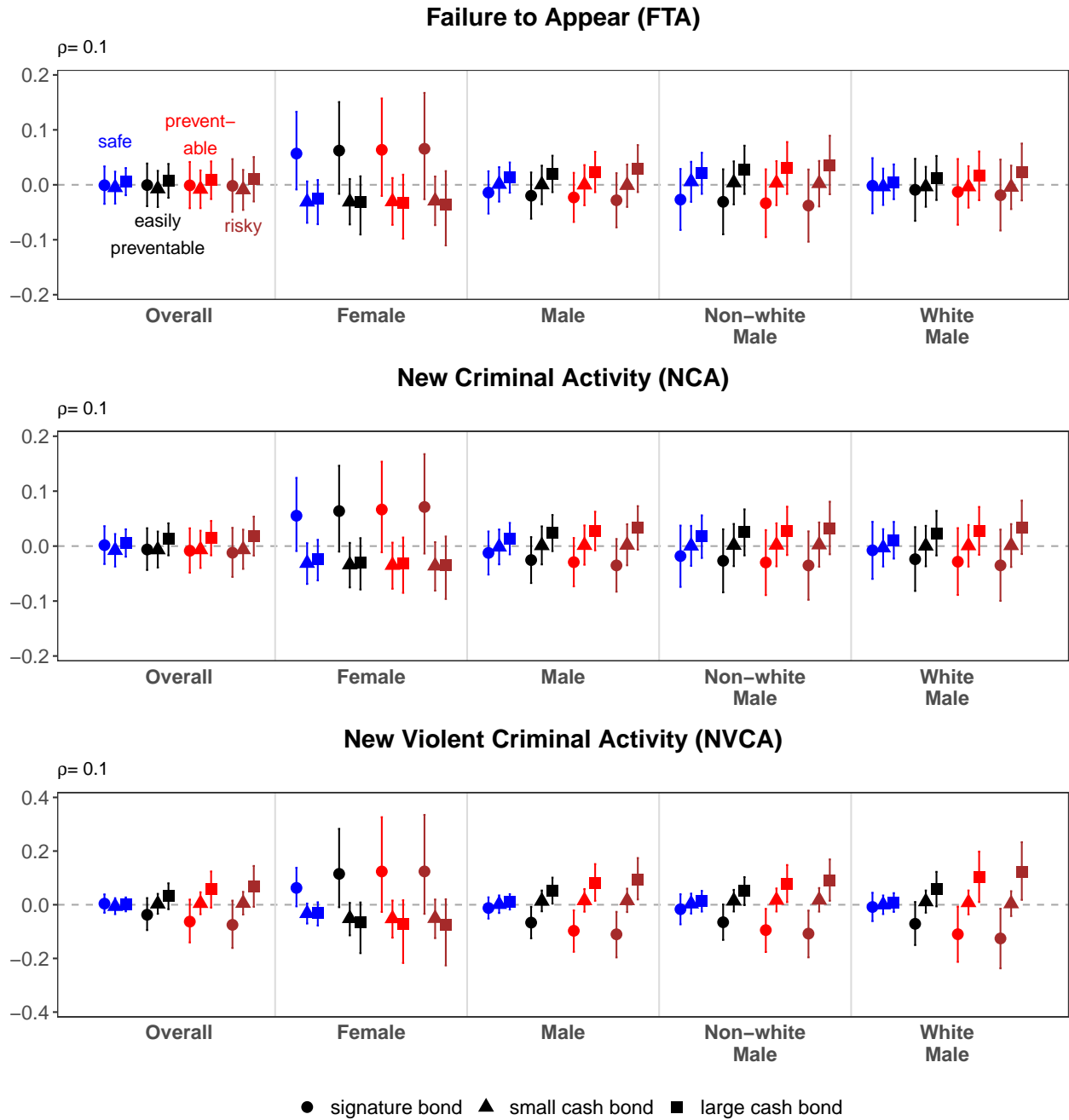


Figure S14: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision with  $\rho = 0.1$ . Each panel presents the overall and subgroup-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the Bayesian 95% credible interval.

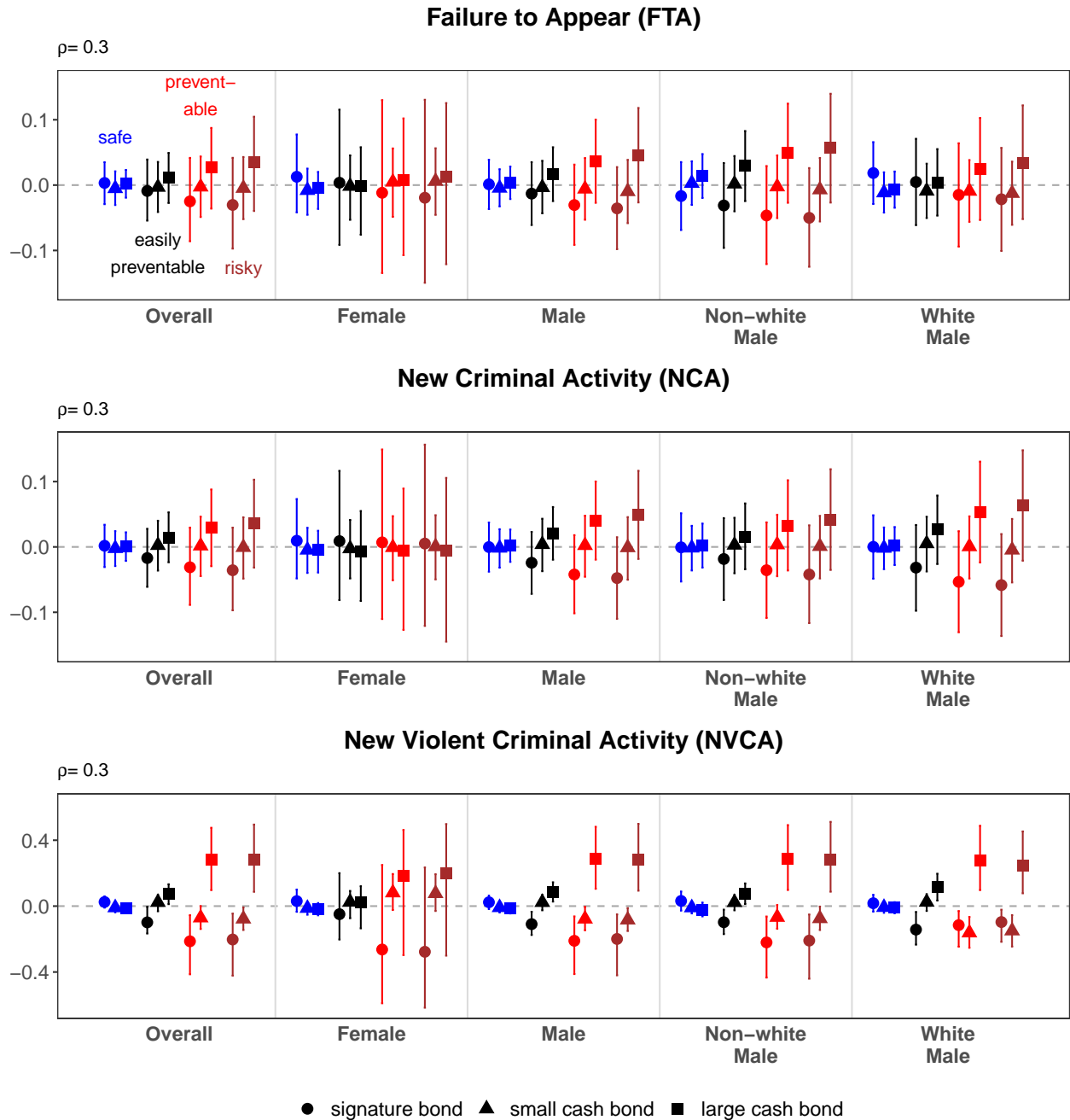
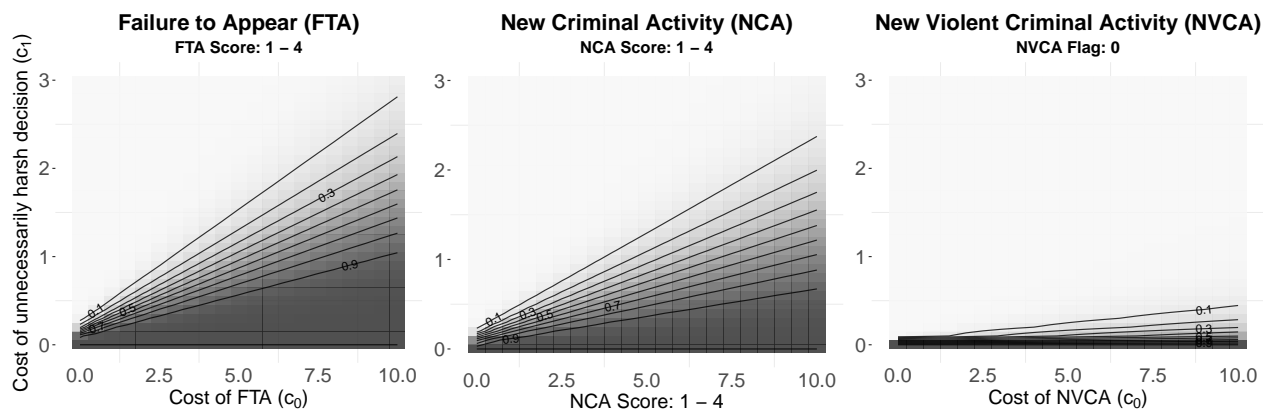


Figure S15: Estimated Average Principal Causal Effects (APCE) of PSA Provision on the Judge’s Decision with  $\rho = 0.3$ . Each panel presents the overall and subgroup-specific results for a different outcome variable. Each column within a panel shows the estimated APCE of PSA provision for safe (blue), easily preventable (black), preventable (red), and risky (brown) cases. For each of these principal strata, we report the estimated APCE on the judge’s decision to impose a signature bond (circles), a small cash bail amount of 1,000 dollars or less (triangles), and a large cash bail amount of greater than 1,000 (squares). The vertical line for each estimate represents the Bayesian 95% credible interval.

## S10 Additional Results for Optimal Decision

(a) The cases whose DMF recommendation is a signature bond



(b) The cases whose DMF recommendation is a cash bond

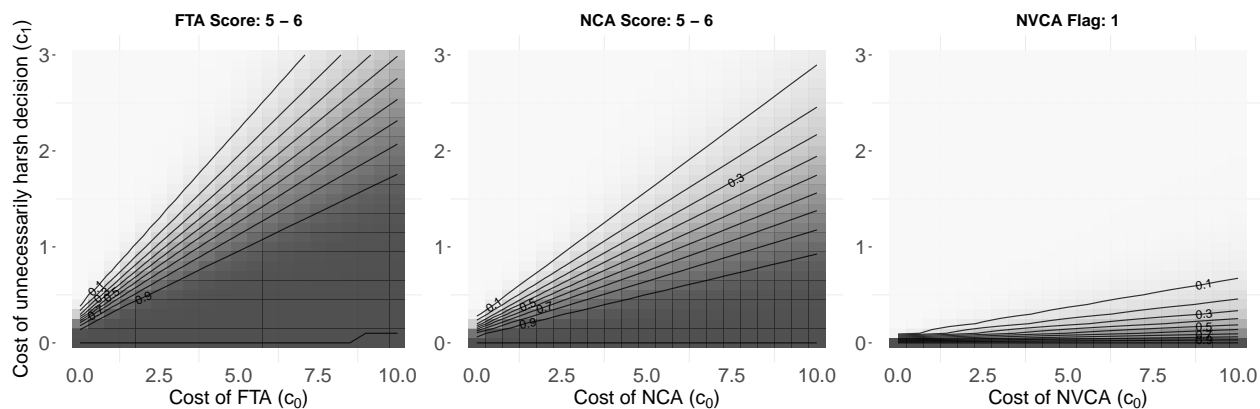


Figure S16: Estimated Proportion of Cases for Which Cash Bond is Optimal. Each column represents the results based on one of the three outcomes (FTA, NCA, and NVCA). The top (bottom) panel shows the results for the cases whose DMF recommendation is a signature (cash) bond. Unlike Figure 6, which uses the overall DMF recommendation, the results are based on the separate DMF recommendation for each outcome. In each plot, the contour lines represent the estimated proportion of cases, for which a cash bond is optimal, given the cost of an unnecessarily harsh decision ( $c_1$ ;  $y$ -axis) and that of a negative outcome ( $c_0$ ;  $x$ -axis). A grey area represents a greater proportion of such cases.

# S11 Additional Results for the Comparison between Judge's Decisions and DMF Recommendations

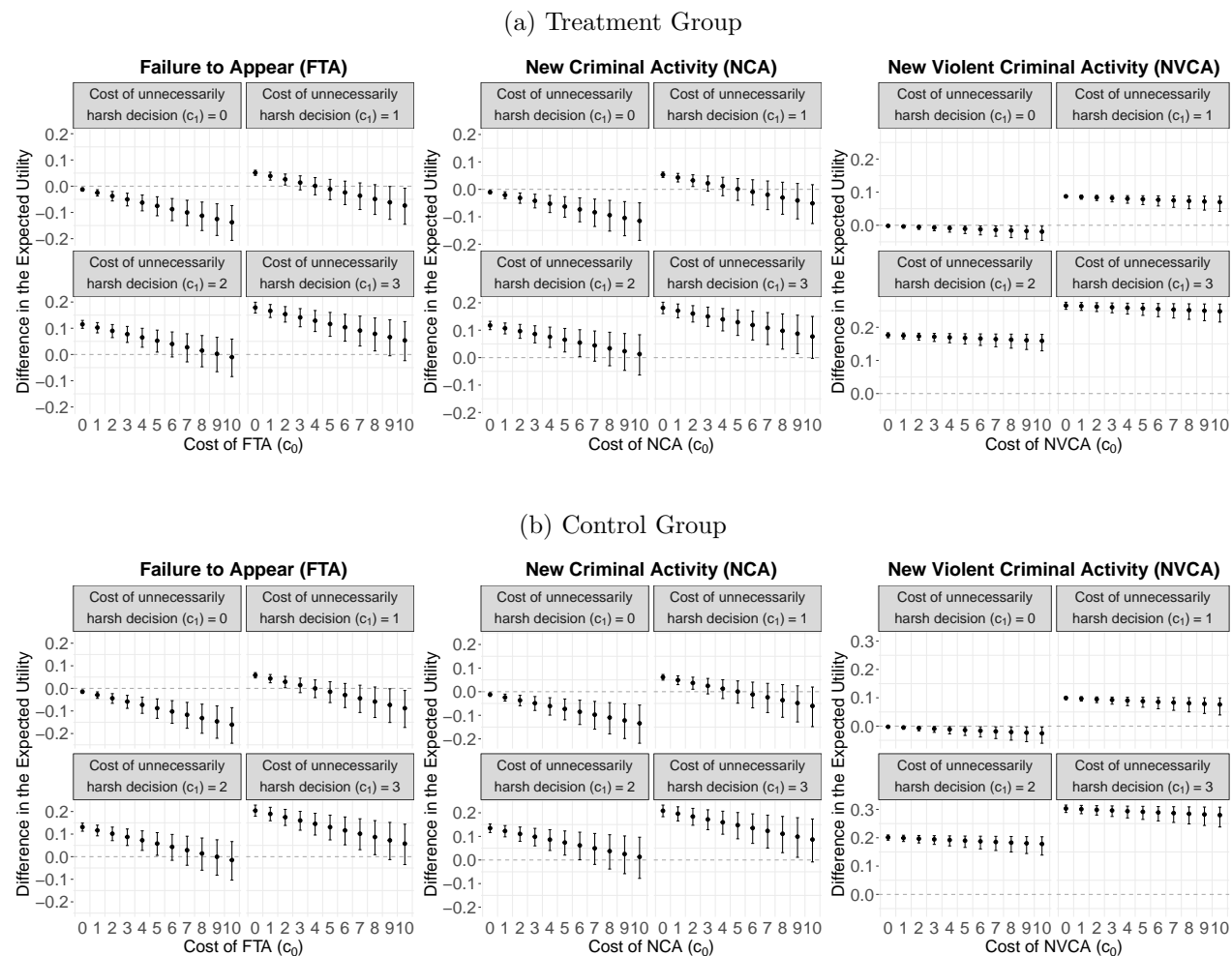


Figure S17: Estimated Difference in the Expected Utility under Selected Values of Cost Parameters between Judge's Decisions and DMF Recommendations for the Treatment (top row) and Control (bottom row) Group. Each column represents the results base on one of the three outcomes, given the cost of an unnecessarily harsh decision ( $c_1$ ; each panel) and that of a negative outcome ( $c_0$ ;  $x$ -axis). A positive value implies that the judge's decision yields a higher expected utility (i.e., more optimal) than the corresponding DMF recommendation. The vertical line for each estimate represents the Bayesian 95% credible interval.