June 2009

DOES SEX OFFENDER REGISTRATION AND NOTIFICATION REDUCE CRIME? A SYSTEMATIC REVIEW OF THE RESEARCH LITERATURE

The Washington State Institute for Public Policy (Institute) was asked by the Sex Offender Policy Board to evaluate the effectiveness of sex offender registration and community notification laws.[1] We conducted a systematic review of the research literature to determine whether these laws reduce crime.

During the 1990s, all states adopted sex offender registration/notification laws. In Washington, the 1990 Legislature passed the Community Protection Act requiring convicted sex offenders who release from custody or are under community supervision and reside in Washington to register with local law enforcement.[2] The 1990 law also authorized officials to notify the public when dangerous sex offenders are released into the community.[3] These measures were intended to "restrict the access of known sex offenders to vulnerable populations, and also to improve law enforcement's ability to identify convicted offenders."[4] The law applies to adults and juveniles convicted of any sex offense.

Four years after Washington passed its registration/notification laws, Congress passed the 1994 Jacob Wetterling Act, requiring states to implement a sex offender registry.[5] The federal act was amended in 1996 to require community notification of offenders convicted

Summary

The Washington State Institute for Public Policy was asked by the Sex Offender Policy Board to evaluate the effectiveness of sex offender registration and community notification laws on reducing crime. We conducted a systematic review of all research evidence throughout the United States and located nine rigorous evaluations.

Seven of these studies address whether the laws influence "specific" deterrence—the effect of a law on the recidivism rates of convicted sex offenders. The other two studies analyze "general" deterrence—the effect of a law on sex offense rates of the general public, as well as recidivism rates of convicted sex offenders.

Regarding specific deterrence, the weight of the evidence indicates the laws have no statistically significant effect on recidivism. This finding, however, should be regarded with caution since we only found seven credible studies and these studies have widely varying results. Additionally, three of the studies have small sample sizes. Thus, at this time, we tentatively conclude that existing research does not offer much policy guidance on the specific deterrent effect of registration/notification laws.

For general deterrence, the two studies provide some indication that registration laws lower sex offense rates in the public at large. Again, caution is warranted when generalizing this result since it is based on only two studies.

Additional research is necessary before definitive conclusions can be drawn.

---

[1] The 2008 Legislature created the Sex Offender Policy Board with the intent that experts and practitioners would coordinate statewide sex offender management and assess the performance of the system's components (RCW 9.94A.8671). The Sex Offender Policy Board is housed within the Sentencing Guidelines Commission. The Institute's Board of Directors approved the contract for this project on June 1, 2009.
[2] RCW 9A.44.130
[3] RCW 4.24.550
[4] Washington Department of Social and Health Services. (1989). *Task Force on Community Protection: Final Report to Booth Gardner, Governor State of Washington.* Olympia: Author.
[5] http://www.ojp.usdoj.gov/BJA/what/2a1jwacthistory.html

of crimes against children or sexually violent offenses.[6] Thus, all 50 states now maintain sex offender registries and have some form of community notification legislation. In July 2006, President Bush signed the Adam Walsh Child Protection and Safety Act, which further standardized state laws.[7]

Theoretical Foundation

There are many reasons why registration and notification laws exist, but one theoretical foundation is deterrence. The principle of deterrence posits that "individuals choose to obey or violate the law by rational calculation."[8] Criminologists often classify deterrence as either "specific" or "general."

Specific deterrence refers to the effect that punishment has on an offender's subsequent criminality. For example, an offender who spent time in confinement for a crime may make the choice not to commit crimes in the future because the individual has experienced punishment in the past.

General deterrence, on the other hand, refers to the effect that punishment has on the general population. For example, an individual may make the choice to remain crime-free because the threat of punishment prevents him or her from committing a crime.

In our review of the research on the impacts of sex offender registration and notification policies on crime, we found that some studies address specific deterrence while others address general deterrence. We present our findings using this framework.

Notification Levels in Washington

In Washington State, sex offenders are assigned a risk level of I, II, or III, with level III offenders being the most likely to reoffend.[9] Decisions regarding the community notification level are first considered by the End of Sentence Review Committee (ESRC). The ESRC was established by the Legislature in 1997 to review the risk level of sex offenders prior to an offender's release from prison.[10] The Secretary of the Department of Corrections (DOC) appoints the ESRC chair who then appoints representatives from state and local agencies.

The ESRC's classification decision is based on two actuarial assessments that predict risk for sexual reoffense: the Minnesota Sex Offender Screening Tool (MnSOST) and the STATIC-99.[11] The ESRC recommends a risk level classification to local law enforcement who ultimately determine the level communicated to the public.

Local law enforcement agencies notify the media, individuals, and organizations in the community regarding released sex offenders assessed at levels II and III. Additionally, these offenders are listed on a website maintained by the Washington Association of Sheriffs and Police Chiefs with information from the Washington State Patrol and the DOC.

---

[6] Ibid.

[7] H. Res. 4472, 109th Cong., 152 Cong. Rec. (2006) (enacted).

[8] R. Akers (1997). *Criminological theories: Introduction and evaluation.* Los Angeles: Roxbury Publishing Company, p. 16.

---

[9] Some research indicates, however, that the risk levels do not necessarily correspond accurately with risk for reoffense. See R. Barnoski (2005). *Sex offender sentencing in Washington State: Notification levels and recidivism.* Olympia: Washington State Institute for Public Policy, Document No. 05-12-1203.

[10] RCW 72.09.345

[11] The ESRC began using this classification system in April 2009. Prior to that time, classification decisions were based on the Washington State Sex Offender Risk Level Classification, which consisted of 21 items that constitute the MnSOST (1995), in addition to four "notification considerations" that were not part of the standard MnSOST instrument.

Study Methods and Findings

The goal of this study is to answer a simple question: Does sex offender registration and notification affect measured crime outcomes? To answer this question, we conducted a comprehensive statistical review of evaluations from over the last 20 years in the United States. The accompanying sidebar "What Does 'Evidence-Based' Mean?" briefly describes the factors we consider in determining the applicability of a particular study for our systematic review. Our research methodology is described fully in the Technical Appendix of this report.

We located nine evaluations of sex offender registration/notification laws with sufficiently rigorous research to be included in our analysis.[12] Seven of these evaluations measured the specific deterrent effect of the laws and two measured the general deterrent effect.[13]

Specific Deterrence

The seven evaluations addressing the specific deterrent effect of sex offender registration/ notification laws typically evaluated a law's effectiveness by comparing recidivism rates of registered sex offenders who release from prison with sex offenders prior to the implementation of the law.

Exhibit 1 displays the results of the seven specific deterrence studies included in our analysis. The studies are grouped into two categories of outcomes: sex offense recidivism and total offense recidivism. Five of the studies focus on adult offenders and two on juvenile offenders.

What Does "Evidence-Based" Mean?

At the direction of the Washington Legislature, the Institute has conducted a number of systematic reviews of evaluation research to determine what public policies and programs work, and which ones do not work. These evidence-based reviews include the policy areas of adult and juvenile corrections, child welfare, mental health, substance abuse, prevention, K–12 education, and pre-K education.

The phrase "evidence-based" is sometimes used loosely in policy discussions. When the Institute is asked to conduct an evidence-based review, we follow a number of steps to ensure a rigorous definition. These criteria include the following:

1) We consider all available studies we can locate on a topic rather than selecting only a few; that is, we do not "cherry pick" studies to include in our reviews. When possible, we then use formal statistical hypothesis testing procedures—meta-analysis—to determine whether the weight of the evidence indicates outcomes are, on average, achieved.

2) To be included in our reviews, we require that an evaluation's research design include control or comparison groups. Random assignment studies are preferred, but we allow quasi-experimental studies when the study uses appropriate statistical techniques.

Our additional criteria are listed in the Technical Appendix.

Results indicate no clear pattern. One study found increased rates of recidivism, two found decreases in recidivism, and four found no statistically significant differences.

For this group of studies, we performed a meta-analysis and found no statistically significant difference in recidivism rates for either sex offenses or total offenses.[14]

---

[12] It is important to note that only two of these nine studies were evaluations of registration/notification laws in Washington State; the remaining evaluations were conducted in other locations. A primary purpose of our study is to take advantage of all these rigorous evaluations and, thereby, learn whether sex offender registration/notification laws impact crime.

[13] Citations of the studies included in this report are located in the Technical Appendix of this report, as are our meta-analytic procedures.

[14] For details of the meta-analysis, see Exhibit A in the Technical Appendix.

This statistically insignificant finding, however, should be regarded with caution. To date, the research literature is limited; we were only able to find seven credible studies to include in our analysis, and three of the studies had quite small sample sizes. Thus, a tentative conclusion is that existing research, focusing on the specific deterrent effect of registration/notification laws, does not offer much policy guidance at this point in time. Additional research studies will be required before more definitive conclusions can be drawn.

Exhibit 1
Specific Deterrence Studies Included in the Analysis and the Effect on Crime Outcomes

Author and Year of Publication	Agan, 2008	Barnoski, 2005[a]	Duwe & Donnay, 2008	Freeman, 2009	Letourneau & Armstrong, 2008	Letourneau, Bandyopadhyay, Sinha & Armstrong, 2008	Schram & Milloy, 1995
State(s) of Research	14 States[b]	Washington	Minnesota	New York	South Carolina	South Carolina	Washington
Population Type	adults	adults	adults	adults	juveniles	juveniles	adults
Offenders in Registration/Notification Group	4,488	5,831	155	10,592	111	574	90
Offenders in Comparison Group	5,135	2,528	125	6,573	111	701	90
	Effect on Crime Outcomes						
Statistical Effect on Sex Offenses[c]	no difference	significant decrease	significant decrease	significant increase	not measured	no difference	no difference
Statistical Effect on Total Offenses	not measured	no difference	significant decrease	significant increase	no difference[e]	no difference	no difference
Crime Outcome Measure[d]	convictions	convictions	convictions	arrests	convictions	convictions	arrests

[a] Evidence in the literature indicates that controlling for time helps to further isolate the effects of registration/notification (see specifically, Shao & Li, 2006). For the current report, we were able to reanalyze Barnoski's 2005 dataset, which did not originally control for time trends. The re-analysis indicates the effect would be smaller than stated in the original Barnoski 2005 study, but still a negative, though statistically insignificant, effect on sex offenses.
[b] States include California, Delaware, Florida, Illinois, Maryland, Michigan, Minnesota, New Jersey, New York, North Carolina, Ohio, Oregon, Texas, and Virginia.
[c] The definition of sex offenses may vary across outcome evaluations.
[d] If authors report more than one type of crime outcome (i.e., arrests, convictions), convictions is our preferred outcome measure.
[e] The authors measure person and nonperson crimes. This effect is the average of those two outcomes.

General Deterrence

General deterrence studies use aggregate-level data such as the Uniform Crime Reports (UCR) or National Incident Based Reporting System (NIBRS) to determine if changes in crime rates have occurred over time as a result of a new policy. Typically, the studies compare crime rates prior to and after implementation of the registration/notification laws. The advantage of general deterrence studies is that they measure overall adult and juvenile crime. Additionally, these studies measure the effect of a policy change on both convicted sex offenders and the general population.

After a thorough search, we located two general deterrence studies that evaluate the impact of registration/notification on crime: Prescott and Rockoff (2008) and Shao and Li (2006). This small number of studies does not allow us to conduct a meta-analysis, but we summarize the authors' findings.

Prescott and Rockoff use NIBRS data to examine the impacts of registration/notification laws in 15 states.[15] Their analysis controls for a comprehensive set of factors including crime rates, county income, demographics, and the heterogeneity of reporting jurisdictions and time. Prescott and Rockoff are the first to examine how the size of a state's sex offender registry may influence crime. They are also the first and only authors to examine, separately, the impacts of registration versus notification.

Overall, Prescott and Rockoff's findings imply that the average registration/notification law produces a statistically significant 10 percent reduction in sex offense rates.[16] That is, taken at the average size of 15 registered sex offenders per 10,000 persons, as reported by Prescott and Rockoff, their finding indicates a 10 percent reduction in sex offense rates.

This result is nuanced, however, because the authors found that registration and notification

have opposing effects. Registration has a statistically significant negative relationship with sex offenses after taking into account the size of the registry. That is, when the authors isolate the effect of registration laws, they find that sex offenses decrease as the size of the registry increases.

This benefit, however, is moderated by the authors' findings on notification laws. Notification has a statistically significant positive relationship with sex offenses after taking into account the size of the registry. That is, when the authors isolate the effect of notification laws, they find that sex offenses increase as the size of the registry increases.

Since all states now have both registration and notification laws, the net effect from Prescott and Rockoff's study would indicate an overall 10 percent decrease in sex offenses. The net effect is negative because registration has a greater effect than notification.

The second general deterrence study included in our analysis was conducted by Shao and Li (2006).[17] The authors use UCR panel data for all 50 states from 1970 to 2002 to investigate the impact of registration laws on reported rapes to police. Shao and Li find a statistically significant negative relationship between registration and reported rapes. The magnitude of their estimate is a 2 percent reduction in reported rapes.[18]

Taken together, these two general deterrence studies provide some indication that sex offender registration laws lower sex offense crime rates. Some caution, however, is warranted when generalizing this result since we only found two rigorous general deterrence studies. Additional research will help develop knowledge on this topic.

---

[15] Of the sex offender registration/notification literature, Prescott and Rockoff (2008) utilize the most rigorous design to date.
[16] Based on Prescott & Rockoff's preferred regression results, we estimate the average effect taken at the mean registry size for a state with both registration and notification laws.

[17] The methodological rigor of Shao and Li's design is also very high quality. The authors control for many factors such as state population, state fixed effects, year fixed effects, and state-specific linear trends.
[18] Shao and Li also examine the effects of the registry over time and find that the negative relationship increases. This finding is consistent with Prescott and Rockoff's analysis that as the size of the registry increases, sex offenses decrease.

1. Research Methodology

The goal of this research is to answer a simple question: Does sex offender registration and notification affect measured crime outcomes? Specifically, does rigorous evaluation evidence indicate that registration and notification laws lower crime rates? To answer this question, we conducted a comprehensive statistical review of evaluations conducted over the last 20 years in the United States. We located nine evaluations of sex offender registration/notification with sufficiently rigorous research to be included in our analysis. Full citations of the studies and summaries are located in Appendix B of this report. We also include summaries of studies that did not meet our minimum methodological standard of rigor.

The research approach we employ is called a "systematic" review of the evidence. In a systematic review, the results of *all* rigorous evaluation studies are analyzed to determine if, on average, it can be stated scientifically that a policy or program achieves an outcome. A systematic review can be contrasted with a "narrative" review of the literature where a writer selectively cites studies to tell a story about a topic. Both types of reviews have their place, but systematic reviews are generally regarded as more rigorous and, because they assess all available studies and employ statistical hypotheses tests, they have less potential for drawing biased or inaccurate conclusions.

In our review of the evidence, we only include "rigorous" evaluation studies. The key criterion for a study to be included is that the evaluation must have a non-treatment or treatment-as-usual comparison group that is well matched to the program group.[19] Appendix B describes the factors we consider in determining the applicability of a particular study for our systematic review.

Researchers have developed a set of statistical tools, called "meta-analysis," to facilitate systematic reviews of the evidence. The technique to portray findings from a meta-analysis is the "effect size" (procedures for calculating the effect size are located later in this Appendix). Effect sizes measure the degree to which a program or law has been shown to change an outcome for participants relative to the comparison group. Calculation of effect sizes allows one to compare the degree of impact across studies. A negative effect size shows a decrease in the measured outcome and a positive effect size shows an increase in the measured outcome. The larger the effect size, the larger the impact on the measured outcome.

Effect sizes of individual studies are displayed in **Exhibit A.** Our meta-analytic findings indicate no statistically significant difference in recidivism rates for either sex offenses or total offenses.

2. Study Selection and Coding Criteria

Search Strategy
We search for all adult and juvenile evaluation studies conducted in the past 20 years are written in English. We use three primary means to identify and locate these studies: (a) we consult the study lists of other systematic and narrative reviews of the adult and juvenile corrections and prevention research literature; (b) we examine the citations in the individual evaluations; and (c) we conduct independent literature searches of research databases using search engines such as Google, Proquest, Ebsco, ERIC, and SAGE. We obtain and examine copies of all individual program evaluation studies we can locate using these search procedures.

Many of these studies are published in peer-reviewed academic journals, while others are from government reports obtained from the agencies themselves. It is important to include non-peer reviewed studies, because it has been suggested that peer-reviewed publications may be biased to show positive program effects.[20] Therefore, our meta-analysis includes all available studies we could locate regardless of published source.

---

[19] A non-treatment comparison group is typical when evaluating programs that have participants and non-participants during the same time period. Since law changes affect everyone at the same time, comparison groups for evaluations of laws typically include people prior to the law change. We found several highly rigorous studies that use econometric methods to evaluate the impacts of sex offender registration/notification, thus they do not use "comparison groups" in the usual program evaluation sense.

[20] M. W. Lipsey & D. B. Wilson (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Exhibit A
Individual Effect Sizes for Rigorous Studies Included in Systematic Review

Author and Year of Publication	State(s) of Research	Number in Treatment	Number in Comparison	Effect Size	Weighted Mean Effect Size[a]
Specific Deterrence Studies					
Sex Offense Outcomes					
Agan, 2008	14 States[b]	4,488	5,135	0.008	-.212 (p=.359)
Barnoski, 2005	Washington	5,831	2,528	-0.442	
Duwe & Donnay, 2008	Minnesota	155	125	-2.044	
Freeman, 2009	New York	10,592	6,573	0.436	
Letourneau et al., 2008	South Carolina	574	701	0.303	
Schram & Milloy, 1995	Washington	90	90	-0.111	
Total Offense Outcomes					
Barnoski, 2005	Washington	5,831	2,528	0.040	.037 (p=.721)
Duwe & Donnay, 2008	Minnesota	155	125	-0.742	
Freeman, 2009	New York	10592	6573	0.260	
Letourneau & Armstrong, 2008[c]	South Carolina	111	111	0.190	
Letourneau et al., 2008	South Carolina	574	701	0.188	
Schram & Milloy, 1995	Washington	90	90	0.243	
General Deterrence Studies					
Sex Offense Outcomes					
Prescott & Rockoff, 2008	15 states[d]	N/A	N/A	N/A	
Shao & Li, 2006	50 States	N/A	N/A	N/A	

[a] We test for homogeneity of the weighted mean effect sizes by calculating the random effects variance. Both sex offense and total offense outcomes indicate heterogeneity, thus a random effects model is used to account for the random variance factor.
[b] States include California, Delaware, Florida, Illinois, Maryland, Michigan, Minnesota, New Jersey, New York, North Carolina, Ohio, Oregon, Texas, and Virginia.
[c] Since there are only two general deterrence studies, we did not conduct a meta-analysis.
[d] States include Colorado, Connecticut, Idaho, Iowa, Kentucky, Massachusetts, Michigan, Nebraska, North Dakota, Ohio, South Carolina, Texas, Utah, Vermont, and Virginia.

Criteria for Inclusion and Exclusion of Studies
Comparison Group. The most important inclusion criterion in our systematic review of the literature is that an evaluation must have a control or comparison group. We do not include studies with a single-group, pre-post research design in order to avoid false inference on causality.[21] Random assignment studies are preferred for inclusion in our review, but we also include non-randomly assigned control groups. We only include quasi-experimental studies if sufficient information is provided to demonstrate reasonable comparability between the treatment and comparison groups on important pre-existing conditions such as age, gender, and prior criminal history. Of the nine individual studies in our review, none of the effects were estimated from well-implemented random assignment studies.

Participant Sampling Procedures. We do not include a study in our meta-analytic review if the treatment group is made up solely of program completers. We adopt this rule to avoid unobserved self-selection factors that distinguish a program completer from a program dropout; these unobserved factors are likely to significantly bias estimated treatment effects.[22] Some comparison group studies of program completers, however, contain information on program dropouts in addition to a comparison group. In these situations, we include the study if sufficient information is provided to allow us to reconstruct an intent-to-treat group that includes both completers and non-completers, or if the demonstrated rate of program non-completion is very small (e.g., under 10 percent). In these cases, the study still needs to meet the other inclusion requirements listed here.

[21] Coalition for Evidence-Based Policy (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* Washington, DC: The Council for Excellence in Government, Author.

[22] M. W. Lipsey (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The ANNALS of the American Academy of Political and Social Science, 587*(1), 69-81.

Outcomes. A crime-related outcome must be reported in the study to be included in our review. Some studies present several types of crime-related outcomes. For example, studies frequently measure one or more of the following outcomes: total arrests, total convictions, felony arrests, misdemeanor arrests, violent arrests, and so on. In these situations, we code the broadest crime outcome measure. Thus, most of the crime outcome measures that we code are total arrests and total convictions. When a study reports both total arrests and total convictions, we calculate an effect size for each measure and then take a simple average of the two effect sizes.

Some studies include two types of measures for the same outcome: a dichotomous outcome and a continuous (mean number) measure. In these situations, we code an effect size for the dichotomous measure. Our rationale for this choice is that in small or relatively small sample studies, continuous measures of crime outcomes can be unduly influenced by a small number of outliers, while dichotomous measures can reduce this problem.[23] Of course, if a study only presents a continuous measure, we code the continuous measure.

When a study presents outcomes with varying follow-up periods, we generally code the effect size for the longest follow-up period. This allows us to gain the most insight into the long-run benefits and costs of various treatments. Occasionally, we do not use the longest follow-up period if it is clear that a longer reported follow-up period adversely affected the attrition rate of the treatment and comparison group samples.

Miscellaneous Coding Criteria. Our unit of analysis is an independent test of a treatment at a particular site. Some studies report outcomes for multiple sites; we include each site as an independent observation if a unique and independent comparison group is also used at each site.

Some studies present two types of analyses: raw outcomes that are not adjusted for covariates such as age, gender, or criminal history; and those that have been adjusted with multivariate statistical methods. In these situations, we code the multivariate outcomes.

3. Procedures for Calculating Effect Sizes

Calculations for Dichotomous and Continuous Outcomes. Effect sizes measure the degree to which a program has been shown to change an outcome for program participants relative to a comparison group. In order to be included in our review, a study has to provide the necessary information to calculate an effect size. Several methods can be used by meta-analysts to calculate effect sizes. We use the standardized mean difference effect size for continuous measures and the D-cox transformation as described in Equation 18 of Sánchez-Meca, Chacón-Moscoso, and Marín-

Martínez[24] to approximate the mean difference effect size for dichotomous outcome variables.

$$(1): \quad d_{Cox} = \ln\left(\frac{P_e(1-p_c)}{P_c(1-p_e)}\right)/1.65$$

In Equation 1, d_{cox} is the estimated effect size, which is derived by dividing the log odds ratio by the constant 1.65. P_e, represents the percentage outcome for the experimental or treatment group and, P_c, is the percentage outcome for the control group.

For continuous outcome measures, we use the standardized mean difference effect size statistic.[25]

$$(2): \quad ES_m = \frac{M_e - M_c}{\sqrt{\frac{SD_e^2 + SD_c^2}{2}}}$$

In the second equation, ES_m is the estimated standardized mean effect size where M_e is the mean outcome for the experimental group, M_c is the mean outcome for the control group, SD_e is the standard deviation of the mean outcome for the experimental group, and SD_c is the standard deviation of the mean outcome for the control group.

Sometimes research studies report the mean values needed to compute ES_m in Equation 2, but they fail to report the standard deviations. Often, however, the research will report information about statistical tests or confidence intervals that can then allow the pooled standard deviation to be estimated.[26]

Some studies have very small sample sizes, and small sample sizes have been shown to upwardly bias effect sizes, especially when samples are less than 20. Therefore, we follow Hedges[27] and Lipsey and Wilson[28] and report the "Hedges correction factor," which we use to adjust all mean difference effect sizes (N is the total sample size of the combined treatment and comparison groups):

$$(3): \quad ES'_m = \left[1 - \frac{3}{4N-9}\right] \times \left[ES_m, or, d_{cox}\right]$$

[23] D. Farrington & R. Loeber (2000). Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health, 10*, 100-122.

[24] J. Sánchez-Meca, S. Chacón-Moscoso, & F. Marín-Martínez (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448-467.
[25] Lipsey & Wilson, op. cit., Table B10, Equation 1.
[26] These procedures are further described in Lipsey & Wilson, op. cit.
[27] L. V. Hedges (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.
[28] Lipsey & Wilson, op. cit., Equation 3.22.

4. Techniques Used to Combine the Evidence

Once effect sizes are calculated for each program effect, the individual measures are summed to produce a weighted average effect size for a program area. We calculate the inverse variance weight for each program effect and these weights are used to compute the average. These calculations involve three steps. First, we calculate the standard error of each mean effect size. For continuous outcomes, the standard error, SE_m, is computed with:[29]

$$(4): \quad SE_m = \sqrt{\frac{n_e + n_c}{n_e n_c} + \frac{(ES'_m)^2}{2(n_e + n_c)}}$$

In Equation 4, n_e and n_c are the number of participants in the experimental and control groups and ES'_m is from Equation 3.

For dichotomous outcomes, the standard error, SEd_{cox}, is computed with:[30]

$$(5) \quad SE_{d_{Cox}} = \sqrt{0.367\left[\frac{1}{O_{1E}} + \frac{1}{O_{2E}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}}\right]}$$

In Equation 5, O_{1E} and O_{1C}, represent the success frequencies of the experimental and control groups. O_{2E} and O_{2C}, represent the failure frequencies of the experimental and control groups.

The second step in calculating the average effect size for a program area is to compute the inverse variance weight w_m for each mean effect size with:[31]

$$(6): \quad w_m = \frac{1}{SE_m^2}$$

The weighted mean effect size for a group of studies is then computed with:[32]

$$(7): \quad \overline{ES} = \frac{\sum (w_m ES'_m)}{\sum w_m}$$

Finally, confidence intervals around this mean are computed by first calculating the standard error of the mean with:[33]

$$(8): \quad SE_{\overline{ES}} = \sqrt{\frac{1}{\sum w_m}}$$

The lower, ES_L, and upper limits, ES_U, of the confidence interval are computed with:[34]

$$(9): \quad \overline{ES_L} = \overline{ES} - z_{(1-\alpha)}(SE_{\overline{ES}})$$

$$(10): \quad \overline{ES_U} = \overline{ES} + z_{(1-\alpha)}(SE_{\overline{ES}})$$

In Equations 9 and 10, $z_{(1-\alpha)}$ is the critical value for the z-distribution.

Techniques Used to Assess Heterogeneity.
Computing Random Effects Weighted Average Effect Sizes and Confidence Intervals. Once the weighted mean effect size is calculated, we test for homogeneity. This provides a measure of the dispersion of the effect sizes around their mean and is given by:[35]

$$(11): \quad Q = \left(\sum w\, ES^2\right) - \frac{\left(\sum wES\right)^2}{\sum w}$$

The Q-test is distributed as a chi-square with k-1 degrees of freedom (where k is the number of effect sizes). When the p-value on the Q-test indicates significance at values of p less than or equal to .05, a random effects model is performed to calculate the weighted average effect size. This is accomplished by first calculating the random effects variance component, v:[36]

$$(12): \quad v = \frac{Q - (k-1)}{\sum w - \left(\sum wsq / \sum w\right)}$$

This random variance factor is then added to the variance of each effect size and all inverse variance weights are recomputed, as are the other meta-analytic test statistics.

Adjustments to Effect Sizes
Methodological Quality. Not all research is of equal quality and this greatly influences the confidence that can be placed in interpreting the policy-relevant results of a study. Some studies are well designed and implemented and the results can be reasonably viewed as causal effects. Other studies are not designed as well and less confidence can be placed in the causal interpretation of any reported differences. Studies with inferior research designs cannot completely control for sample selection bias or other unobserved threats to the validity of reported research results. This does not mean that results from these studies are of no value, but it does mean that less confidence can be placed in any cause-and-effect conclusions drawn from the results.

To account for the differences in the quality of research designs, we use a 5-point scale as a way to adjust the raw effect sizes. The scale is based closely on the 5-point scale developed by researchers at the University of

---

[29] Lipsey & Wilson, op. cit., Equation 3.23.
[30] Sánchez-Meca et al., op. cit., Equation 19.
[31] Lipsey & Wilson, op. cit., Equation 3.24.
[32] Lipsey & Wilson, op. cit., p. 114.
[33] Ibid.

[34] Ibid.
[35] Lipsey & Wilson, op. cit., p. 116.
[36] Lipsey & Wilson, op. cit., p. 134.

Maryland.[37] On the 5-point scale as interpreted by the Institute, each study is rated with the following numerical ratings.

A **"5"** is assigned to an evaluation with well-implemented random assignment of subjects to a treatment group and a control group that does not receive the treatment/ program. A good random assignment study should also report how well the random assignment actually occurred by reporting values for pre-existing characteristics for the treatment and control groups.

A **"4"** is assigned to a study that employs a rigorous quasi-experimental research design with a program and matched comparison group, controlling with statistical methods for self-selection bias that might otherwise influence outcomes. These quasi-experimental methods may include estimates made with a convincing instrumental variables or regression discontinuity modeling approach or other techniques such as a Heckman self-selection model.[38] A value of 4 may also be assigned to an experimental random assignment design that reported problems in implementation, perhaps because of significant attrition rates.

A **"3"** indicates a non-experimental evaluation where the program and comparison groups are reasonably well matched on pre-existing differences in key variables. There must be evidence presented in the evaluation that indicates few, if any, significant differences were observed in these salient pre-existing variables. Alternatively, if an evaluation employs sound multivariate statistical techniques to control for pre-existing differences, and if the analysis is successfully completed and reported, then a study with some differences in pre-existing variables can qualify as a level 3.

A **"2"** involves a study with a program and matched comparison group where the two groups lack comparability on pre-existing variables and no attempt was made to control for these differences in the study. A **"1"** involves an evaluation study where no comparison group is utilized.

In our meta-analytic review, we only consider evaluations that rate at least a 3 on this 5-point scale. We do not use the results from program evaluations rated as a "1" on this scale, because they do not include a comparison group and thus provide no context to judge program effectiveness. We also regard evaluations with a rating of "2" as highly problematic and, as a result, do not consider their findings in our analyses.

An explicit adjustment factor is assigned to the results of individual effect sizes based on the Institute's judgment concerning research design quality. The specific adjustments made for these studies are based on our knowledge of research in particular fields. For example, in criminal justice program evaluations, there is strong evidence that random assignment studies (i.e., level 5 studies) have, on average, smaller absolute effect sizes than studies with weaker designs.[39] We use the following default adjustments to account for studies of different research design quality. The effect size of a level 3 study is discounted by 50 percent and the effect size of a level 4 study is discounted by 25 percent, while the effect size of a level 5 study is not discounted. While these factors are subjective, we believe not making some adjustments for studies with varying research design quality would severely over-estimate the true causal effect of the average program.

Researcher Involvement in the Program's Design and Implementation. The purpose of the Institute's work is to identify and evaluate programs that can make cost-beneficial improvements to Washington's actual service delivery system. There is some evidence that programs closely controlled by researchers or program developers have better results than those that operate in "real world" administrative structures.[40] For example, in our evaluation of a real-world implementation of a research-based juvenile justice program in Washington, we found that the actual results were considerably lower than the results obtained when the intervention was conducted by the originators of the program.[41] Therefore, we make an adjustment to effect sizes to reflect this distinction. As a general parameter, the Institute discounts effect sizes by 50 percent for all studies deemed not to be "real world" trials.

[37] L. Sherman, D. Gottfredson, D. MacKenzie, J. Eck, P. Reuter, & S. Bushway (1998). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress* (chap. 2). College Park: University of Maryland at College Park, Department of Criminology and Criminal Justice.
[38] W. Rhodes, B. Pelissier, G. Gaes, W. Saylor, S. Camp, & S. Wallace (2001). Alternative solutions to the problem of selection bias in an analysis of federal residential drug treatment programs. *Evaluation Review, 25*(3), 331-369.

[39] Lipsey, op. cit.
[40] Lipsey (2003); and A. Petrosino & H. Soydan (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1(4), 435-450.
[41] R. Barnoski (2004). *Outcome evaluation of Washington State's research-based programs for juvenile offenders* Olympia: Washington State Institute for Public Policy, Document No. 04-01-1201).

Exhibit B
Citations and Summary of Nine Rigorous Studies Used in the Systematic Review

Author and Year of Publication	Description and Methods	Summary of Findings	Full Citation
Agan, 2007	The author investigates the impact of registration/notification in 14 states (CA, DE, FL, IL, MD, MI, MN, NJ, NY, NC, OH, OR, TX, VA) by conducting an outcome evaluation on sex offenders who released from prison. The author compares the outcomes of offenders required to register to offenders who were not required to register prior to the law change. The author conducts multivariate regression analysis to control for differences between the two groups and also controls for differences across states.	Author finds no statistically significant effect of registration/notification on sex offense convictions.	Agan, A. (2007). *Sex offender registries: Fear without function?* Unpublished manuscript, University of Chicago.
Barnoski, 2005	The author examines the influence of Washington's registration/notification law by conducting an outcome evaluation on sex offenders who released from prison before and after the passage of the law. The author conducts multivariate regression analysis to control for differences between the groups and analyzes felony, sex, and violent felony reconvictions within five years of being at-risk in the community.	The author finds a statistically significant decrease in felony sex and violent felony convictions for offenders subject to Washington's registration/notification laws. The author cautions that the causal link between the laws and crime is not proven by this research and that other factors, such as drops in national and state crime rates, may contribute to the decreases in recidivism. The author finds no difference between the groups for felony recidivism.	Barnoski, R. (2005). *Sex offender sentencing in Washington State. Has community notification reduced recidivism?* Olympia: Washington State Institute for Public Policy, Document No. 05-12-1202.
Duwe & Donnay, 2008	The authors examine the effects of Minnesota's registration/notification law on convicted sex offenders who release from prison. The authors compare the highest risk offenders (level 3) subject to broad community notification to offenders who would have scored as level 3 offenders prior to the law. The authors use survival analysis over an average eight-year follow-up period.	The authors find statistically significant reductions in sexual, non-person, and general reconvictions.	Duwe, G., & Donnay, W. (2008). The impact of Megan's Law on sex offender recidivism: the Minnesota experience. *Criminology, 46*(2), 411-446.
Freeman, 2009	The author examines the influence of New York's registration/notification law on convicted sex offenders. The author compares outcomes of registered sex offenders with offenders who released into the community during the same time period, but were not subject to community registration/notification because their offenses were committed prior to the enactment of the law. The author uses survival analysis over a five-year follow-up period.	The author find that the registration/notification group had higher recidivism rates and were arrested faster than the comparison group.	Freeman, N.J. (2009, May 18). The public safety impact of community notification laws: Rearrest of convicted sex offenders. *Crime & Delinquency OnlineFirst,* published as doi: 10.1177/0011128708330852.

Letourneau & Armstrong, 2008	The authors examine the effects of South Carolina's registration/notification law on convicted juvenile offenders. The authors compare youth required to register under the law to a matched comparison group who did not have to register because judges used discretion not to impose the law. Although this is a selection bias, the study groups are matched on key variables and there are no significant differences between the groups. The authors used survival analysis over a 4.3 year follow-up period.	The authors find no significant differences between the groups for nonsexual person convictions. Registered youth had significantly higher nonperson convictions.	Letourneau, E.J., & Armstrong, K.S. (2008). Recidivism rates for registered and nonregistered juvenile sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 20 (4), 393-408.
Letourneau, Bandyopadhyay, Sinha & Armstrong, 2008	The authors examine influence of South Carolina registration/notification law on juvenile recidivism. The authors compare registered offenders to other youth prior to the implementation of the law using survival analysis over a nine-year follow-up period.	No statistically significant differences between registered and non-registered offenders on sexual or other adjudications.	Letourneau, E. J., Bandyopadhyay, D., Sinha, D., & Armstrong, K. S. (2008). The influence of sex offender registration on juvenile sexual recidivism [Online]. *Criminal Justice Policy Review*.
Prescott & Rockoff, 2008	The authors examine the impacts of registration/notification on general and specific deterrence using official crime data in fifteen states (CO, CT, ID, IA, KY, MA, MI, NE, ND, OH, SC, TX, UT, VT, and VA). The authors are the only authors to date who analyze the effects of registration and notification separately. The authors also examine the influence of registry size.	The authors find opposing effects. Registration, while taking into account the size of the registry, has a statistically significant negative relationship on sex offenses. Notification, while taking into account the size of the registry, has a statistically significant positive relationship on sex offenses. That is, as the size of the registry increases, sex offenses increase. The net effect is a decrease in sex offenses.	Prescott, JJ., & Rockoff, J. E. (2008, February). *Do sex offender registration and notification laws affect criminal behavior?* Retrievable from http://ssrn.com/abstract=1100663.
Schram & Milloy, 2005	The authors examine the influence of Washington's registration/notification law on arrest recidivism by conducting an outcome evaluation on the highest risk sex offenders (level III) subject to community notification. The treatment group was matched to a comparison group of offenders convicted prior to the passage of the law (the two variables include multiple sex convictions and victim type). The authors use survival analysis over a 4.5 year period.	The authors find no statistically significant differences between the study groups on sex offense arrests or all arrests. Offenders subject to registration were arrested for new crimes more quickly than the comparison group.	Schram, D. D., & Milloy, C. (1995). Community notification: A study of offender characteristics and recidivism. Olympia: Washington State Institute for Public Policy, Document No. 95-10-1101.
Shao & Li, 2006	The authors use econometric methods to examine the impacts of registration/notification on general and specific deterrence of rapes using official crime data in 50 states over a 32-year period and control for differences across states and time. The authors also test the effectiveness of registration/notification over time.	The authors find marginally significant reduction in reported rapes to the police. The authors also find that the effect gets larger as time goes on (i.e., registry size increases).	Shao, L. & Li, J. (2006). *The effect of sex offender registration laws on rape victimization.* Unpublished manuscript, University of Alabama.

Exhibit C
Citations and Summary of Studies Reviewed But Not Included in the Analysis Due to Methodological Rigor

Author and Year of Publication	Study Description and Methods	Reason for Exclusion	Full Citation
Adkins, Huff, & Stageberg, et al., 2000	The author examines the influence of Iowa's registration/notification law by conducting an outcome evaluation on sex offenders who released from prison before and after the passage of the law.	The registration/notification group is inherently less risky than the comparison group (as shown by criminal history and risk assessment data) and the authors do not use multivariate analysis to control for these differences. Results of multivariate analysis could be included in our analysis if conducted.	Adkins, G., Huff, D., Stageberg, P., Prell, L., & Musel, S. (2000, December). *The Iowa sex offender registry and recidivism*. Des Moines, IA: Iowa Department of Human Rights, Division of Criminal and Juvenile Justice Planning, Statistical Analysis Center.
Agan, 2007	The author investigates the impact of registration/notification on rape and other sexual offense rates using state-level UCR panel data.	The author does not include state-specific linear trends, which Shao & Li, 2006 demonstrate as a necessary variable to control for linear differences across states. Additionally, in order to code this study, the Institute received information from the author via email (mean reported rapes and standard deviation), which appeared low compared with Shao & Li's figures and also the Institute's analysis of UCR data.	Agan, A. (2007). *Sex offender registries: Fear without function?* Unpublished manuscript, University of Chicago.
Petrosino & Petrosino, 1999	The authors conduct a retrospective analysis to determine who would have been eligible for registration/notification had it existed at the time and how many crimes could have potentially been deterred.	This study is not an evaluation with a comparison group.	Petrosino, A. J., & Petrosino, C. (1999). The public safety potential of Megan's Law in Massachusetts: An assessment from a sample of criminal sexual psychopaths. *Crime & Delinquency,* 45, (1), 140-158.
Sandler, Freeman, & Socia, 2008.	The authors examine the impact of registration/notification conducting a time-series analysis using monthly arrest count data.	Time-series analysis is a statistical method used to determine a "break point" in time. In this case, the authors did not use multivariate controls, which help to isolate the cause of the break.	Sandler, J. C., Freeman, N. J., & Socia, K. M. (2008). Does a watched pot boil? A time-series analysis of New York State's Sex Offender Registration Notification law. *Psychology, Public Policy, and Law*, 14, 284-302.
Sienkiewicz, 2007	The authors conduct a retrospective analysis to determine who would have been eligible for registration/notification had it existed at the time and how many crimes could have potentially been deterred.	This study is not an evaluation with a comparison group.	Sienkiewicz, D. (2007). *Connecticut sex offender registry: The potential impact of a proactive community notification requirement* (Master's thesis, Central Connecticut State University).
Yessine & Bonta, 2006	The authors analyze the effectiveness of the Canadian National Flagging System to properly identify violent and sex offenders using actuarial risk scores. The analysis is not an outcome evaluation and does not have an appropriate comparison group to determine effectiveness of the registration law.	No comparison group.	Yessine, A. K., & Bonta, J. (2006). Tracking High-Risk, Violent Offenders: An Examination of the National Flagging. *Canadian Journal of Criminology and Criminal Justice*, 48 (4), 573 - 607.

Vásquez, Maddan, & Walker, 2008	The authors conduct a time-series analysis to determine the general deterrent effect of registration/ notification for 10 states.	Time-series analysis is a statistical method used to determine a "break point" in time. In this case, the authors did not use multivariate controls, which help to isolate the cause of the break.	Vásquez, B. E., Maddan, S., & Walker, J. T. (2008). The influence of sex offender registration and notification laws in the United States: A time-series analysis. *Crime & Delinquency, 54*(2), 175-192.
Zevitz, 2006	The author examines the impact of community notification on sex offenders in Wisconsin. The state recommends a notification level for a sex offender, but local jurisdictions implement at their discretion. The author compares offenders who received extensive notification to offenders who may or may not have been exposed to limited notification.	The author is unclear about what the study groups were exposed to. For example the comparison group may or may not have received limited notification. Further, the author conducts multivariate analysis, but does not report the results.	Zevitz, R. G. (2006). Sex offender community notification: Its role in recidivism and offender reintegration. *Criminal Justice Studies, 19*(2), 193-208.
Zgoba, Witt, Dalessandro, et al., 2008	The author investigates the impact of registration/notification in New Jersey by analyzing rearrest rates of registered sex offenders who released from prison compared with offenders not required to register prior to the law change. The authors conduct survival analysis of rearrest rates over a six-year period. The authors also conduct a time-series analysis using official crime rate data.	The registration group has statistically fewer prior sex offenses than the comparison group and the authors do not use multivariate analysis to control for differences between the groups. For the time-series analysis, the authors also do not use multivariate controls.	Zgoba, K., Witt, P., Dalessandro, M., & Veysey, B. (2008, December). *Megan's Law: Assessing the practical and monetary efficacy*. Trenton, NJ: New Jersey Department of Corrections, Office of Policy & Planning, The Research & Evaluation Unit.

For further information, contact Elizabeth Drake at
(360) 586-2767 or ekdrake@wsipp.wa.gov

Washington State
Institute for
Public Policy