United States General Accounting Office

GAO

Report to Congressional Requesters

March 2002

JUSTICE IMPACT EVALUATIONS

One Byrne Evaluation Was Rigorous; All Reviewed Violence Against Women Office Evaluations Were Problematic

G A O

Accountability * Integrity * Reliability

GAO-02-309

Contents

Letter 1

Results in Brief 2
Background 4
Scope and Methodology 7
Number, Type, Status of Completion, and Award Amount of Byrne and VAWO Discretionary Grant Program Evaluations 8
Methodological Problems Have Adversely Affected Three of Four Impact Evaluations 10
Conclusions 18
Recommendations for Executive Action 18
Agency Comments and Our Evaluation 19

Appendix I **Summaries of the Impact Evaluations of Byrne and VAWO Programs** 25

Appendix II **NIJ's Guidelines for Disseminating Discretionary Grant Program Evaluation Results** 29

NIJ's Dissemination Practices 29
Dissemination of NIJ's Byrne Discretionary Grants Comprehensive Communities Program and Children at Risk Program 31

Appendix III **Comments from the Department of Justice** 32

GAO Comments 43

Appendix IV **GAO Contacts and Staff Acknowledgments** 47

GAO Contacts 47
Acknowledgments 47

Tables

Table 1: NIJ Evaluations of Byrne Discretionary Grant Programs 9
Table 2: NIJ Evaluations of VAWO Discretionary Grant Programs 10

Abbreviations

BJA	Bureau of Justice Assistance
NIJ	National Institute of Justice
OJP	Office of Justice Programs
VAWO	Violence Against Women Office

G A O
Accountability * Integrity * Reliability

United States General Accounting Office
Washington, DC 20548

March 7, 2002

The Honorable Charles E. Grassley
Ranking Minority Member
Subcommittee on Crime and Drugs
The Honorable Jeff Sessions
Ranking Minority Member
Subcommittee on Administrative Oversight and the Courts
Committee on the Judiciary
United States Senate

This report responds to your request that we review selected aspects of the U.S. Department of Justice's (Justice) Office of Justice Programs' (OJP) program evaluations of discretionary grants awarded by OJP's Bureau of Justice Assistance's (BJA) Byrne Program (Byrne)[1] and the Violence Against Women Office (VAWO). Between fiscal years 1997 and 2000, Byrne and VAWO discretionary grant awards grew, in constant fiscal year 2000 dollars, about 85 percent—from about $105 million to approximately $194 million. These funds were awarded directly to various organizations, such as state and local governments, either on a competitive basis or pursuant to legislation allocating funds through congressional earmarks or direction. Discretionary grants awarded under the Byrne program were designed to help state and local governments make communities safe and improve criminal justice. Discretionary grants awarded under VAWO programs are aimed at improving criminal justice system responses to domestic violence, sexual assault, and stalking. Questions have been raised, however, regarding what these and other Justice grant programs have accomplished.

To address your request, we are reporting on (1) the number, type, status of completion, and award amount of Byrne and VAWO discretionary grant program evaluations during fiscal years 1995 through 2001 and (2) the methodological rigor of the impact evaluation studies of Byrne and VAWO discretionary grant programs during fiscal years 1995 through 2001. In addition, information that you requested on OJP's approaches to

[1]The Edward Byrne Memorial State and Local Law Enforcement Assistance Program. The Byrne discretionary grant program represents the largest single discretionary grant program within BJA.

disseminating evaluation results about Byrne and VAWO discretionary grant programs is presented in appendix II. Our review covered discretionary grant program evaluations managed by Justice's National Institute of Justice (NIJ). Program evaluations are systematic studies that are conducted periodically or ad hoc to assess how well a program is working. These studies can include impact evaluations—designed to assess the net effect of a program by comparing program outcomes with an estimate of what would have happened in the absence of the program—and process evaluations—designed to assess the extent to which a program is operating as intended. NIJ is OJP's principal research and development agency and is responsible for evaluating Byrne and VAWO programs. For Byrne program evaluations, NIJ awarded funding to evaluators using its own funds and funds made available through BJA. For VAWO program evaluations, NIJ awarded funding to evaluators using funds made available exclusively through VAWO.

Results in Brief

During fiscal years 1995 through 2001, NIJ awarded about $6 million for five Byrne and five VAWO discretionary grant program evaluations. Of the 10 program evaluations, all 5 VAWO evaluations were designed to be both process and impact evaluations of the VAWO programs. By contrast, only one of the five Byrne evaluations was designed as an impact evaluation and the four other Byrne evaluations were process evaluations.

Our in-depth review of the four impact evaluations that have progressed beyond the formative stage since fiscal year 1995 showed that only one of these, the evaluation of the Byrne Children at Risk (CAR) Program, was methodologically sound. The other three evaluations, all of which examined VAWO programs, had methodological problems that raise concerns about whether the evaluations will produce definitive results. Although program evaluation is an inherently difficult task, in all three VAWO evaluations, the effort is particularly arduous because of variations across grantee sites in how the programs are implemented. In addition, VAWO sites participating in the impact evaluations have not been shown to be representative of their programs, thereby limiting the evaluators' ability to generalize results, and the lack of comparison groups hinders evaluators' ability to minimize the effects of factors that are external to the program. Furthermore, data collection and analytical problems (e.g., related to statistical tests, assessment of change) compromise the evaluator's ability to draw appropriate conclusions from the results. Peer review committees found methodological problems in two of these three evaluation proposals that we examined, but it is unclear what NIJ has done to effectively resolve those problems.

We are making a recommendation regarding the two VAWO impact evaluations in the formative stage of development, and for all future impact evaluations, to ensure that potential methodological design and implementation problems are mitigated.

In commenting on a draft of this report, OJP's Assistant Attorney General stated that she agreed with the substance of our recommendations and said that NIJ has begun or plans to take steps to address them. Although it is still too early to determine whether NIJ's actions will be effective in preventing or resolving the problems we identified, they appear to be steps in the right direction. The Assistant Attorney General also said that the report could have gone further in acknowledging the challenges that evaluators face when conducting research in the complex environment of criminal justice programs and interventions. In addition, she stated that the report contrasts the Byrne evaluation with the three VAWO evaluations and obscures important programmatic differences that affect an evaluator's ability to achieve GAO's conditions for methodological rigor.

Our report applies standards of methodological rigor that are well defined in scientific literature. We recognize that there are substantive differences in the intent, structure, and design of the various discretionary grant programs managed by OJP and its bureaus and offices and that impact evaluation can be an inherently difficult and challenging task, especially since the Byrne and VAWO programs are operating in an ever-changing, complex environment. Not all evaluation issues that can compromise results are easily resolvable, but with more up-front attention to design and implementation issues, there is a greater likelihood that NIJ evaluations will provide meaningful results for policymakers. Absent this up-front attention, questions arise as to whether NIJ is (1) positioned to provide the definitive results expected from an impact evaluation and (2) making sound investments given the millions of dollars spent on these evaluations. The full text of the Assistant Attorney General's comments and our evaluation of them are presented in appendix III and elsewhere in this report, as appropriate.

Background

The Justice Assistance Act of 1984 (P.L. 98-473) created OJP to provide federal leadership in developing the nation's capacity to prevent and control crime, administer justice, and assist crime victims.[2] OJP carries out its responsibilities by providing grants to various organizations, including state and local governments, Indian tribal governments, nonprofit organizations, universities, and private foundations. OJP comprises five bureaus, including BJA, and seven program offices, including VAWO.[3] In fulfilling its mission, BJA provides grants for programs and for training and technical assistance to combat violent and drug-related crime and help improve the criminal justice system. VAWO administers grants to help prevent and stop violence against women, including domestic violence, sexual assault, and stalking. During fiscal years 1995 through 2001, BJA and VAWO awarded about $943 million to fund 700 Byrne and 1,264 VAWO discretionary grants.

One of BJA's major grant programs is the Byrne Program.[4] BJA administers the Byrne program, just as its counterpart, VAWO, administers its programs. Under the Byrne discretionary grants program, BJA provides federal financial assistance to grantees for educational and training programs for criminal justice personnel; for technical assistance to state and local units of government; and for projects that are replicable in more than one jurisdiction nationwide. During fiscal years 1995 through 2001, Byrne discretionary grant programs received appropriations of about $385 million.

[2]This act amended the Omnibus Crime Control and Safe Streets Act of 1968 to eliminate the Law Enforcement Assistance Administration and established OJP, which was to be headed by an assistant attorney general.

[3]OJP's other four bureaus are Bureau of Justice Statistics, NIJ, Office of Juvenile Justice and Delinquency Prevention, and Office for Victims of Crime. OJP's other six program offices are American Indian & Alaska Native Affairs Desk, Executive Office for Weed and Seed, the Corrections Program Office, the Drug Courts Program Office, the Office for Domestic Preparedness (formerly the Office for State and Local Domestic Preparedness Support), and the Office of the Police Corps and Law Enforcement Education.

[4]Local law enforcement programs established by the Anti-Drug Abuse Act of 1986 (P.L. 99-570) were amended and renamed the Edward Byrne Memorial State and Local Law Enforcement Assistance Program by the Anti-Drug Abuse Act of 1988 (P.L. 100-690). The Byrne grant program is codified at 42 U.S.C. 3750-3766b. The Byrne program is 1 of approximately 20 programs administered by BJA. The program funds formula grants, which are awarded directly to state governments, and discretionary grants for 28 legislatively authorized purposes.

VAWO was created in 1995 to carry out certain programs created under the Violence Against Women Act of 1994.[5] The Victims of Trafficking and Violence Prevention Act of 2000 reauthorized most of the existing VAWO programs and added new programs.[6] VAWO programs seek to improve criminal justice system responses to domestic violence, sexual assault, and stalking by providing support for law enforcement, prosecution, courts, and victim advocacy programs across the country. During fiscal years 1995 through 2001, VAWO's five discretionary grant programs that were subject to program evaluation were (1) STOP (Services, Training, Officers, and Prosecutors) Violence Against Indian Women Discretionary Grants, (2) Grants to Encourage Arrest Policies, (3) Rural Domestic Violence and Child Victimization Enforcement Grants, (4) Domestic Violence Victims' Civil Legal Assistance Grants, and (5) Grants to Combat Violent Crimes Against Women on Campuses. During fiscal years 1995 through 2001, about $505 million was appropriated to these discretionary grant programs.[7]

As already mentioned, NIJ is the principal research and development agency within OJP, and its duties include developing, conducting, directing, and supervising Byrne and VAWO discretionary grant program evaluations. Under 42 U.S.C. 3766, NIJ is required to "conduct a reasonable number of comprehensive evaluations" of the Byrne discretionary grant program. In selecting programs for review under section 3766, NIJ is to consider new and innovative approaches, program costs, potential for replication in other areas, and the extent of public awareness and community involvement. According to NIJ officials, the implementation of various types of evaluations, including process and impact evaluations, fulfills this legislative requirement. Although legislation creating VAWO does not require evaluations of the VAWO discretionary grant programs, Justice's annual appropriations for VAWO during fiscal years 1998 through 2002 included monies for NIJ research and evaluations of violence against

[5]Title IV of the Violent Crime Control and Enforcement Act of 1994 (P.L. 103-322).

[6]P.L. 106-386.

[7]This amount does not include about $19 million for the Sex Offender Management Training discretionary program. During the period of our review, NIJ did not evaluate this training program. From fiscal year 1995 through fiscal year 1999 this program was administered by VAWO. As of fiscal year 2000, responsibility for the program was shifted to OJP's Corrections Program Office.

women.[8] In addition, Justice has promulgated regulations requiring that NIJ conduct national evaluations of two of VAWO's discretionary grant programs.[9] As with the Byrne discretionary programs, NIJ is not required by statute or Justice regulation to conduct specific types of program evaluations, such as impact or process evaluations.

The Director of NIJ is responsible for making the final decision on which Byrne and VAWO discretionary grant programs to evaluate; this decision is based on the work of NIJ staff in coordination with Byrne or VAWO program officials. Once the decision has been made to evaluate a particular program, NIJ issues a solicitation for proposals for grant funding from potential evaluators. When applications or proposals are received, an external peer review panel comprising members of the research and relevant practitioner communities is convened. Peer review panels identify the strengths, weaknesses, and potential methodologies to be derived from competing proposals. When developing their consensus reviews, peer review panels are to consider the quality and technical merit of the proposal; the likelihood that grant objectives will be met; the capabilities, demonstrated productivity, and experience of the evaluators; and budget constraints. Each written consensus review is reviewed and discussed with partnership agency representatives (e.g., staff from BJA or VAWO). These internal staff reviews and discussions are led by NIJ's Director of the Office of Research and Evaluation who then presents the peer review consensus reviews, along with agency and partner agency input, to the NIJ Director for consideration and final grant award decisions. The NIJ Director makes the final decision regarding which application to fund.

[8]In fiscal year 1998, Congress appropriated $7 million, and in subsequent fiscal years 1999 through 2002, Congress appropriated $5.2 million annually of VAWO funds for NIJ research and evaluations of violence against women. According to NIJ's Director, Office of Research and Evaluation, none of these funds were used to award evaluation grants of the five VAWO discretionary grant programs discussed in this report.

[9]28 C.F.R. 90.58 and 90.65.

Scope and Methodology

To meet our objectives, we conducted our work at OJP, BJA, VAWO, and NIJ headquarters in Washington, D.C. We reviewed applicable laws and regulations, guidelines, reports, and testimony associated with Byrne and VAWO discretionary grant programs and evaluation activities. In addition, we interviewed responsible OJP, NIJ, BJA, and VAWO officials regarding program evaluations of discretionary grants. As agreed with your offices, we focused on program evaluation activities associated with the Byrne and VAWO discretionary grant programs. In particular, we focused on the program evaluations of discretionary grants that were funded during fiscal years 1995 through 2001.

To address our first objective, regarding the number, type, status of completion, and award amount of Byrne and VAWO discretionary grant program evaluations, we interviewed NIJ, BJA, and VAWO officials and obtained information on Byrne and VAWO discretionary grant programs and program evaluations. Because NIJ is responsible for carrying out program evaluations of Byrne and VAWO discretionary grant programs, we also obtained and analyzed NIJ data about specific Byrne and VAWO discretionary grant program evaluations, including information on the number of evaluations as well as the type, cost, source of funding, and stages of implementation of each evaluation for fiscal years 1995 through 2001. We did not independently verify the accuracy or completeness of the data that NIJ provided.

To address the second objective, regarding the methodological rigor of the impact evaluation studies of Byrne and VAWO discretionary grant programs during fiscal years 1995 through 2001, we initially identified the impact evaluations from the universe of program evaluations specified by NIJ. We excluded from our analysis any impact evaluations that were in the formative stage of development—that is, the application had been awarded but the methodological design was not yet fully developed. As a result, we reviewed four program evaluations.

For the four impact evaluations that we reviewed, we asked NIJ to provide any documentation relevant to the design and implementation of the impact evaluation methodologies, such as the application solicitation, the grantee's initial and supplemental applications, progress notes, interim reports, requested methodological changes, and any final reports that may have become available during the data collection period. We also provided NIJ with a list of methodological issues to be considered in our review and requested them to submit any additional documentation that addressed these issues. We used a data collection instrument to obtain information systematically about each program being evaluated and about the features

of the evaluation methodology. We based our data collection and assessments on generally accepted social science standards. We examined such factors as whether evaluation data were collected before and after program implementation; how program effects were isolated (i.e., the use of nonprogram participant comparison groups or statistical controls); and the appropriateness of sampling and outcome measures. Two of our senior social scientists with training and experience in evaluation research and methodology separately reviewed the evaluation documents and developed their own assessments before meeting jointly to discuss the findings and implications. This was done to promote a grant evaluation review process that was both independent and objective.

To obtain information on the approaches that BJA, VAWO, and NIJ used to disseminate program evaluation results, we requested and reviewed, if available, relevant handbooks and guidelines on information dissemination, including, for example, NIJ's guidelines. We also reviewed BJA, VAWO, and NIJ's available print and electronic products as related to their proven programs and evaluations, including two NIJ publications about Byrne discretionary programs and their evaluation methodologies and results.

We conducted our work between February 2001 and December 2001 in accordance with generally accepted government auditing standards.

We requested comments from Justice on a draft of this report in January 2002. The comments are discussed near the end of this letter and are reprinted as appendix III.

Number, Type, Status of Completion, and Award Amount of Byrne and VAWO Discretionary Grant Program Evaluations

During fiscal years 1995 through 2001, NIJ awarded about $6 million to carry out five Byrne and five VAWO discretionary grant program evaluations. NIJ awarded evaluation grants using mostly funds transferred from BJA and VAWO. Specifically, of the approximately $1.9 million awarded for one impact and four process evaluations of the Byrne discretionary program, NIJ contributed about $299,000 (16 percent) and BJA contributed about $1.6 million (84 percent). VAWO provided all of the funding (about $4 million) to NIJ for all program evaluations of five VAWO discretionary grant programs. According to NIJ, the five VAWO program evaluations included both impact and process evaluations. Our review of information provided by NIJ showed that 6 of the 10 program evaluations—all 5 VAWO evaluations and 1 Byrne evaluation—included impact evaluations. The remaining four Byrne evaluations were exclusively process evaluations that measured the extent to which the

programs were working as intended. [10] As of December 2001, only one of these evaluations, the impact evaluation of the Byrne CAR Program, had been completed. The remaining evaluations were in various stages of implementation. Table 1 lists each of the five Byrne program evaluations and shows whether it was a process or an impact evaluation, its stage of implementation, the amount awarded during fiscal years 1995 through 2001, and the total amount awarded since the evaluation was funded.

Table 1: NIJ Evaluations of Byrne Discretionary Grant Programs

Discretionary grant programs	Impact evaluation	Process evaluation	Status/stage of completion as of 12/01	Amount awarded for evaluation since FY 1995	Total amount awarded[a]
Children at Risk Program	Y	N	Completed	$452,780	**$1,034,732**
Comprehensive Communities Program[b]	N	Y	Final report under peer review	482,762	**782,294**
Tribal Strategies Against Violence Initiative	N	Y	Final report under peer review	280,169	**280,169**
Boston's Safe Neighborhood Initiative	N	Y	Evaluation under way	274,223	**274,223**
Red Hook Community Court Program	N	Y	Evaluation under way	374,981	**374,981**
Total evaluation dollars awarded				**$1,864,915**	**$2,746,399**

[a]Two of the Byrne program evaluation grants began prior to FY 1995. The evaluation of the Children at Risk Program was initially awarded in FY 1992 for $581,952. The evaluation of the Comprehensive Communities Program was initially awarded in FY 1994 for $299,532.

[b]NIJ documents identify the Byrne Comprehensive Communities Program evaluation as a process evaluation, but NIJ officials indicated that some elements of the evaluation were impact-oriented. Our review of the evaluation subsequently showed that this was a process evaluation because it addressed whether the program was working as intended rather than identifying a net effect of the program.

Source: GAO analysis of data provided by the National Institute of Justice

Table 2 lists each of the five VAWO program evaluations and shows that it was both a process and an impact evaluation, its stage of implementation, and the amount awarded during fiscal years 1995 through 2001, which is the total amount awarded.

[10]NIJ documents identify the Byrne Comprehensive Communities Program evaluation as a process evaluation, but NIJ officials indicated that some elements of the evaluation were impact oriented. Our review of the evaluation subsequently showed that this was a process evaluation, because it addressed whether the program was working as intended rather than identifying the net effect of the program.

Table 2: NIJ Evaluations of VAWO Discretionary Grant Programs

Discretionary grant programs	Impact evaluation	Process evaluation	Status/stage of completion as of 12/01	Total amount awarded for evaluation
STOP Violence Against Indian Women Discretionary Grants	Y	Y	Evaluation under way	**$468,552**
Grants to Encourage Arrests Policies	Y	Y	Evaluation under way	**1,130,574**
Rural Domestic Violence and Child Victimization Enforcement Grants	Y	Y	Evaluation under way	**719,949**
Domestic Violence Victims' Civil Legal Assistance Grants	Y	Y	Formative stage	**800,154**
Grants to Combat Crimes Against Women on Campuses	Y	Y	Formative stage	**849,833**
Total Evaluation Award Amounts				**$3,969,032**

Note: We did not review the program evaluation methodologies of the Civil Legal Assistance Program and Crimes Against Women on Campus Program because they were in their formative stage of development. That is, the application was awarded but the methodological design was not yet fully developed.

Source: GAO analysis of data provided by the National Institute of Justice.

Methodological Problems Have Adversely Affected Three of Four Impact Evaluations

Our review showed that methodological problems have adversely affected three of the four impact evaluations that have progressed beyond the formative stage. All three VAWO evaluations that we reviewed demonstrated a variety of methodological limitations, raising concerns as to whether the evaluations will produce definitive results. The one Byrne evaluation was well designed and used appropriate data collection and analytic methods. We recognize that impact evaluations, such as the type that NIJ is managing, can encounter difficult design and implementation issues. In the three VAWO evaluations that we reviewed, program variation across sites has added to the complexity of designing the evaluations. Sites could not be shown to be representative of the programs or of particular elements of these programs, thereby limiting the ability to generalize results; the lack of comparison groups hinders the ability to minimize the effects of factors external to the program. Furthermore, data collection and analytical problems compromise the ability of evaluators to draw appropriate conclusions from the results. In addition, peer review committees found methodological problems in two of the three VAWO evaluations that we considered.

The four program evaluations are multiyear, multisite impact evaluations. Some program evaluations used a sample of grants, while others used the entire universe of grants. For example, the Grants to Encourage Arrests Policies Program used 6 of the original 130 grantee sites. In contrast, in the

Byrne Children at Risk impact evaluation, all five sites participated. As of December 2001, NIJ had already received the impact findings from the Byrne Children at Risk Program evaluation but had not received impact findings from the VAWO discretionary grant program evaluations.

Impact Evaluations Are Difficult to Successfully Design and Implement

An impact evaluation is an inherently difficult task, since the objective is to isolate the effects of a particular program or factor from all other potential contributing programs or factors that could also effect change. Given that the Byrne and VAWO programs are operating in an ever changing, complex environment, measuring the impact of these specific Byrne and VAWO programs can be arduous. For example, in the evaluation of VAWO's Rural Domestic Violence Program, the evaluator's responsibility is to demonstrate how the program affected the lives of domestic violence victims and the criminal justice system. Several other programs or factors besides the Rural Domestic Violence Program may be accounting for all or part of the observed changes in victims' lives and the criminal justice system (e.g., a co-occurring program with similar objectives, new legislation, a local economic downturn, an alcohol abuse treatment program). Distinguishing the effects of the Rural Domestic Violence Program requires use of a rigorous methodological design.

Project Variation within the VAWO Programs Complicates Evaluation Design and Implementation

All three VAWO programs permitted their grantees broad flexibility in the development of their projects to match the needs of their local communities. According to the Assistant Attorney General, this variation in projects is consistent with the intent of the programs' authorizing legislation. We recognize that the authorizing legislation provides VAWO the flexibility in designing these programs. Although this flexibility may make sense from a program perspective, the resulting project variation makes it more difficult to design and implement a definitive impact evaluation of the program. Instead of assessing a single, homogeneous program with multiple grantees, the evaluation must assess multiple configurations of a program, thereby making it difficult to generalize about the entire program. Although all of the grantees' projects under each program being evaluated are intended to achieve the same or similar goals, an aggregate analysis could mask the differences in effectiveness among individual projects and thus not result in information about which configurations of projects work and which do not.

The three VAWO programs exemplify this situation. The Arrest Policies Program provided grantees with the flexibility to develop their respective projects within six purpose areas: implementing mandatory arrest or

proarrest programs and policies in police departments, tracking domestic violence cases, centralizing and coordinating police domestic violence operations, coordinating computer tracking systems, strengthening legal advocacy services, and educating judges and others about how to handle domestic violence cases. Likewise, the STOP Grants Program encouraged tribal governments to develop and implement culture-specific strategies for responding to violent crimes against Indian women and provide appropriate services for those who are victims of domestic abuse, sexual assault, and stalking. Finally, the Rural Domestic Violence Program was designed to provide sites with the flexibility to develop projects, based on need, with respect to the early identification of, intervention in, and prevention of woman battering and child victimization; with respect to increases in victim safety and access to services; with respect to enhancement of the investigation and prosecution of crimes of domestic violence; and with respect to the development of innovative, comprehensive strategies for fostering community awareness and prevention of domestic abuse. Because participating grant sites emphasized different project configurations, the resulting evaluation may not provide information that could be generalized to a broader implementation of the program.

Site Selection Not Shown to be Representative within the VAWO Programs

The sites participating in the three VAWO evaluations were not shown to be representative of their programs. Various techniques are available to help evaluators choose representative sites and representative participants within those sites. Random sampling of site and participant selection are ideal, but when this is not feasible, other purposeful sampling methods can be used to help approximate the selection of an appropriate sample (e.g., choosing the sample in such proportions that it reflects the larger population--stratification). At a minimum, purposeful selection can ensure the inclusion of a range of relevant sites.

As discussed earlier, in the case of the Arrest Policies Program, six purpose areas were identified in the grant solicitation. The six grantees chosen for participation in the evaluation were not however, selected on the basis of their representativeness of the six purpose areas or the program as a whole. Rather, they were selected on the basis of factors related solely to program “stability;” that is; they were considered likely to receive local funding after the conclusion of federal grant funding, and key personnel would continue to participate in the coordinated program effort. Similarly, the 10 Rural Domestic Violence impact evaluation grantees were not selected for participation on the basis of program representativeness or the specific purpose areas discussed earlier. Rather, sites were selected

by the grant evaluator on the basis of "feasibility"; specifically, whether the site would be among those participants equipped to conduct an evaluation.[11] Similarly, the STOP Violence Against Indian Women Program evaluation used 3 of the original 14 project sites for a longitudinal study; these were not shown to be representative of the sites in the overall program. For another phase of the evaluation, the principal investigator indicated that grantee sites were selected to be geographically representative of American Indian communities. While this methodology provides for inclusion of a diversity of Indian tribes in the sample from across the country, geography as a sole criterion does not guarantee representativeness in relation to many other factors.

Lack of Comparison Groups in the VAWO Evaluations Hinders Ability to Isolate the Effects of the Programs

Each of the three VAWO evaluations was designed without comparison groups—a factor that hinders the evaluator's ability to isolate and minimize external factors that could influence the results of the study. Use of comparison groups is a standard practice employed by evaluators to help determine whether differences between baseline and follow-up results are due to the program under consideration or to some other programs or external factors. For example, as we reported in 1997,[12] to determine whether a drug court program has been effective in reducing criminal recidivism and drug relapse, it is not sufficient to merely determine whether those participating in the drug court program show changes in recidivism and relapse rates. Changes in recidivism and relapse variables between baseline and program completion could be due to other external factors, irrespective of the drug court program (e.g., the state may have developed harsher sentencing procedures for those failing to meet drug court objectives). If, however, the drug court participant group is matched at baseline against another set of individuals, "the comparison group" who are experiencing similar life circumstances but who do not qualify for drug court participation (e.g., because of area of residence), then the comparison group can help in isolating the effects of the drug court program. The contrasting of the two groups in relation to recidivism and relapse can provide an approximate measure of the program's impact.

[11]According to NIJ officials, sites would also be selected based on VAWO's recommendations.

[12]U.S. General Accounting Office, *Drug Courts: Overview of Growth, Characteristics, and Results*, GAO/GGD-97-106, (Washington, D.C.: July 1997).

All three VAWO program impact evaluations lacked comparison groups. One issue addressed in the Arrest Policies Program evaluation, for example, was the impact of the program on the safety and protection of the domestic violence victim. The absence of a comparison group, however, makes it difficult to firmly conclude that change in the safety and protection of participating domestic abuse victims is due to the Arrest Policies Program and not to some other external factors operating in the environment (e.g., economic changes, nonfederal programs such as safe houses for domestically abused women, and church-run support programs). Instead of using comparison groups, the Arrest Policies Program evaluation sought to eliminate potential competing external factors by collecting and analyzing extensive historical and interview data from subjects and by conducting cross-site comparisons; the latter method proved unfeasible.

The STOP Violence Against Indian Women Discretionary Grant Program has sought in part, to reduce violent crimes against Indian women by changing professional staff attitudes and behaviors. To do this, some grantees created and developed domestic violence training services for professional staff participating in site activities. Without comparison groups, however, assessing the effect of the STOP training programs is difficult. Attitudes and behaviors may change for myriad reasons unrelated to professional training development initiatives. If a treatment group of professional staff receiving the STOP training had been matched with a comparison group of professional staff that was similar in all ways except receipt of training, there would be greater confidence that positive change could be attributed to the STOP Program. Similarly, the lack of comparison groups in the Rural Domestic Violence evaluation makes it difficult to conclude that a reduction in violence against women and children in rural areas can be attributed entirely, or in part, to the Rural Domestic program. Other external factors may be operating.

VAWO Data Collection and Analytical Problems Evident during Grant Implementation

All three VAWO impact evaluations involved data collection and analytical problems that may affect the validity of the findings and conclusions. For example, we received documentation from NIJ on the STOP Grant Program for Reducing Violence Against Indian Women showing that only 43 percent of 127 grantees returned a mail survey.[13] In addition, only 25 percent of 127 tribes provided victim outcome homicide and hospitalization rates—far less than the percentage needed to draw broad-based conclusions about the intended goal of assessing victim well being. In the Arrest Policies evaluation, NIJ reported that the evaluators experienced difficulty in collecting pre-grant baseline data from multiple sites and the quality of the data was oftentimes inadequate, which hindered their ability to statistically analyze change over time. In addition, evaluators were hindered in several work areas by lack of automated data systems; data were missing, lost, or unavailable; and the ability to conduct detailed analyses of the outcome data was sometimes limited. For the Rural Domestic Violence evaluation, evaluators proposed using some variables (e.g., number and type of awareness messages disseminated to the community each month, identification of barriers to meeting the needs of women and children, and number of police officers who complete a training program on domestic violence) that are normally considered to relate more to a process evaluation than an impact evaluation. NIJ noted that outcome measurement indicators varied by site, complicating the ability to draw generalizations. NIJ further indicated that the evaluation team did not collect baseline data prior to the start of the program, making it difficult to identify change resulting from the program.

VAWO Peer Review Committees Expressed Concerns about Two Evaluations

NIJ does not require applicants to use particular evaluation methodologies.[14] NIJ employs peer review committees in deciding which evaluation proposals to fund. The peer review committees expressed concerns about two of the three VAWO program evaluation proposals (i.e.,

[13]Guidance from the Office of Management and Budget indicated that certain statistical errors begin to arise when response rates are in the 50- to 75-percent range. Below 50 percent, these errors rise rapidly, and they are not significant when the response rates are 75 percent or higher. Office of Management and Budget, Office of Information and Regulatory Affairs, *The Paperwork Reduction Act of 1995: Implementing Guidance for OMB Review of Agency Information Collection*. [Washington, D.C.: 1999], 127 and 128.

[14]NIJ's Director of Planning and Management indicated several reasons for the agency's decision not to specify particular evaluation methodologies, including the need to maintain the grantee's independence, objectivity, innovation, and methodological flexibility under varying situations.

those for the Arrest Policies and Rural Domestic Violence programs) that were subsequently funded by NIJ. Whereas NIJ funded the Arrest Policies evaluation as a grant, NIJ funded the Rural Domestic Violence evaluation as a cooperative agreement so that NIJ could provide substantial involvement in conducting the evaluation.[15]

A peer-review panel and NIJ raised several concerns about the Arrest Policies Program evaluation proposal. These concerns included issues related to site selection, victim interviewee selection and retention in the sample, and the need for additional impact measures and control variables. The grant applicant's responses to these issues did not remove concerns about the methodological rigor of the application, thus calling into question the ability of the grantee to assess the impact of the Arrest Policies Program. For example, the grantee stated that victim interviewee selection was to be conducted through a quota process and that the sampling would vary by site. This would not allow the evaluators to generalize program results. Also, the evaluators said that they would study communities at different levels of "coordination" when comparison groups were not feasible, but they did not adequately explain (1) how the various levels of coordination would be measured, (2) the procedures used to select the communities compared, and (3) the benefits of using this method as a replacement for comparison groups. NIJ subsequently funded this evaluation, and it is still in progress.

A peer review committee for the Rural Domestic Violence and Child Victimization Enforcement Grant Program evaluation also expressed concerns about whether the design of the evaluation application, as proposed, would demonstrate whether the program was working. In its consensus review notes, the peer review committee indicated that the "ability to make generalizations about what works and does not work will be limited." The peer review committee also warned of outside factors (e.g., unavailability of data, inaccessibility of domestic violence victims) that could imperil the evaluation efforts of the applicant. Based on the peer review committee's input, NIJ issued the following statement to the applicant: "As a national evaluation of a major programmatic effort we hope to have a research design and products on what is working, what is not working, and why. We are not sure that the proposed design will get us to that point." We reviewed the grant applicant's response to NIJ's concern

[15]According to OJP, a cooperative agreement includes substantial involvement by the agency.

in its application addendum and found that the overall methodological design was still not discussed in sufficient detail or depth to determine whether the program was working. Although the Deputy Director of NIJ's Office of Research and Evaluation asserted that this initial application was only for process evaluation funding, our review of available documents showed that the applicant had provided substantial information about both the process and impact evaluation methodologies in the application and addendum. We believe that the methodological rigor of the addendum was not substantially improved over that of the original application. The Deputy Director told us that, given the "daunting challenge faced by the evaluator," NIJ decided to award the grant as a cooperative agreement. Under this arrangement, NIJ was to have substantial involvement in helping the grantee conduct the program evaluation. The results of that evaluation have not yet been submitted. The evaluator's draft final report is expected no earlier than April 2002.

Byrne Evaluation Was Successfully Designed and Implemented

In contrast to the three VAWO impact evaluations, the Byrne impact evaluation employed methodological design and implementation procedures that met a high standard of methodological rigor, fulfilling each of the criteria indicated above. In part, this may reflect the fact that Byrne's CAR demonstration program, unlike the VAWO programs, was according to the Assistant Attorney General, intended to test a research hypothesis, and the evaluation was designed accordingly. CAR provided participants with the opportunity to use a limited number of program services (e.g., family services, education services, after-school activities) that were theoretically related to the impact variables and the prevention and reduction of drug use and delinquency. As a result, the evaluation was not complicated by project heterogeneity. All five grantees participated in the evaluation. High-risk youths within those projects were randomly selected from targeted neighborhood schools, providing student representation. Additionally, CAR evaluators chose a matched comparison group of youths with similar life circumstances (e.g., living in distressed neighborhoods and exposed to similar school and family risk factors) and without access to the CAR Program. Finally, no significant data collection implementation problems were associated with the CAR Program. The data were collected at multiple points in time from youths (at baseline, at completion of program, and at one year follow-up) and their caregivers (at baseline and at completion of program). Self-reported findings from youths were supplemented by the collection of more objective data from school, police, and court records on an annual basis, and rigorous test procedures were used to determine whether changes over time were statistically significant.

Additionally, CAR's impact evaluation used control groups, a methodologically rigorous technique not used in the three VAWO evaluations. To further eliminate the effects of external factors, youths in the targeted neighborhood schools were randomly assigned either to the group receiving the CAR Program or to a control group that did not participate in the program. Since the CAR Program group made significant gains over the same-school group and the matched comparison group not participating in the program, there was good reason to conclude that the CAR Program was having a beneficial effect on the targeted audience.

Appendix I provides summaries of the four evaluations.

Conclusions

Despite great interest in assessing results of OJP's discretionary grant programs, it can be extremely difficult to design and execute evaluations that will provide definitive information. Our in-depth review of one Byrne and three VAWO impact evaluations that have received funding since fiscal year 1995 has shown that, in some cases, the flexibility that can be beneficial to grantees in tailoring programs to meet their communities' needs has added to the complexities of designing impact evaluations that will result in valid findings. Furthermore, the lack of site representativeness, appropriate comparison groups, and problems in data collection and analysis may compromise the reliability and validity of some of these evaluations. We recognize that not all evaluation issues that can compromise results are easily resolvable, including issues involving comparison groups and data collection. To the extent that methodological design and implementation issues can be overcome, however, the validity of the evaluation results will be enhanced. NIJ spends millions of dollars annually to evaluate OJP grant programs. More up-front attention to the methodological rigor of these evaluations will increase the likelihood that they will produce meaningful results for policymakers. Unfortunately, the problematic evaluation grants that we reviewed are too far along to be radically changed. However, two of the VAWO evaluation grants are still in the formative stage; more NIJ attention to their methodologies now can better ensure useable results.

Recommendations for Executive Action

We recommend that the Attorney General instruct the Director of NIJ to assess the two VAWO impact evaluations that are in the formative stage to address any potential methodological design and implementation problems and, on the basis of that assessment, initiate any needed interventions to help ensure that the evaluations produce definitive results. We further recommend that the Attorney General instruct the

Director of NIJ to assess its evaluation process with the purpose of developing approaches to ensure that future impact evaluation studies are effectively designed and implemented so as to produce definitive results.

Agency Comments and Our Evaluation

We provided a copy of a draft of this report to the Attorney General for review and comment. In a February 13, 2002, letter, the Assistant Attorney General commented on the draft. Her comments are summarized below and presented in their entirety in appendix III.

The Assistant Attorney General agreed with the substance of our recommendations and said that NIJ has begun, or plans to take steps, to address them. Although it is still too early to tell whether NIJ's actions will be effective in preventing or resolving the problems we identified, they appear to be steps in the right direction. With regard to our first recommendation—that NIJ assess the two VAWO impact evaluations in the formative stage to address any potential design and implementation problems and initiate any needed intervention to help ensure definitive results—the Assistant Attorney General noted that NIJ has begun work to ensure that these projects will provide the most useful information possible. She said that for the Crimes Against Women on Campus Program evaluation, NIJ is considering whether it will be possible to conduct an impact evaluation and, if so, how it can enhance its methodological rigor with the resources available. For the Civil Legal Assistance Program evaluation, the Assistant Attorney General said that NIJ is working with the grantee to review site selection procedures for the second phase of the study to enhance the representativeness of sites. The Assistant Attorney General was silent about any additional steps that NIJ would take during the later stages of the Civil Legal Assistance Program process evaluation to ensure the methodological rigor of the impact phase of the study. However, it seems likely that as the process evaluation phase of the study continues, NIJ may be able to take advantage of additional opportunities to address any potential design and implementation problems.

With regard to our second recommendation—that NIJ assess its evaluation process to develop approaches to ensure that future evaluation studies are effectively designed and implemented to produce definitive results—the Assistant Attorney General stated that OJP has made program evaluation, including impact evaluations of federally funded programs, a high priority. The Assistant Attorney General said that NIJ has already launched an examination of NIJ's evaluation process. She also noted that, as part of its reorganization, OJP plans to measurably strengthen NIJ's capacity to manage impact evaluations with the goal of making them more useful for

Congress and others. She noted as an example that OJP and NIJ are building measurement requirements into grants at the outset, requiring potential grantees to collect baseline data and track the follow-up data through the life of the grant. We have not examined OJP's plans for reorganizing, nor do we have a basis for determining whether OJP's plans regarding NIJ would strengthen NIJ's capacity to manage evaluations. However, we believe that NIJ and its key stakeholders, such as Congress and the research community, would be well served if NIJ were to assess what additional actions it could take to strengthen its management of impact evaluations regardless of any reorganization plans.

In her letter, the Assistant Attorney General pointed out that the report accurately describes many of the challenges facing evaluators when conducting research in the complex environment of criminal justice programs and interventions. However, she stated that the report could have gone further in acknowledging these challenges. The Assistant Attorney General also stated that the report contrasts the Byrne evaluation with the three VAWO evaluations and obscures important programmatic differences that affect an evaluator's ability to achieve "GAO's conditions for methodological rigor." She pointed out that the Byrne CAR Program was intended to test a research hypothesis and that the evaluation was designed accordingly, i.e., the availability of baseline data were ensured; randomization of effects were stipulated as a precondition of participation; and outcome measures were determined in advance on the basis of the theories to be tested. She further stated that, in contrast, all of the VAWO programs were (1) highly flexible funding streams, in keeping with the intention of Congress, that resulted in substantial heterogeneity at the local level and (2) well into implementation before the evaluation started. The Assistant Attorney General went on to say that it is OJP's belief that evaluations under less than optimal conditions can provide valuable information about the likely impact of a program, even though the conditions for methodological strategies and overall rigor of the CAR evaluation were not available.

We recognize that there are substantive differences in the intent, structure, and design of the various discretionary grant programs managed by OJP and its bureaus and offices. And, as stated numerous times in our report, we acknowledge not only that impact evaluation can be an inherently difficult and challenging task but also that measuring the impact of these specific Byrne and VAWO programs can be arduous, given that they are operating in an ever changing, complex environment. We agree that not all evaluation issues that can compromise results are easily resolvable, but we firmly believe that, with more up-front attention to design and

implementation issues, there is a greater likelihood that NIJ evaluations will provide meaningful results for policymakers. Absent this up-front attention, questions arise as to whether NIJ is (1) positioned to provide the definitive results expected from an impact evaluation and (2) making sound investments given the millions of dollars spent on these evaluations.

The Assistant Attorney General also commented that although our report discussed "generally accepted social science standards," it did not specify the document that articulates these standards or describe our elements of rigor. As a result, the Assistant Attorney General said, OJP had to infer that six elements had to be met to achieve what "GAO believes" is necessary to "have a rigorous impact evaluation." Specifically, she said that she would infer that, for an impact evaluation to be rigorous would require (1) selection of homogenous programs, (2) random or stratified site sampling procedures (or selection of all sites), (3) use of comparison groups, (4) high response rates, (5) available and relevant automated data systems that will furnish complete and accurate data to evaluators in a timely manner, and (6) funding sufficient to accomplish all of the above. Furthermore, the Attorney General said that it is rare to encounter all of these conditions or be in a position to engineer all of these conditions simultaneously; and when all of these conditions are present, the evaluation would be rigorous. She also stated that it is possible to glean useful, if not conclusive, evidence of the impact of a program from an evaluation that does not rise to the standard recommended by GAO because of the unavoidable absence of "one or more elements."

We agree that our report did not specify particular documents that articulate generally accepted social science standards. However, the standards that we applied are well defined in scientific literature.[16] All assessments of the impact evaluations we reviewed were completed by social scientists with extensive experience in evaluation research. Throughout our report, we explain our rationale and the criteria we used in measuring the methodological rigor of NIJ's impact evaluations. Furthermore, our report does not suggest that a particular standard or set

[16]Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally & Company, 1963); Carol H. Weiss, *Evaluation Research: Methods for Assessing Program Effectiveness* (Englewood Cliffs: Prentice-Hall, Inc., 1972); Edward A. Suchman, *Evaluative Research: Principles and Practice in Public Service & Social Action Programs* (New York: Russell Safe Foundation, 1967); U.S. General Accounting Office, *Designing Evaluations*, GAO/PEMD-10.1.4 (Washington, D.C.: May 1991).

of standards is necessary to achieve rigor, nor does it suggest that other types of evaluations, such as comprehensive process evaluations, are any less useful in providing information on how a program is operating. In this context, it is important to point out that the scope of our work covered impact evaluations of Byrne and VAWO discretionary grant programs—those designed to assess the net effect of a program by comparing program outcomes with an estimate of what would have happened in the absence of the program.

We differ with the Assistant Attorney General with respect to the six elements cited as necessary elements for conducting an impact evaluation. Contrary to the Assistant Attorney General's assertion, our report did not state that a single homogeneous program is a necessary element for conducting a rigorous impact evaluation. Rather, we pointed out that heterogeneity or program variation is a challenge that adds to the complexity of designing an evaluation. In addition, contrary to her assertion, the report did not assert that random sampling or stratification was a necessary element for conducting a rigorous evaluation; instead it stated that when random sampling is not feasible, other purposeful sampling methods can be used. With regard to comparison groups, the Assistant Attorney General's letter asserted that GAO standards required using groups that do not receive program benefits as a basis of comparison with those that do receive such benefits. In fact, we believe that the validity of evaluation results can be enhanced through establishing and tracking comparison groups. If other ways exist to effectively isolate the impacts of a program, comparison groups may not be needed. However, we saw no evidence that other methods were effectively used in the VAWO impact evaluations we assessed.

The Assistant Attorney General also suggested that we used a 75 percent or greater response rate for evaluation surveys as a standard of rigor. In fact, we did not—we simply pointed out that NIJ documents showed a 43 percent response rate on one of the STOP Grant Program evaluation surveys. This is below OMB's threshold response rate level—the level below which OMB particularly believes nonresponse bias and statistical problems could affect surveys. Given OMB guidance, serious questions could be raised about program conclusions drawn from the results of a survey with a 43 percent response rate. In addition, the Assistant Attorney General suggested that, by GAO standards, she would have to require state, local, or tribal government officials to furnish complete and accurate data in a timely manner. In fact, our report only points out that NIJ reported that evaluators were hindered in carrying out evaluations because of the lack of automated data systems or because data were

missing, lost, or unavailable—again, challenges to achieving methodologically rigorous evaluations that could produce meaningful and definitive results.

Finally, the Assistant Attorney General's letter commented that one of the elements needed to meet "all of GAO's conditions" of methodological rigor is sufficient funding. She stated that more rigorous impact evaluations cost more than those that provide less scientific findings, and she said that OJP is examining the issue of how to finance effective impact evaluations. We did not assess whether funding is sufficient to conduct impact evaluations, but we recognize that designing effective and rigorous impact evaluations can be expensive—a condition that could affect the number of impact evaluations conducted. However, we continue to believe that with more up-front attention to the rigor of ongoing and future evaluations, NIJ can increase the likelihood of conducting impact evaluations that produce meaningful and definitive results.

In addition to the above comments, the Assistant Attorney General made a number of suggestions related to topics in this report. We have included the Assistant Attorney General's suggestions in the report, where appropriate. Also, the Assistant Attorney General provided other comments in response to which we did not make changes. See appendix III for a more detailed discussion of the Assistant Attorney General's comments.

We are sending copies of this report to the Chairman and the Ranking Minority Member of the Senate Judiciary Committee; to the Chairman and Ranking Minority Member of the House Judiciary Committee; to the Chairman and Ranking Minority Member of the Subcommittee on Crime, House Committee on the Judiciary; to the Chairman and the Ranking Minority Member of the House Committee on Education and the Workforce; to the Attorney General; to the OJP Assistant Attorney General; to the NIJ Director; to the BJA Director; to the VAWO Director; and to the Director, Office of Management and Budget. We will also make copies available to others on request.

If you or your staff have any questions about this report, please contact John F. Mortin or me at (202) 512-8777. Key contributors to this report are acknowledged in appendix IV.

Laurie E. Ekstrand
Director, Justice Issues

Appendix I: Summaries of the Impact Evaluations of Byrne and VAWO Programs

Evaluation	National Evaluation of the Rural Domestic Violence and Child Victimization Grant Program
Principal investigator	COSMOS Corporation
Program evaluated	The Violence Against Women Office's (VAWO) Rural Domestic Violence Program, begun in fiscal year 1996, has funded 92 grantees through September 2001. The primary purpose of the program is to enhance the safety of victims of domestic abuse, dating violence, and child abuse. The program supports projects that implement, expand, and establish cooperative efforts between law enforcement officers, prosecutors, victim advocacy groups, and others in investigating and prosecuting incidents of domestic violence, dating violence, and child abuse; provide treatment, counseling, and assistance to victims; and work with the community to develop educational and prevention strategies directed toward these issues.
Evaluation components	The impact evaluation began in July 2000, with a final report expected no earlier than April 2002. Initially, 10 grantees were selected to participate in the impact evaluation; 9 remain in the evaluation. Two criteria were used in the selection of grant participants: the "feasibility" of earlier site-visited grantees to conduct an outcome evaluation and VAWO recommendations based on knowledge of grantee program activities. Logic models were developed, as part of the case study approach, to show the logical or plausible links between a grantee's activities and the desired outcomes. The specified outcome data were to be collected from multiple sources, using a variety of methodologies during 2- to-3-day site visits (e.g., multiyear criminal justice, medical, and shelter statistics were to be collected from archival records; community stakeholders were to be interviewed; and grantee and victim service agency staff were to participate in focus groups). At the time of our review, this evaluation was funded at $719,949. The National Institute of Justice (NIJ) could not separate the cost of the impact evaluation from the cost of the process evaluation.
Evaluation findings	Too early to assess.
Assessment of evaluation	This evaluation has several limitations. (1) The choice of the 10 impact sites is skewed toward the technically developed evaluation sites and is not representative of either all Rural Domestic Violence Program Grantees, particular types of projects, or delivery styles. (2) The lack of comparison groups will make it difficult to exclude the effect of external factors, such as victim safety and improved access to services, on perceived change. (3) Several so-called short-term outcome variables are in fact process variables (e.g., number of police officers who complete a training program on domestic violence, identification of barriers to meeting the needs of women and children). (4) It is not clear how interview and focus group participants are to be selected, (5) Statistical procedures to be used in the analyses have not been sufficiently identified. The NIJ peer review committee had concerns about whether the evaluation could demonstrate that the program was working. NIJ funded the application as a cooperative agreement because a substantial amount of agency involvement was deemed necessary to meet the objectives of the evaluation.

Evaluation	National Evaluation of the Arrest Policies Program
Principal investigator	Institute for Law and Justice (ILJ)
Program evaluated	The purpose of VAWO's Arrest Policies Program is to encourage states, local governments, and Indian tribal governments to treat domestic violence as a serious violation of criminal law. The program received a 3-year authorization (fiscal years 1996 through 1998) at approximately $120 million to fund grantees under six purpose areas: implementing mandatory arrest or proarrest programs and policies in police departments, tracking domestic violence cases, centralizing and coordinating police domestic violence operations, coordinating computer tracking systems, strengthening legal advocacy services, and educating judges and others about how to handle domestic violence cases. Grantees have flexibility to work in several of these areas. At the time the NIJ evaluation grant was awarded, 130 program grantees had been funded; the program has since expanded to 190 program grantees.
Evaluation components	The impact evaluation began in August 1998, with a draft final report due in March 2002. Six grantees were chosen to participate in the impact evaluation. Each of the six sites was selected on the basis of program "stability," not program representativeness. Within sites, both quantitative and qualitative data were to be collected and analyzed to enable better understanding of the impact of the Arrest Program on offender accountability and victim well being. This process entailed reviewing data on the criminal justice system's response to domestic violence; tracking a random sample of 100 offender cases, except in rural areas, to determine changes in offender accountability; conducting content analyses of police incident reports to assess change in police practices and documentation; and interviewing victims or survivors at each site to obtain their perceptions of the criminal justice system's response to domestic violence and its impact on their well-being. ILJ had planned cross-site comparisons and the collection of extensive historical and interview data to test whether competing factors could be responsible for changes in arrest statistics. At the time of our review, this evaluation was funded at $1,130,574. NIJ could not separate the cost of the impact evaluation from the cost of the process evaluation.
Evaluation findings	Too early to assess.
Assessment of evaluation	This evaluation has several limitations: the absence of a representative sampling frame for site selection, the lack of comparison groups, the inability to conduct cross-site comparisons, and the lack of a sufficient number of victims in some sites to provide a perspective on the changes taking place in domestic violence criminal justice response patterns and victim well-being. In addition, there was difficulty collecting pre-grant baseline data, and the quality of the data was oftentimes inadequate, limiting the ability to measure change over time. Further, automated data systems were not available in all work areas, and data were missing, lost, or unavailable. An NIJ peer review committee also expressed some concerns about the grantee's methodological design.

Evaluation	Impact Evaluation of STOP Grant Programs for Reducing Violence Against Indian Women
Principal investigator	The University of Arizona
Program evaluated	VAWO's STOP (Services, Training, Officers, and Prosecutors) Grant Programs for Reducing Violence Against Indian Women Discretionary Grant Program was established under Title IV of the Violent Crime Control and Law Enforcement Act of 1994. The program's principal purpose is to reduce violent crimes against Indian women. The program, which began in fiscal year 1995 with 14 grantees, encourages tribal governments to develop and implement culture-specific strategies for responding to violent crimes against Indian women and providing appropriate services for those who are victims of domestic abuse, sexual assault, and stalking. In this effort, the program provided funding for the services and training, and required the joint coordination, of nongovernmental service providers, law enforcement officers, and prosecutors hence the name, the STOP Grant Programs for Reducing Violence Against Indian Women.
Evaluation components	The University of Arizona evaluation began in October 1996 with an expected final report due in March 2002. The basic analytical framework of this impact evaluation involves the comparison of quantitative and qualitative pre-grant case study histories of participating tribal programs with changes taking place during the grant period. Various data collection methodologies have been adopted (at least in part, to be sensitive to the diverse Indian cultures): 30-minute telephone interviews, mail surveys, and face-to-face 2- to 3 day site visits. At the time of our review, this evaluation was funded at $468,552. NIJ could not separate the cost of the impact evaluation from the cost of the process evaluation.
Evaluation findings	Too early to assess.
Assessment of evaluation	Methodological design and implementation issues may cause difficulties in attributing program impact. A number of methodological aspects of the study remain unclear: the site selection process for "in-depth case study evaluations;" the methodological procedures for conducting the longitudinal evaluation; the measurement, validity, and reliability of the outcome variables; the procedures for assessing impact; and the statistical tests to be used for determining significant change. Comparison groups are not included in the methodological design. In addition, only 43 percent of the grantees returned the mail survey, only 25 percent could provide the required homicide and hospitalization rates; and only 26 victims of domestic violence and assault could be interviewed (generally too few to measure statistical change). Generalization of evaluation results to the entire STOP Grant Programs for Reducing Violence Against Indian Women will be difficult, given these problems.

Evaluation	Longitudinal Impact Evaluation of the Strategic Intervention for High Risk Youth (a.k.a. The Children at Risk Program)
Principal investigator	The Urban Institute
Program evaluated	The Children at Risk (CAR) Program, a comprehensive drug and delinquency prevention initiative funded by the Bureau of Justice Assistance (BJA), the Office of Juvenile Justice and Delinquency Prevention (OJJDP), the Center on Addiction and Substance Abuse, and four private foundations, was established to serve as an experimental demonstration program from 1992 to 1996 in five grantee cities. Low-income youths (11 to 13 years old) and their families, who lived in severely distressed neighborhoods at high-risk for drugs and crime, were targeted for intervention. Eight core service components were identified: case management, family services, education services, mentoring, after-school and summer activities, monetary and nonmonetary incentives, community policing, and criminal justice and juvenile intervention (through supervision and community service opportunities). The goals of the program were to reduce drug use among targeted families and improve the safety and overall quality of life in the community.
Evaluation components	The evaluation actually began in 1992, and the final report was submitted in May 1998. The study used both experimental and quasi-experimental evaluation designs. A total of 671 youths in target neighborhood schools were randomly assigned to either a treatment group (which received CAR services and the benefit of a safer neighborhood) or to a control group (which received only a safer neighborhood). Comparison groups (n=203 youths) were selected from similar high-risk neighborhoods by means of census tract data; comparison groups did not have access to the CAR Program. Interviews were conducted with youth participants at program entry (baseline), program completion (2 years later), and 1-year after program completion. A parent or caregiver was interviewed at program entry and completion. Records from schools, police, and courts were collected annually for each youth in the sample as a means of obtaining more objective data. The total evaluation funding was $1,034,732.
Evaluation findings	Youths participating in CAR were significantly less likely than youths in the control group to have used gateway and serious drugs, to have sold drugs, or to have committed violent crimes in the year after the program ended. CAR youths were more likely than youths in the control and comparison groups to report attending drug and alcohol abuse programs. CAR youths received more positive peer support than controls, associated less frequently with delinquent peers, and were pressured less often by peers to behave in antisocial ways. CAR households used more services than control group households, but the majority of CAR households did not indicate using most of the core services available.
Assessment of evaluation	CAR is a methodologically rigorous evaluation in both its design and implementation. The evaluation findings demonstrate the value of the program as a crime and drug prevention initiative.

Appendix II: NIJ's Guidelines for Disseminating Discretionary Grant Program Evaluation Results

NIJ has the primary role of disseminating Byrne and VAWO discretionary grant program evaluation results of evaluations managed by NIJ, according to NIJ, BJA, and VAWO officials, because NIJ is responsible for conducting these types of evaluations.[1] NIJ is authorized to share the results of its research with federal, state, and local governments.[2] NIJ also disseminates information on methodology designs.

NIJ's Dissemination Practices

NIJ's practices for disseminating program evaluation results are specified in its guidelines. According to the guidelines, once NIJ receives a final evaluation report from the evaluators and the results of peer reviews have been incorporated, NIJ grant managers are to carefully review the final product and, with their supervisor, recommend to the NIJ Director which program results to disseminate and the methods for dissemination. Before making a recommendation, grant managers and their supervisors are to consider various criteria, including policy implications, the nature of the findings and research methodology, the target audience and their needs, and the cost of various forms of dissemination. Upon receiving the recommendation, the Director of NIJ is to make final decisions about which program evaluation results to disseminate. NIJ's Director of Planning and Management said that NIJ disseminates program evaluation results that are peer reviewed, are deemed successful, and add value to the field.

Once the decision has been made to disseminate program evaluation results and methodologies with researchers and practitioners, NIJ can choose from a variety of publications, including its *Research in Brief; NIJ Journal–At a Glance: Recent Research Findings; Research Review; NIJ Journal–Feature Article;* and *Research Report*. In addition, NIJ provides research results on its Internet site and at conferences. For example, using its *Research in Brief* publication, NIJ disseminated impact evaluation results on the Byrne Children at Risk (CAR) program to 7,995 practitioners and researchers, including state and local government and law enforcement officials; social welfare and juvenile justice professionals; and criminal justice researchers. In addition, using the same format, NIJ stated that it distributed the results of its process evaluation of the Byrne Comprehensive Communities Program (CCP) to 41,374 various constituents, including local and state criminal and juvenile justice agency

[1]BJA and VAWO also disseminate evaluation results of their discretionary grant programs.

[2]42 U.S.C. 3721.

administrators, mayors and city managers, leaders of crime prevention organizations, and criminal justice researchers. NIJ and other OJP offices and bureaus also disseminated evaluation results during NIJ's annual conference on criminal justice research and evaluation. The July 2001 conference was attended by 847 public and nonpublic officials, including criminal justice researchers and evaluation specialists from academic institutions, associations, private organizations, and government agencies; federal, state, and local law enforcement, court, and corrections officials; and officials representing various social service, public housing, school, and community organizations.

In addition to NIJ's own dissemination activities, NIJ's Director of Planning and Management said that it allows and encourages its evaluation grantees to publish their results of NIJ-funded research via nongovernmental channels, such as in journals and through presentations at professional conferences. Although NIJ requires its grantees to provide advance notice if they are publishing their evaluation results, it does not have control over its grantees' ability to publish these results. NIJ does, however, require a Justice disclaimer that the "findings and conclusions reported are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice." For example, although NIJ has not yet disseminated the program evaluation results of the three ongoing VAWO impact evaluations that we reviewed, one of the evaluation grantees has already issued, on its own Internet site, 9 of 20 process evaluation reports on the Arrests Policies evaluation grant. The process evaluations were a component of the NIJ grantee's impact evaluation of the Arrest Policies Program. Because the evaluations were not completed, NIJ required that the grantee's publication of the process evaluations be identified as a draft report pending final NIJ review.

Dissemination of NIJ's Byrne Discretionary Grants: Comprehensive Communities Program and Children at Risk Program

As discussed earlier, NIJ publishes the results of its evaluations in several different publications. For example, NIJ used the *Research in Brief* format to disseminate evaluation results for two of the five Byrne discretionary grant programs Comprehensive Communities Program (CCP) and Children at Risk Program (CAR) that were evaluated during fiscal years 1995 through 2001. Both publications summarize information including each program's evaluation results, methodologies used to conduct the evaluations, information about the implementation of the programs themselves, and services that the programs provided.[3] CCP's evaluation results were based on a process evaluation. Although a process evaluation does not assess the results of the program being evaluated, it can provide useful information that explains the extent to which a program is operating as intended. The NIJ *Research in Brief* on the Byrne CAR Discretionary Grant Program provides a summary of issues and findings regarding the impact evaluation. That summary included findings reported one year after the end of the program, in addition to a summary of the methodology used to conduct the evaluation, the outcomes, the lessons learned, and a major finding from the evaluation.

[3]Although Justice funds these grants and publishes them, its publications carry a disclaimer that says "findings and conclusions of the research reported here are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice."

Appendix III: Comments from the Department of Justice

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

U.S. Department of Justice

Office of Justice Programs

Office of the Assistant Attorney General

Washington, D.C. 20531

FEB 13 2002

Laurie Ekstrand
Director, Justice Issues
Tax Administration and Justice
General Accounting Office
441 G Street, NW, Room 2A38
Washington, DC 20548

Dear Ms. Ekstrand:

This letter is in response to the General Accounting Office (GAO) draft report entitled, *"JUSTICE IMPACT EVALUATIONS: One Byrne Evaluation Was Rigorous; All Reviewed Violence Against Women Office Evaluations Were Problematic"(GAO-02-309).*

The report makes two specific recommendations to the National Institute of Justice (NIJ) regarding its ongoing evaluation work. These recommendations reflect shared values that NIJ and GAO place on monitoring closely all ongoing evaluations and strengthening evaluation designs whenever possible. The recommendations are re-stated in bold below, followed by our response.

> **We recommend the Attorney General instruct the Director of NIJ to assess the two VAWO impact evaluations that were in the formative stage to address any potential methodological design and implementation problems, and, on the basis of that assessment, initiate any needed interventions to help ensure that the evaluations produce definitive results.**

With regard to the two ongoing Violence Against Women Act (VAWA) evaluations that GAO characterizes as in formative stages, NIJ has already begun work to ensure that these projects will provide the most useful results possible. For the Crimes Against Women on Campus Program evaluation, NIJ is considering whether it will be possible to conduct an impact evaluation and if so, how to enhance its methodological rigor with the resources made available. For the Civil Legal Assistance Program evaluation – halfway through its process evaluation phase – NIJ will work with the grantee to review site selection procedures for the second phase of the study in order to enhance the representativeness of the sites.

We further recommend that the Attorney General instruct the Director of NIJ to assess its evaluation process with the purpose of developing approaches to ensure future impact evaluation studies are effectively designed and implemented so as to produce definitive results.

With respect to GAO's second recommendation, we have made program evaluation, including impact evaluations of Federally-funded programs, a high priority. Since assuming her position in August 2001, NIJ Director Hart has launched an examination of NIJ's evaluation process. As part of OJP's reorganization, Director Hart and I plan to measurably strengthen NIJ's capacity to manage impact evaluations so that they can be of maximum use to Congress, Federal, State, local, and tribal officials. For example, we are building measurement requirements into our grants at the outset, requiring potential grantees to collect baseline data and track the follow-up data through the life of the grant.

While we agree with the substance of the recommendations, we believe GAO's draft report could have gone further in acknowledging the challenges that evaluators face when conducting research in the complex environment of criminal justice programs and interventions. The report accurately describes many of these challenges, including extensive local "customization" of VAWA discretionary programs, the unwillingness of many victims to provide information, the unavailability of basic "baseline data,"[1] and the lack of meaningful comparison groups for some program sites such as specific rural and Indian Country communities.

The report contrasts the Byrne Children At Risk (CAR) evaluation with the three VAWA evaluations in a way that obscures important programmatic differences that affect an evaluator's ability to achieve GAO's conditions for methodological rigor. Specifically, in a discretionary demonstration project like CAR, the program is intentionally designed by a multi-agency team as a means to test a research hypothesis. Knowledge objectives take priority and program implementation is subservient to these objectives. As a result, the availability of baseline data is ensured; randomization of effects can be stipulated as a condition of participation; site selection and any local flexibility are dictated by the national evaluation design; and outcome measures are determined in advance based on the theoretical propositions to be tested.

In contrast, each VAWA "program" being evaluated, in keeping with the intentions of Congress, consists of a highly flexible funding stream. One result of this flexibility is substantial program heterogeneity at the local level.[2] Furthermore, VAWA program activities were already

[1]Baseline data represent conditions prior to program implementation.

[2] For example, the same VAWA program has been used in one community to pay for half the salary of one police officer in a pre-existing dedicated domestic violence unit, while in another community VAWA funds were used to create a comprehensive, coordinated community response to domestic violence that would support victim

2

well into implementation before the evaluation started. Thus, the conditions for methodological strategies and overall rigor of the CAR evaluation were simply not available. We believe, however, that an evaluation, even under less than optimal conditions, can still provide valuable information about the likely impact of such a program.

Now on p. 8.

In evaluating NIJ's evaluations, GAO applied what it said were "generally accepted social science standards"(p. 6). However, GAO did not specify the document that contains these standards or directly describe its elements of rigor. Based on GAO's draft report, we infer that GAO believes all of the following elements are required to have a rigorous impact evaluation:

Now on p. 11.

Now on p. 12.

Now on p. 13.

Now on p. 15.

Now on p. 15.

1. A single homogenous program even with multiple grantees -- in other words, a lack of local variation (heterogeneity) in program policies and implementation; (p. 10)
2. Evaluation sites selected using random sampling or stratification (or evaluating all sites); (p. 11)
3. Comparison groups employed to match the characteristics of those receiving the program, except that these groups do not receive the program benefits; (p. 12)
4. High response rates (>75 %) to all evaluation surveys; (p. 14)
5. Available and relevant automated data systems maintained by State, local, and/or tribal officials, who will furnish complete and accurate data (both past and current) to evaluators in a timely manner; (p. 14) and
6. Funding sufficient to accomplish all of the above.

It is rare to encounter or be in a position to engineer all of GAO's conditions (as listed above) simultaneously when evaluating Federally funded criminal justice "programs" or funding streams. We agree that if all of GAO's conditions are present, the evaluation would be rigorous. Without question, randomized trials have their place, but so do comprehensive process evaluations, qualitative studies, and a host of other evaluation designs. We believe that it is possible to glean useful, if not conclusive, evidence of the impact of a program from an evaluation which does not rise to the standard recommended by GAO because of the unavoidable absence of one or more elements.

More rigorous impact evaluations also cost more than those which may admittedly provide less scientifically valid findings. The combined process and impact evaluations of VAWA programs reviewed by GAO cost between $468,552 and $1,130,574 each, and this is by no means the upper limit; in contrast, the impact-only evaluation of the CAR project costs $1,034,737. Within OJP, we are examining the issue of how to finance effective impact evaluations.

OJP, through NIJ, is committed to providing research and evaluation knowledge to officials at all levels of government to reduce and prevent crime and improve the operations of

---

advocates, prosecutors, probation officers, and an entirely new dedicated police unit.

3

crime and justice agencies throughout the country. We are mindful of the significant responsibilities that fall to us as we strive to bring the tools of social science to the urgent questions of how best to improve programs and processes. We share the enthusiasm apparent in the GAO report for evaluations that use rigorous methodologies and we are committed to employing such methodologies, where feasible and within available funds.

OJP appreciates the opportunity to comment on the draft report. Additional specific comments are enclosed for GAO's consideration.

Sincerely,

Deborah J. Daniels
Assistant Attorney General

Enclosure

cc: Sarah V. Hart, Director
National Institute of Justice

Richard Nedelkoff, Director
Bureau of Justice Assistance

Diane Stuart, Director
Violence Against Women Office

Cynthia J. Schwimer
Comptroller

LeToya Bryant
OJP Audit Liaison

Vickie L. Sloan
DOJ Audit Liaison

OAAG Executive Secretariat
Control No. 20020204

4

Enclosure

OJP's Specific Comments on GAO Draft Report Entitled
"Justice Impact Evaluations: One Byrne Evaluation Was Rigorous; All Reviewed Violence Against Women Office Evaluations Were Problematic" (GAO-02-309)

See comment 1.

1. Throughout the report, the authors have failed to make and maintain the essential distinction between legislative initiatives and agency programs. In particular, references in the title of the report to "Byrne" and "VAWO" suggests an equivalence in origin and intent. This confusion is compounded by linking Byrne (a discretionary funding program established by legislation) to a program independently designed and implemented by an inter-agency team (Children at Risk) and linking VAWO (a grant making office that administers VAWA programs) to programs and initiatives whose scope, focus, and implementation are derived directly from legislation. This inconsistency recurs throughout the report, obscuring the statutory derivation of these programs.

   The counterpart to VAWO is the Bureau of Justice Assistance (BJA), not Byrne. We recommend that, for accuracy, throughout the report, GAO change references to "VAWO programs" and "VAWO evaluations" to "VAWA programs" and "VAWA evaluations."

See comment 2.

2. Throughout the report, comments referencing the "variations across grantee sites in how [VAWA] programs are implemented" fail to note that this variability was by intention, reflecting provisions of the original legislation. Congress passed VAWA after four years of exhaustive investigation focused on the extent and severity of domestic violence, sexual assault, and stalking committed in the United States. Having concluded that violence against women is a problem of national scope, Congress enacted a comprehensive legislative package targeting violence against women. Rather than imposing a strictly federal response to these crimes, Congress mandated that VAWA grant funds be used to support state- and community-based responses to domestic violence, sexual assault, and stalking.

   The statutory framework of the Rural Domestic Violence and Child Victimization Enforcement Grants ("Rural Program") illustrates the program variation and flexibility that Congress built into the VAWA grant programs. Under the Rural Program, eligible grantees include States, Indian tribal governments, local governments of rural States and other public or private entities of rural states. 42 U.S.C. § 13971(a). Grantees may use these grants for one or more of three statutory purposes: 1) to implement, expand, and establish cooperative efforts and projects between law enforcement officers, prosecutors, victim advocacy groups, and other related parties to investigate and prosecute incidents of domestic violence, dating violence and child abuse; (2) to provide treatment, counseling, and assistance to victims of domestic violence, dating violence and child abuse, including in immigration matters; and (3) to work in cooperation with the community to develop

1

education and prevention strategies directed toward such issues. Id.[3] Consequently, Rural Program grantees include such diverse entities as schools, church organizations, city, county and state governments, district attorneys' offices, sheriffs' offices, shelters, legal advocacy organizations, and immigration organizations. Moreover, funded activities may include a combination of such diverse activities as community or public education, direct victim services, direct advocacy, specialized domestic violence investigation or prosecution units, community organizing, and school intervention.

Although the draft report recognizes at a number of points that there are wide variations in the projects supported by VAWA funds under the Rural, Arrest and STOP Violence Against Indian Women Programs, the report never explains that, in large part, Congress itself mandated these variations. Instead, the report implies that the VAWA programs' "heterogeneity" - and the obstacles that such heterogeneity creates for researchers - is the result of choices that VAWO made in structuring its programs. This treatment of program variation is both misleading and unfair. As the report acknowledges, program variation does make program evaluation "particularly arduous." It should not, however, be categorized as a "methodological problem." Rather, it should be discussed as a reason why the VAWA program evaluations could not meet, and should not be held to, the standards of the CAR evaluation.

We agree that demonstration projects have great value and VAWO is funding or contributing funding to four such initiatives in partnership with NIJ and others: the Judicial Oversight Demonstration Initiative; Collaborations to Address Domestic Violence and Child Maltreatment: A Public-Private Initiative (also known as the Greenbook Initiative); the Comprehensive Indian Resources for Community and Law Enforcement Project; and the Safe Kids/Safe Streets Initiative.

Now on pp.15-16.
See comment 3.

3. On page 2 of the report, the statement that "peer review committees found methodological problems in two of these three [VAWA] evaluations" is misleading. Significant clarification would be added by changing this to read "...in two of the three VAWA evaluation *proposals* that were eventually funded by NIJ," or wording similar to that used on page 15. Also, please note, peer review panels are instructed to raise concerns and cautions and do so for virtually every proposal submitted. As NIJ discussed with GAO officials, in each case, the peer review panel identified the eventual grantee as the superior application and recommended that NIJ fund it. In other words, NIJ selected these proposals based on the findings of the peer review committees, not despite them, as the report seems to suggest.

[3] This statutory language includes changes made to the Rural Program's purpose areas by the Violence Against Women Act of 2000 (VAWA 2000), Pub. L. No. 106-386, §§ 1109(d) and 1512(c). Prior to the passage of VAWA 2000, and during the fiscal years included in NIJ's Rural Program evaluation, the Rural Program's purpose areas did not address explicitly dating violence and victim services in immigration matters.

2

See comment 4.

4. It is incorrect that "peer review panels ... make recommendations to NIJ's Director of the Office of Research and Evaluation..." Written consensus reviews, authored by the peer review panel, are submitted unaltered to the NIJ Director, who has final authority for making all NIJ grants. Prior to presentation of these reviews to the Director, each is reviewed and discussed with partnership agency representatives (e.g., staff from VAWO or BJA). These internal staff reviews and discussions are lead by the Director of the Office of Research and Evaluation who then presents the reviews, along with agency and partner agency input, to the NIJ Director for consideration and final grant award decisions.

Now on p. 11.

Now on pp. 14 and 18.

See comment 5.

5. On page 10, the report discusses external factors that could account for changes that the Rural Program evaluation observed in victims' lives and the criminal justice system. (See also similar comments on pages 13 and 19). The external factors listed here and elsewhere in the report, however, include some of the very types of activities that may be attributable to Rural Program funding. For example, a co-occurring program in a church could be the result of education in churches funded by the Rural Program. Likewise, new state or local legislation may be the result of public and community education. The Rural Program's evaluation itself specifically examined co-occurring events, sorted out those not related to the Program, and demonstrated in a number of instances that co-occurring events could be linked to education funded through the Rural Program.

In addition, the draft report cites an alcohol abuse treatment program as a possible external factor that might account for perceived changes. While alcohol may be an exacerbating factor that contributes to violence against women, alcohol consumption does not cause domestic violence, stalking or sexual assault. Perpetrators who are substance abusers have two distinct problems -- abusing alcohol and committing violence against women -- requiring two separate solutions. Addressing alcohol abuse solves only one problem. The other continues to exist because of beliefs and attitudes that result in the physical and sexual abuse of women, whether or not alcohol is involved. We therefore recommend that GAO remove or replace the alcohol abuse example, which is not necessary to its larger point and could be seen as endorsing the stereotype that alcohol abuse "causes" domestic violence.

See comment 6.

6. On pages 10 and 11, the report draws a contrast in site selection methodology between the Arrest Policies Evaluation (where 6 of 130 sites were studied) and the CAR project (where "all five sites participated") which overlooks a fundamental programmatic difference between the CAR and the VAWA evaluations that necessitated use of different evaluation strategies. In essence, CAR was a research-driven field test or demonstration project designed and implemented specifically for research purposes permitting – necessitating, in fact – the collection of baseline and follow-up data, randomization of schools, and strict compliance to program parameters, data protocols, and analysis strategies as dictated by the research design. The VAWA evaluations were not designed or conducted in a fashion even remotely resembling this research-driven approach. The

3

Now on p. 28.

Now on p. 17.

description of CAR on page 22 in the appendix acknowledges this important difference, though the text on page 16 obscures the distinction.

In addition, the report does not address, however, how immensely expensive the Arrest evaluation would have become if it had included all 130 sites. By focusing exclusively on the impact aspect of the Arrest evaluation, the report may mislead a reader regarding the scope of the Arrest process evaluation.

At the evaluators' request, VAWO staff assisted with the Rural Program evaluation to ensure that the sites selected were representative, to the greatest degree possible, of the vast range of grantees funded under the Rural Program. The selected sites represent varying types of grantees, different size grantees, and projects that address different Rural program goals. We note that 111 of 130 project directors responded to the evaluator's survey on services, 20 Arrest sites were selected, based on diversity in type of project and geography, for the process evaluation. Of those 20, 6 were selected for the impact evaluation, maximizing geographical and purpose area diversity while focusing on sites with high program integrity.

The statement that "all three VAWO programs permitted their grantees broad flexibility in the development of their projects to match the needs of their local communities" seems to suggest that this flexibility was a result of choices made by the VAW Office that administers the grants. In fact, this flexibility originates in the VAWA itself. While the authors of the VAWA may not have anticipated that allowing participating grant sites to emphasize different project configurations would make generalizing research findings difficult, the practical benefits to communities of such a flexible programmatic approach must not be minimized. Noting that, as a result, the evaluation "may not provide information that could be generalized to a broader implementation of the program" seems to suggest that this limitation derives from the evaluation design, rather than its true origin: the statutes governing the VAWA grant program.

Now on p. 13.

See comment 7.

7. On page 12 , the report describes the advantages of comparison groups for isolating the effect of an intervention, describing as an example the comparison of drug court clients to another "set of individuals, 'the comparison group,' who are experiencing similar life circumstances but who do not qualify for drug court participation." This analogy fails to recognize important differences in the unit of analysis (i.e., individual clients of a drug court vs. agencies for victim services or whole communities). In several cases, the VAWA evaluations studied networks of agencies and service providers who were to develop and implement a comprehensive strategy out of which specific implementations might flow. Comparison groups of individuals would tell us little about whether communities had been successful in building cohesive partnerships for responding to victims or whether adequate victim incident reporting systems had been developed and implemented.

4

Now on p. 14.
See comment 8.

8. On page 13, the report states that "the lack of comparison groups in the Rural Domestic Violence evaluation makes it difficult to conclude that a reduction in violence against women and children in rural areas can be attributed . . . to the Rural Domestic program." Reduction of violence, however, is not the only indicator that VAWA grant programs are working. Although the long-term goal of VAWA funding is to reduce violence, in the first years of a successful grant we expect instead to see outputs such as more programs and assistance for victims, more community involvement, more orders of protection, more arrests, and more prosecutions.

Now on p. 15.
See comment 9.

9. On page 14, the report accurately describes the low response rates for some study elements of the VAWA STOP evaluation. It also correctly notes that "the evaluation team did not collect baseline data prior to the start of the program." Both of these statements seem to suggest that the problems resulted from a flaw in the evaluation design or a shortcoming in the efforts of the evaluator. Those familiar with victim research will recognize the challenges that the most expert researchers face in securing victim data. This is especially true in the area of domestic violence and sexual assault where victims often do not want to participate in research, fear for their safety, or live such residentially unstable lives as a result of abuse that they can not be located.

Practitioners in the violence against women field also recognize that, prior to the VAWA, standardized baseline data on the prevalence of violence against women were virtually nonexistent in the U.S. Particularly in many rural and Indian Country communities, resources still do not exist to collect, compile, retain, and report basic victim data. Addressing this serious lack of empirical incident data was one of the explicit objectives of the legislation.

The GAO report should discuss the inherent difficulties of collecting information in Indian Country in order to provide a more balanced context for the problems faced by the STOP VAIW evaluators.[4] First, many tribes are only beginning to collect statistics and have very little technology or training in this field. Second, because of criminal jurisdiction issues in Indian country, many of the statistics about crimes against women are actually collected and analyzed by outside agencies and are not easily accessible to the tribes. Finally, tribes are often reluctant to turn over information to researchers because, in their view, this information historically has been used against tribes. While tribes should realize that accepting federal grants means some minor relinquishment of

[4] The following resources provide in-depth analysis regarding why research, and particularly social science research, in Indian country is so challenging: Thorton, Russell, ed., Studying Native America: Problems and Prospects, University of Wisconsin Press (1999); Mihesuah, Devon, Native and Academics: Writing About American Indians, University of Nebraska Press (1998); Smith, Linda Tuhiwai, Decolonizing Methodologies: Research and Indigenous Peoples, Zed Books (1999).

5

sovereignty, there is still a great deal of distrust and suspicion about providing information, particularly negative information about crime rates, to the United States government.

Now on p. 5.
See comment 10.

10. On page 14, Footnote 7 should explain that the Sex Offender Management Training Discretionary Grant program is administered by the Corrections Program Office.

Now on p. 25.
See comment 11.

11. On page 19, the report states that the Rural Program has funded 92 grants from 1996 through 2001. Please note, the Rural Program has made 317 awards (including continuation awards) to 140 rural jurisdictions. Furthermore, a better description of the program would more closely track the statute and would use gender-neutral language:

"The primary purpose of the Rural Program is to enhance the safety of victims of domestic violence, dating violence and their children by supporting projects designed to address and prevent domestic violence, dating violence, and child victimization in rural America. The Rural Program welcomes applications that propose innovative solutions to obstacles for abused victims and children created by the rural nature of a particular community. Unique partnerships, typically not found in urban settings, are encouraged. The Rural Program will support projects that: implement, expand, and establish cooperative efforts and projects between law enforcement officers, prosecutors, victim advocacy groups, and other related parties to investigate and prosecute incidents of domestic violence, dating violence, and child abuse; provide treatment, counseling and assistance to victims of domestic violence, dating violence, and child abuse, including in immigration matters; and work in cooperation with the community to develop education and prevention strategies directed toward such issues."

Now on pp. 25-27.
See comment 12.

12. Notes in the appendices (pages 20-21) that "NIJ could not separate the cost of the impact evaluation from the cost of the process evaluation" seems to suggest that such a distinction is merely a matter of bookkeeping. In fact, much of the effort toward the process evaluation was interconnected with the impact evaluation effort. For instance, studying and reporting on sites' achievements regarding basic measures and methods for data collection and data-sharing, typically completed during a process phase, would have important (and efficient) utility for the impact evaluation to come after. A clearer statement would be: "Attempts to estimate separately the costs associated with the process and impact evaluations might obscure the way in which process evaluation work contributed directly to shape, inform, and guide the impact evaluation."

Now on p. 26.
Now on p. 25.
Now on p. 27.
See comment 13.

13. The statement on page 20 that VAWA Arrest grantees "have flexibility to work in several" program areas is also true of the VAWA Rural grantees (page 19) and the VAWA STOP grantees (page 21). Most important, however, is that this flexibility reflects the intention of the legislation that created these programs and is not purely a choice made by the Violence Against Women Office.

6

Now on p. 31.

See comment 14.

14. On page 25, in discussing the Comprehensive Communities Program evaluation findings, the report undervalues the utility of process evaluations to uncover how well a program works and what its programmatic outcomes are. The report suggests that "the extent to which a program is operating as intended" is not particularly important among the "results" of a program. Sometimes, these are exactly the kinds of "results" needed, particularly to facilitate replication of a program to other communities.

7

Following are GAO's comments on the Department of Justice's February 13, 2002, letter.

GAO Comments

1. We have amended the text to further clarify that BJA administers the Byrne program, just as its counterpart, VAWO, administers its programs (see page 4). However it is important to point out that regardless of the issues raised by OJP, the focus of our work was on the methodological rigor of the evaluations we reviewed, not the purpose and structure of the programs being evaluated. As discussed in our Scope and Methodology section, our work focused on program evaluation activities associated with Byrne and VAWO discretionary grant programs generally and the methodological rigor of impact evaluation studies associated with those programs in particular. To make our assessment, we relied on NIJ officials to identify which of the program evaluations of Byrne and VAWO grant programs were, in fact, impact evaluation studies. We recognize that there are substantial differences among myriad OJP programs that can make the design and implementation of impact evaluations arduous. But, that does not change the fact that impact evaluations, regardless of differences in programs, can benefit from stronger up-front attention to better ensure that they provide meaningful and definitive results.

2. We disagree with OJP's assessment of our report's treatment of program variation. As discussed earlier, the scope of our review assessed impact evaluation activities associated with Byrne and VAWO discretionary grant programs, not the programs themselves. We examined whether the evaluations that NIJ staff designated as impact evaluations were designed and implemented with methodological rigor. In our report we observe that variations in projects funded through VAWO programs complicate the design and implementation of impact evaluations.

   According to the Assistant Attorney General, this variation in projects is consistent with the intent of the programs' authorizing legislation. We recognize that the authorizing legislation provides VAWO the flexibility in designing these programs. In fact, we point out that although such flexibility may make sense from a program perspective, project variation makes it much more difficult to design and implement a definitive impact evaluation. This poses sizable methodological problems because an aggregate analysis, such as one that might be constructed for an impact evaluation, could mask the differences in effectiveness among individual projects and therefore not result in information about which configurations of projects work and which do not.

3. We have amended the Results in Brief to clarify that peer reviews evaluated proposals. However, it is important to note that while the peer review committees may have found the two VAWO grant applications to be the most superior, this does not necessarily imply that the impact evaluations resulting from these

applications were well designed and implemented. As discussed in our report, the peer review panel for each of the evaluations expressed concerns about the proposals that were submitted, including issues related to site selection and the need for additional impact measures and control variables. Our review of the documents NIJ made available to us, including evaluators' responses to peer review comments, led to questions about whether the evaluators' proposed methodological designs were sufficient to allow the evaluation results to be generalized and to determine whether the program was working.

4. We have amended the Background section of the report to add this information (see page 6).

5. As discussed in OJP's comments, we discussed external factors that could account for changes that the Rural Program evaluation observed in victims' lives and the criminal justice system. We did so not to critique or endorse activities that the program was or was not funding, but to demonstrate that external factors may influence evaluation findings. To the extent that such factors are external, the Rural Program evaluation methodology should account for their existence and attempt to establish controls to minimize their affect on results (see page 14). We were not intending to imply that alcohol is a cause for domestic violence, as suggested by the Assistant Attorney General, but we agree that it could be an exacerbating factor that contributes to violence against women.

6. As discussed earlier, we recognize that there are substantive differences in the intent, structure, and design of the various discretionary grant programs managed by OJP and its bureaus and offices. Also, as stated numerous times in our report, we acknowledge not only that impact evaluation can be an inherently difficult and challenging task but also that measuring the impact of these specific Byrne and VAWO programs can be arduous given that they are operating in an ever changing, complex environment. We agree that not all evaluation issues that can compromise results are easily resolvable, but we firmly believe that with more up-front attention to design and implementation issues, there is a greater likelihood that NIJ impact evaluations will provide meaningful results for policymakers.

   Regarding the representativeness of sites, NIJ documents that were provided during our review indicated that sites selected during the Rural Program evaluation were selected on the basis of feasibility, as discussed in our report—specifically, whether the site would be among those participants equipped to conduct an evaluation. In its comments, OJP stated that the 6 sites selected for the impact evaluation were chosen to maximize geographical and purpose area diversity while focusing on sites with high program priority. OJP did not provide any additional information that would further indicate that the sites were selected on a representative basis. OJP did, however, point out that the report does not address how immensely expensive the Arrest evaluation would have become if it had included all 130 sites. We did not address specific evaluation site costs

because we do not believe that there is a requisite number of sites needed for any impact evaluation to be considered methodologically rigorous. Regarding OJP's comment about the flexibility given to grantees in implementing VAWO grants, our report points out that project variation complicates evaluation design and implementation. Although flexibility may make sense from a program perspective, it makes it difficult to generalize about the impact of the entire program.

7. We used the drug court example to illustrate, based on our past work, how comparison groups can be used in evaluation to isolate and minimize external factors that could influence the study results. We did not, nor would we, suggest that any particular unit of analysis is appropriate for VAWO evaluations since the appropriate unit of analysis is dependent upon the specific circumstances of the evaluation. We were only indicating that since comparison groups were not utilized in the studies, the evaluators were not positioned to demonstrate that change took place as a result of the program.

8. We do not dispute that VAWO grant programs may provide valuable outputs over the short term. However, as we have stated previously, the focus of our review was on the methodological rigor of impact evaluations--those evaluations that are designed to assess the net effect of a program by comparing program outcomes with an estimate of what would have happened in the absence of the program. Given the methodological issues we found, it is unclear whether NIJ will be able to discern long-term effects due to the program.

9. As stated in our report, we acknowledge not only that impact evaluation can be an inherently difficult and challenging task, but that measuring the impact of Byrne and VAWO programs can be arduous given the fact that they are operating in an ever changing, complex environment. We agree that not all evaluation issues that can compromise results are easily resolvable, but we firmly believe that, with more up-front attention to design and implementation issues, there is a greater likelihood that NIJ evaluations will provide meaningful results for policymakers. As we said before, absent this up-front attention, questions arise as to whether NIJ is (1) positioned to provide the definitive results expected from an impact evaluation and (2) making sound investments given the millions of dollars spent on these evaluations. If NIJ believes that the circumstances of a program are such that it cannot be evaluated successfully (in relation to impact) they should not proceed with an impact evaluation.

10. We have amended the footnote to state that from fiscal year 1995 through fiscal year 1999, this program was administered by VAWO. As of fiscal year 2000, responsibility for the program was shifted to OJP's Corrections Program Office (see page 5).

11. In regard to the number of grants, we have amended the text to reflect that the information NIJ provided during our review is the number of grantees, not the

number of grants (see pages 25 and 26). We have also amended our report to reflect some of the information provided in VAWO's description of the Rural Domestic Violence Program to further capture the essence of the program (see page 25).

12. We disagree. We believe that separating the cost of the impact and process evaluations is more than a matter of bookkeeping. Even though the work done during the process phase of an evaluation may have implications for the impact evaluation phase of an evaluation, it would seem that, given the complexity of impact evaluations, OJP and NIJ would want to have in place appropriate controls to provide reasonable assurance that the evaluations are being effectively and efficiently carried out at each phase of the evaluation. Tracking the cost of these evaluation components would also help reduce the risk that OJP's, NIJ's, and, ultimately, the taxpayer's investment in these impact evaluations is not wasted.

13. As discussed earlier, we recognize that there are substantive differences in the intent, structure, and design of the various discretionary grant programs managed by OJP and its bureaus and offices, including those managed by VAWO. Our report focuses on the rigor of impact evaluations of grant programs administered by VAWO and not on the program's implementing legislations. Although flexibility may make sense from a program perspective, it makes it difficult to develop a well designed and methodologically rigorous evaluation that produces generalizeable results about the impact of the entire program.

14. Our report does not suggest that other types of evaluations, such as comprehensive process evaluations, are any less useful in providing information about how well a program is operating. The scope of our review covered impact evaluations of Byrne and VAWO discretionary grant programs—those designed to assess the net effect of a program by comparing program outcomes with an estimate of what would have happened in the absence of the program.

Appendix IV: GAO Contacts and Staff Acknowledgments

GAO Contacts

Laurie E. Ekstrand, (202) 512-8777
John F. Mortin, (202) 512-8777

Acknowledgments

In addition to the above, Wendy C. Simkalo, Jared A. Hermalin, Chan My J. Battcher, Judy K. Pagano, Grace A. Coleman, and Ann H. Finley made key contributions to this report.

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily e-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
P.O. Box 37050
Washington, D.C. 20013

To order by Phone:	Voice:	(202) 512-6000
	TDD:	(202) 512-2537
	Fax:	(202) 512-6061

Visit GAO's Document Distribution Center

GAO Building
Room 1100, 700 4th Street, NW (corner of 4th and G Streets, NW)
Washington, D.C. 20013

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm,
E-mail: fraudnet@gao.gov, or
1-800-424-5454 or (202) 512-7470 (automated answering system).

Public Affairs

Jeff Nelligan, Managing Director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G. Street NW, Room 7149,
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER