University of Pennsylvania
ScholarlyCommons

Publicly accessible Penn Dissertations

Fall 12-22-2009

# Unpacking the impact of Restorative Justice in the RISE experiments: Facilitators, offenders, and conference non-delivery

Daniel J. Woods
*University of Pennsylvania*, woodser.dan@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Criminology Commons

Recommended Citation
Woods, Daniel J., "Unpacking the impact of Restorative Justice in the RISE experiments: Facilitators, offenders, and conference non-delivery" (2009). *Publicly accessible Penn Dissertations.* Paper 73.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/73
For more information, please contact repository@pobox.upenn.edu.

# Unpacking the impact of Restorative Justice in the RISE experiments: Facilitators, offenders, and conference non-delivery

**Abstract**
Restorative Justice (RJ) programs are often evaluated in terms of their outcomes, with little attention to the process. Typically we analyze average effects across individuals who experience RJ differently. The present dissertation unpacks these different effects in three separate inquiries utilizing data from the Reintegrative Shaming Experiments (RISE) conducted in Canberra, Australia from 1995 - 2000.

First, we descriptively assess the extent RJ conference facilitators engender perceptions of procedural justice and legitimacy in offenders. We examine the number of conferences delivered (experience), sequential conferences (practice-makes-perfect) and the timing between conferences (skill maintenance). Certain conference facilitators are better than others from the outset. We recommend the identification of RJ facilitators who are good at promoting perceptions of procedural justice and legitimacy. Second, we utilize trajectory analysis and find the impact of RJ varies by offending group, with negative effects observed for Aboriginal offenders. Finally, utilizing multinomial logistic regression, we examine the characteristics associated with non-delivery of RJ. Randomized controlled trials, such as RISE, rely on treatment integrity to best assess the impact of the assigned treatment. From a policy standpoint, the most efficient use of resources would rely on successful conference delivery. We find that the time between random assignment and the first conference attempt is significantly related to successful delivery.

This dissertation takes important steps in understanding the importance of unpacking the impact of RJ and helps inform who should conduct RJ conferences, what groups of individuals to include in future studies, and what impacts non-delivery of RJ conferences.

**Degree Type**
Dissertation

**Degree Name**
Doctor of Philosophy (PhD)

**Graduate Group**
Criminology

**First Advisor**
Lawrence Sherman

**Second Advisor**
Richard Berk

**Third Advisor**
John MacDonald

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/73

**Keywords**
Restorative Justice, Evaluations, RISE, crime, treatment integrity

**Subject Categories**
Criminology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/73

# UNPACKING THE IMPACT OF RESTORATIVE JUSTICE IN THE RISE EXPERIMENTS: FACILITATORS, OFFENDERS, AND CONFERENCE NON-DELIVERY

Daniel James Woods

A DISSERTATION

in

Criminology

Presented to the Faculties of the University of Pennsylvania

In

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2009

Supervisor of Dissertation

_____

Lawrence Sherman, Professor of Criminology

Graduate Group Chairperson

_____

William Laufer, Professor of Criminology, Sociology, and Legal Studies & Business Ethics

Dissertation Committee:

Lawrence Sherman, Professor, Criminology

Richard Berk, Professor, Criminology and Statistics

John MacDonald, Assistant Professor, Criminology

Dedicated to Hannah Elizabeth Woods & Emme Catherine Woods

# ACKNOWLEDGEMENTS:

I would like to thank my committee; Lawrence Sherman, Richard Berk, and John MacDonald for their assistance, comments, and patience in this process.

In addition, I would like to thank my family for their continual support, inquiry, pestering, etc. in all of my academic endeavors.

I need to give a special thank you to Janel McCaffrey and Knakiya Hagans. They both tolerated me and made the office a brighter place to be in. Janel might not realize it, but her support made sure this dissertation was not “frah-gee-lay.” Without Knakiya’s help navigating the bureaucracy, I might not have made it. In addition, her take-no-guff attitude is an asset to us all. Thank y’all very much.

Most important, I would like to thank Charlotte Gill for her amazing support and love throughout this process. Her editorial support was invaluable. Writing the dissertation has coincided with some of the biggest tests of my life and there were times when I was ready to throw in the towel, but she has stood by me, encouraged me, and loved me to the point where I am forever in her debt. Thank you so much.

# ABSTRACT

UNPACKING THE IMPACT OF RESTORATIVE JUSTICE IN THE RISE EXPERIMENTS: FACILITATORS, OFFENDERS, AND CONFERENCE NON-DELIVERY

Daniel James Woods

Lawrence Sherman

Restorative Justice (RJ) programs are often evaluated in terms of their outcomes (e.g. impact on crime), with little attention to the process. RJ's impact is often assessed by examining the difference between RJ "treated" groups with controls who typically experience standard criminal justice system processing. This strategy yields average effects across individuals who experience RJ differently. The present dissertation unpacks these different effects in three separate inquiries utilizing data from the Reintegrative Shaming Experiments (RISE) conducted in Canberra, Australia from 1995 - 2000.

First, borrowing a medical model, we descriptively assess the extent to which conference facilitators engender perceptions of procedural justice and legitimacy in offenders who attend RJ. We graphically examine the number of conferences delivered per facilitator (experience), sequential conferences (practice-makes-perfect) and the timing between conferences (skill maintenance). Certain conference facilitators are just

better than others from the outset. We recommend the identification of RJ facilitators who are good at promoting perceptions of procedural justice and legitimacy, two concepts that are linked with reduced criminal behavior. Second, we utilize trajectory analysis to assess the impact of assignment to RJ on subsequent criminal offending for individuals following different offending trajectories. The impact of RJ varies by offending group, with negative effects observed for Aboriginal offenders. Finally, utilizing multinomial logistic regression, we examine the characteristics associated with non-delivery of RJ, either due to offender-related or administrative-related reasons. From a practical standpoint, randomized controlled trials, such as RISE, rely on treatment integrity to best assess the impact of the assigned treatment. From a policy standpoint, the most efficient use of resources would rely on successful conference delivery. We find that the time between random assignment and the first conference attempt is significantly related to successful delivery. Specifically, organizing a conference too soon or too late increases the chance that the conference will fail.

This dissertation takes important steps in understanding the importance of unpacking the impact of RJ and helps inform who should conduct RJ conferences, what groups of individuals to include in future studies, and what impacts non-delivery of RJ conferences.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS: iii
ABSTRACT iv
TABLE OF CONTENTS vi
LIST OF TABLES viii
LIST OF FIGURES ix
PREFACE: x
CHAPTER 1: Police-led Restorative Justice Conferencing: Experience, learning, or skill maintenance? 1
INTRODUCTION 1
RESEARCH QUESTIONS 8
RESEARCH DESIGN 10
RESULTS 14
Questions 1 & 2 - Practice makes perfect vs. Expertise: Procedural Justice & Legitimacy 14
Question 3 - Impact of Skill Maintenance on Procedural Justice & Legitimacy 16
Question 4 - Procedural Justice & Legitimacy: Consistency or Development 18
Question 5 - Impact of Procedural Justice & Legitimacy on Offending 20
DISCUSSION/LIMITATIONS/POLICY RECOMMENDATIONS 22
CHAPTER 2: The Differential Impact of Restorative Justice on Offending Trajectories 38
INTRODUCTION: 38
DATA AND METHOD 45
Outcome Variable: 47
Data and Analytic Strategy: 47
RESULTS 49
Selecting and describing the model 49
The impact of RJ on offending trajectories 52
Racial differences in the experience of RJ 56
DISCUSSION & CONCLUSIONS 59
CHAPTER 3: Characteristics of treatment non-delivery in the Reintegrative Shaming Experiments 70
INTRODUCTION 70
METHODS 73
Data 73

Measures 75
RESULTS 78
Analytic Strategy 79
An Illustration 82
DISCUSSION & CONCLUSIONS 83
Appendices 90
APPENDIX A: FIT STATISTICS FOR THE MULTINOMIAL LOGISTIC REGRESSION MODEL 90
APPENDIX B: TEST OF THE INDEPENDENCE OF IRRELEVANCE ASSUMPTION (IIA) 91
APPENDIX C: LIKELIHOOD-RATIO TEST FOR COMBINING ALTERNATIVES 92
APPENDIX D: VARIANCE INFLATION FACTORS OF THE INCLUDED VARIABLES 93
Bibliography: 94

# LIST OF TABLES

Table 1.1 Descriptive Statistics for Outcome Variables 28
Table 1.2: Cross-tabulation of Average Procedural Justice & Legitimacy Quartile 28
Table 1.3: Change in Personal Victim Crimes 1 Year Pre-/Post Random Assignment by Quartile of Average Procedural Justice 28
Table 1.4: Change in Personal Victim Crimes 1 Year Pre-/Post Random Assignment by Quartile of Average Legitimacy 29
Table 2.1: Fit Statistics for 2-7 Trajectories of Personal Victim Offending 62
Table 2.2: Regression Results for Trajectory Slope Parameters 63
Table 2.3: Posterior Probability Group Assignments for 6 Group Model 64
Table 2.4: Characteristics of Group Members 64
Table 3.1: Descriptive Statistics for Analysis Variables 86
Table 3.2: Probability of Non-delivery (Offender related or Administrative) given sample characteristics 87
Table 3.3: Multinomial Regression Results for Non-Treatment 88

# LIST OF FIGURES

Figure 1.1: Number of conference facilitators per quarter 30
Figure 1.2: Average N of conducted conferences per facilitator 30
Figure 1.3: Procedural Justice by Type of Facilitator Experience 30
Figure 1.4: Awareness of the Process by Type of Facilitator Experience 31
Figure 1.5: Consistency & Fairness by Type of Facilitator Experience 31
Figure 1.6: Correctability of the Process by Type of Facilitator Experience 31
Figure 1.7: Control of the Process by Type of Facilitator Experience 32
Figure 1.8: Impartiality of Conference Attendees by Type of Facilitator Experience 32
Figure 1.9: Ethicality of Conference Attendees by Type of Facilitator Experience 33
Figure 1.10: Respect by Type of Facilitator Experience 33
Figure 1.11: Legitimacy by Type of Facilitator Experience 33
Figure 1.12: Procedural Justice by Weeks between Conferences 34
Figure 1.13: Awareness of the Process by Weeks between Conferences 34
Figure 1.14: Consistency & Fairness by Weeks between Conferences 34
Figure 1.15: Correctability of the Process by Weeks between Conferences 35
Figure 1.16: Control of the Process by Weeks between Conferences 35
Figure 1.17: Impartiality of Conference Attendees by Weeks between Conferences 35
Figure 1.18: Ethicality of Conference Attendees by Weeks between Conferences 36
Figure 1.19: Respect by Weeks between Conferences 36
Figure 1.20: Legitimacy by Weeks between Conferences 36
Figure 1.21: The Development of Procedural Justice over Sequential Conferences by Quartile of Average Procedural Justice. 37
Figure 1.22: The Development of Legitimacy over Sequential Conferences by Quartile of Average Legitimacy 37
Figure 2.1: Predicted trajectories for the 6 group ZIP model 65
Figure 2.2: Impact of RJ on offending trajectories – 6 group model 65
Figure 2.3: Group 1 (Low Drifters) – Impact of RJ on offending 66
Figure 2.4: Group 2 (High Drifters) – Impact of RJ on offending 66
Figure 2.5: Group 3 (Little Blips) – Impact of RJ on offending 67
Figure 2.6: Group 4 (Big Blips) – Impact of RJ on offending 67
Figure 2.7: Group 5 (Low Plateau) – Impact of RJ on offending 68
Figure 2.8: Group 6 (Steep Risers) – Impact of RJ on offending 68
Figure 2.9: Non-Aboriginal subsample – Impact of RJ on offending 69
Figure 3.1: Prediction Profiler from Multinomial Logistic Regression Results 89

# PREFACE:

This dissertation represents three separate inquiries into the differential impact of Restorative Justice (RJ). Evaluations of RJ, in general, focus on the average effect across individuals participating in the process. This strategy generally masks the differential effects RJ can have on different sets of individuals. This dissertation unpacks the impact of RJ under different circumstances. Each of the following chapters represents a stand-alone paper, and within each paper we look at different aspects of the RJ process utilizing data from the Reintegrative Shaming Experiments (RISE) conducted in Canberra, Australia (Australian Capital Territory) from 1995 – 2000.

Chapter 1 looks at the impact of conference facilitators on offender perceptions of procedural justice and legitimacy. Although, in general, RJ is more likely to contain elements of procedural justice which, in turn, can increase legitimacy than conventional criminal justice system processing, the ability to influence these psychological mechanisms will certainly vary depending on the ability of the conference facilitator. We examine the extent to which those facilitators with the most experience in conference delivery (expertise) differed from the ongoing learning process accrued through subsequent conferences (practice-makes-perfect). Finally, as with all skills, periodic practice is meant to keep one's skills sharp, thus we look at the time between subsequent conferences (skill maintenance) and its impact on perceptions of procedural justice and legitimacy. We utilize a graphic examination of these mechanisms for expertise to determine if any pattern exists between conference facilitators' experience and offenders' perceptions. We find the relationship between facilitator expertise and offender perceptions is not related to any of the above mechanisms. Those facilitators who are in

the top quartile of procedural justice tend to remain there throughout their conference careers. We offer that identification of RJ conference facilitators who are proficient at engendering perceptions of procedural justice and legitimacy is important in the most effective delivery of restorative justice.

In Chapter 2 we examine the extent to which the impact of RJ on future offending is dependent on the criminal "career" of the offenders involved. We utilize trajectory modeling to determine if there are groups of individuals following similar offending pathways over time using 10 years of official criminal history. We follow this up by examining the impact of assignment to RJ as compared to standard criminal justice system (CJS) processing on future offending post-random assignment. That is to say, is the impact of RJ dependent on the developmental pathway of the offenders involved? Examining a six group model, initially we find no difference for two groups, a reduction in crime in favor of RJ for two groups, and a backfire effect for RJ in the final two groups (which included the most prolific offenders). Further examination reveals the two "backfire" groups contain the largest percentage of Aboriginal offenders. When we focus on the "white" sample, RJ produces a decrease in offending for these relatively high offending groups. We can see that the impact of RJ on future offending varies depending on the offending pathway of the offenders involved. In addition, RJ with Aboriginal offenders produces a backfire effect, at least in the ACT. Practitioners who want an RJ program to have the largest crime reduction should focus on these individuals. In addition, RJ may provide a low-cost, less intrusive alternative to court among low-level or first-time offenders.

Finally, Chapter 3 examines the characteristics related to non-delivery of assigned treatment in RISE experiments. In randomized controlled trials, the impact of assigned treatment is most relevant on assigned cases actually receiving the assigned treatment. Departures from the assigned treatment, certainly in the case of field experiments, should be expected and measures should be taken beforehand to minimize them. Although the RISE experiments report a relatively high rate of treatment integrity, approximately 24 percent of cases assigned to RJ did not receive the assigned treatment. Using multinomial logistic regression, we examine treatment failure due to administrative reasons (e.g. deciding the case did not warrant a conference or the victim refused to participate) and due to offender-related reasons (e.g. the offender moved out of the territory or was arrested prior to treatment delivery) as compared to a successfully completed conference. We find that the most important characteristic is the time it takes to organize the initial RJ treatment attempt. This relationship is curvilinear in that in the time following random assignment is not optimal to organize a conference; however, not being able to organize an conference until several months post-random assignment has an increasing probability of treatment non-delivery. There are post hoc statistical corrections available to correct for treatment non-delivery, but if we know the characteristics related to this phenomenon beforehand, we can maximize our chances of delivering a completed conference from a practical standpoint and improve our estimates of the impact of RJ from a research standpoint.

# CHAPTER 1: Police-led Restorative Justice Conferencing: Experience, learning, or skill maintenance?

## INTRODUCTION

Restorative Justice (RJ) is "a process whereby parties with a stake in a specific offense collectively resolve how to deal with the aftermath of the offence and its implications for the future" (Marshall, 1999, p. 5). Offenders and victims, sometimes with supporters of each, meet in the presence of a facilitator to discuss a criminal event and decide what needs to happen to repair the harm caused. As a process it is both offender- and victim-focused. Theoretically, the RJ process provides a better forum, where offenders and victims have a voice in the proceedings leading to "better outcomes," than the typical experience of people who go through the criminal justice system. Generally the outcome of interest is the change in criminal behavior; however, other outcomes have included procedural justice and legitimacy, and changes in victim quality of life. On paper, RJ programs would appear to be a better use of resources for policy makers.

Despite the optimism for RJ, evaluations of the programs that have been implemented generally have produced mixed results. The question of whether RJ "works" may be an oversimplification. Some limited evidence that RJ may reduce crime[1] among some offenders, while for others the treatment yields no improvement over the

[1] Although other outcomes are available (and relevant), evaluation of RJ as a criminal justice policy continues to rely heavily on reductions in crime (or at least no evidence of increasing crime) to assess the effectiveness of RJ.

standard care (typically judicial processing).[2] In addition, RJ seems to be a blanket term covering a wide range of practices from face-to-face interactions between victim and offender to "shuttle communication" between parties through an intermediary. Arguably, face-to-face conferences are the most commonly evaluated type of RJ, and provide the setting for this paper.

RJ programs are often evaluated in terms of how they are doing (i.e. the outcomes), with little regard to what they are doing (i.e. the process). As Sherman, Strang and Woods (2003) note, "[w]hile theory and practice have grown at a rapid rate in recent years, the science of restorative justice (RJ) has lagged far behind" (p. 229). Most evaluations rely on outcome differences between RJ cases and comparison group cases. Few of them focus on the process of RJ.[3] Because the conduct of RJ can be highly variable, research that examines average treatment effects may mask important heterogeneity in the RJ process and how this, in turn, influences treatment outcomes. Stated differently, the average treatment effect of RJ may be made up of un-average RJ practices. Alternatively, RJ can be delivered with a "cookie-cutter" approach, whereby each case, with its unique elements, is treated similarly, similar to an assembly line. "When RJ…is rolled out quickly on a wide scale, there is a risk that many conferences will just 'go through the motions' to 'tick off a box', rather than treating each case as a kind of surgical procedure requiring careful advance planning, preparation, and follow-up" (Sherman & Strang, 2007, p. 21). Both possibilities necessitate examining how RJ conferences are conducted. Face-to-face RJ conferences typically consist of a facilitator

---

[2] In rare cases RJ appears to have backfired, e.g., Aboriginal offenders in Canberra, Australia; see Sherman & Strang (2007) for a discussion of instances where RJ has increased offending.
[3] However, see Hayes and Daly (2003)

(often a police officer, but not exclusively), the offender(s) and supporters, the victim(s) and supporters, and sometimes an interested community representative. Each case therefore entails a unique dynamic created in each instance as the principal actors (and facts) change.

Sherman et al. (2003) likened the role of the RJ conference facilitator to that of an eighteenth century ship captain: "[l]ike the captains of sailing ships, the leaders of RJ processes may have enormous influence on the success or failure of those processes" (p. 230). Conference facilitators are trained in the concepts and rationale of RJ, whereas the conference attendees are not. In addition, facilitators have the opportunity to conduct multiple conferences, whereas the other attendees may have only a single encounter with RJ. As such, conference facilitators play an important role in guiding the flow of the conference. Through multiple occurrences of conference dynamics, ideally a facilitator will learn to navigate through the process in order to "repair the harm" most effectively. Although the goal remains the same, the route may be unique. Each facilitator can develop his or her own style, being directive or passive as they may deem necessary.

There are two (psychological) mechanisms by which RJ conference facilitators can navigate the process: procedural justice and legitimacy. Procedural justice (Tyler, 1990) hypothesizes that people comply with the law because the governed accept the authority of the government as legitimate and submit to it rules. Legitimacy reflects the belief that the government has the just and proper authority to enact laws regulating the behavior of the citizens. Fair processes are defined as those in which citizens are treated respectfully, given a voice in the process and outcome, and in which corrections can be made and issues addressed which in turn invoke a belief in the legitimacy of the

sanctioning body. Tyler theorized that such fairness increased compliance with the law. When justice is procedurally fair, Tyler argues, people are likely to obey the law (more so than if only the outcome is viewed as fair). Several studies have demonstrated that RJ processes are often viewed by offenders as procedurally fair, thus increasing feelings of legitimacy (Strang, Barnes, Braithwaite & Sherman (1999); Hipple (2003); Tyler, Sherman, Strang, Barnes & Woods (2007)).

Legitimacy is a central concept in defiance theory. Sherman (1993) argues sanctions can have different effects depending on the frame of reference of the sanction target (i.e. the offender). Sanctions provoke criminal behavior "to the extent that offenders experience sanctioning conduct as illegitimate, that offenders have weak bonds to the sanctioning agent and community, and that offenders deny their shame and become proud of their isolation from the community" (Sherman, 1993, p. 448). Based on the preceding statement, it is clear why restorative justice should reduce "defiant" reactions. Through reintegrative shame (Braithwaite, 1989) that seeks to strengthen one's bonds with others, and procedural justice instilling a sense of legitimacy in the sanctioning agent and community, participation in RJ should, in theory, be more likely to reduce further crime than conventional justice processing.

Conference facilitators who conduct more conferences should obtain better results in terms of crime reduction through enhanced perceptions of procedural justice and legitimacy than those conducting fewer conferences. However, this hypothesis is too simplistic and requires a mechanism. There are three potential explanations for this hypothesized relationship: "practice makes perfect," "initial expertise or trait", and "skill maintenance," that we have borrowed from the medical literature.

Mechanisms of Expertise:

There are at least three mechanisms by which one can obtain and maintain expertise. One can practice a technique to hone one's skill; however one must also continually use the skill in order to maintain it[4]. Some skills are obtained quickly, while others require constant practice in order to maintain expertise. Practice and skill-maintenance are mediated by natural ability. Natural ability for some skills is not evenly distributed in the population, thus some are better than others in almost any skill, technique, or ability one can name. The following argument on RJ conference facilitator expertise is informed by research in the medical field, as similar studies in the social sciences are rare.

In 1979 Luft, Bunker, and Enthoven examined the mortality rates of 12 surgical procedures and determined there was a relationship between the hospital (institutional) volume and mortality, which varied by procedure. For example, they observed decreasing mortality with increased volume for open-heart, vascular, and coronary bypass surgery, supporting a "practice makes perfect"[5] argument. The complexity of the procedure drove the practice makes perfect result. Those surgeons with more opportunity to practice the technique improved their skill and thus reduced mortality. Analogous to surgeons learning new techniques, RJ conference facilitators are initially inexperienced in the principles of RJ and the dynamics of RJ conferences. It would be natural to assume the ability to facilitate RJ conferences would improve over time.

---

[4] E.g. learning a foreign language. Once obtained, one must continually practice or the ability to produce or comprehend the foreign language degrades over time.

[5] Luft, Hunt, and Maerki (1987) examined competing hypotheses to explain the volume-outcome relationship. They examined whether physicians and hospitals improved in ability due to the volume of cases ("practice makes perfect") versus "selective referral," whereby physicians and hospitals with known competence for certain procedures drew more cases of that type. They conclude both explanations existed in concert and the relative importance of each was related to the type of procedure.

Arguably there are complex interrelationships within an RJ conference, which would lend itself to better outcomes through increased facilitator practice. Devereaux et al. (2005) note that all procedures involve a learning curve, and mastery of a skill, procedure, etc. will depend on its complexity. To date, the number of conducted conferences required to achieve competence as a facilitator is an open question. If a "practice makes perfect" argument is correct, procedural justice and legitimacy measures should increase as time (and conferences) progresses.

All skill proficiency will degrade over time without continual practice. The need or importance of refreshing one's abilities will likely vary. Luft et al (1987) note "[a] more likely aspect of experience is the deterioration of skills with lack of practice. In comparing two equally trained surgical residents, the one who consistently performs many of the procedures in question will maintain - or continue to improve - his or her skills, while the one who performs few such procedures will become progressively less proficient" (p. 160). If there is a particular volume of cases (i.e., learning curve) associated with expertise, a facilitator who conducts the requisite number in a shorter timeframe would arguably be more adept, to an extent, [6] than one who took longer to achieve the requisite number. In addition, if one delivered conferences skillfully, it would stand to reason those skills should be maintained through practice.[7] Practice or maintenance of a skill is likely dependent on the initial ability of the individual.

---

[6] Regarding RJ conferences, this comes with a caveat: Skill maintenance, in my opinion, would require reflection and feedback (either self-directed or elicited from others) in order to be effective. A conference facilitator who stacks his or her conferences within the same day or a rapid succession of days would not gain the same degree of competence as one who had time to reflect between conferences over the course of a few days or weeks.

[7] As a caveat, it is possible to have too much practice, especially in a RJ context. RJ is an inherently emotional process, and it is possible too many conferences within a relatively short period of time could result in "burn out." This paper can examine the possibility of a curvilinear relationship between skill-

Some people are generally better at a skill than others. Luft, Hunt, and Maerki (1987) found that the relationship between hospital volume and mortality for various procedures was consistent with selective referral (i.e. expertise). Patient outcomes were better because certain hospitals had more expert staff. Such expertise would draw similar cases, thus increasing the volume of surgeries performed. In a similar vein, although conference attendees cannot select their facilitator, it could be the case that those who conduct more conferences are generally more skillful.

Sherman et al. (2003) sought to test the assumption that mastery over RJ conferencing would produce better outcomes. They compared individual facilitators on the basis of their averaged procedural justice and legitimacy measures across all conferences from the beginning of their "conference career" to the end. Thus, if the volume of conducted conferences was associated with better outcomes, this would suggest that better facilitators conduct more conferences. Alternatively, conference facilitators who possess better initial skill should conduct procedurally just conferences from the outset.

The preceding explanations are not necessarily mutually exclusive. Expertise, whether present *a priori* or learned, will benefit from practice. Likewise, a facilitator who "learns" his or her "craft" should be able to achieve comparable competence to someone with pre-existing ability. The Reintegrative Shaming Experiments (RISE) conducted in the Australian Capital Territory (ACT), Australia from 1995 to 2000 provide an excellent source to test whether facilitator competence is associated with procedural fairness, legitimacy, and ultimately crime reduction and to examine the route to competency.

---

maintenance and outcomes. However, we do not have the ability to address reasons why police officers cease their facilitator "career."

**RESEARCH QUESTIONS**

- Do facilitators with more experience generate more perceptions of procedural justice and legitimacy?
- Do facilitators perform better in procedural justice and legitimacy generation as their experience increases (i.e., as they accumulate more conferences)?
- What is the relationship between time-between-conferences and procedural justice and legitimacy?
- Can facilitators consistently generate feelings of procedural justice and legitimacy or does it change over time?
- Does greater procedural justice and legitimacy engendered by conference facilitators correlate with less recidivism?

<u>The Reintegrative Shaming Experiments:</u>

The RISE experiments represent the largest randomized field trial in criminology in Australia. RISE consisted of four different experiments: adult drink-driving, juvenile shoplifting (under 18), juvenile personal property (under 18), and youth violence (under 30). The present study will focus on the Juvenile/Youth experiments. According to Knight (1999), when a criminal suspect is in custody of the police, and, during interview, fully admits to the offense, an Australian Federal Police (AFP) officer has essentially three options:

- Formal caution
- Put the offender before the court
- Refer the offender to the Diversionary Conference Team

With the cooperation of the AFP and during the study period from 1995 to 2000, an officer had a fourth option of referring the case to RISE. RISE researchers subjected the treatment decision to random assignment. Case referral was left to the discretion of the AFP officer. The matter had to be serious enough to warrant a court appearance, yet not so serious that the officer would have reservations about sending the case to a diversionary conference (DC). In addition, the offender had to agree to attend the conference and the AFP officer had to agree to abide by the random assignment decision. Exactly 173 juvenile property, 113 shoplifting, and 100 youth violence cases, representing 249, 143, and 121 offenders, respectively, were randomly assigned to either diversionary conference or typical court processing.

Diversionary Conferencing in the ACT

As described in Sherman et al (2003), diversionary conferencing in the ACT was meant to be a generalized strategy for police work, that is, all police officers should have been ready to do it. However, over the course of the experiment, of more than one-hundred AFP police officers (out of a force of over six-hundred) who conducted at least one RISE conference, only two continued doing diversionary conferences until the completion of the random assignment period (one of whom was the only conference facilitator to accumulate over 50 known conferences over the duration).[8,9] Figure 1.1 shows the number of conference facilitators per quarter. The number of facilitators

---

[8] The current DC team consists of 2 officers.

[9] This situation could possibly reflect a flaw in RISE conference delivery. Devereaux et al (2005) called for expertise in procedure delivery (in the context of medicine). They believed that surgeons with different levels of competence in performing a new procedure would bias results. In circumstances where competence would be achieved by some and not others, they proposed random assignment to "new" techniques performed by experts. Lim (2005) countered their argument, saying that randomized controlled trials (RCTs) should be balanced on experience levels in order to reflect who will ultimately perform the surgery. In the context of RISE, although in the end conferences were performed by "experts," this was not the original intent of the program.

dropped after a high of 42 conference facilitators in the first quarter of 1996 to two in the final months of the experiments. In addition, as time passed the average number of conferences conducted per facilitator increased (see Figure 1.2). This reflects the loss of facilitators with relatively few conferences and the retention of a select few who had the opportunity to complete more. Although facilitators were not randomly assigned, the number of facilitators offers the chance to examine the relationship between conference expertise and outcomes.

## RESEARCH DESIGN

### Sample and Measures

The sample consists of subjects assigned to RJ where the facilitator was known. Of the 513 total offenders, 266 of them were randomly assigned to participate in RJ. We know the facilitator of 73 percent (193 offenders) of them.[10] In addition, 194 RJ assigned offenders were interviewed at Year 0.[11] The intersection between conferences where the facilitator was known and the offender was interviewed yields a final sample of 159 offenders.

Data used in this study come from a variety of sources. Information on conference facilitation is derived from RISE administrative data. Process and outcome measures come from an initial interview of offenders a few weeks after treatment. The responses are limited to those offenders who both agreed to an interview and where the facilitator

---

[10] Cases in which the facilitator was unknown could arise when RISE observers were not notified of the conference or were not otherwise available to observe. Every attempt was made to ascertain the identity of the facilitator utilizing AFP Diversionary Conference team data. Most of the "unknown" facilitators consist of conferences where one of the two final DC team members conducted the conference. However, for accuracy it is unclear which of the two actually conducted the conference.

[11] Year 0 refers to the interview conducted as soon as possible after the completed assigned treatment. RISE researchers made every attempt to interview offenders within a few weeks of the final treatment event.

was known. Finally, criminal history data come from the AFP PROMIS (Police Real-time Online Management Information System) database. Criminal behavior is limited to crimes against a personal victim, as distinct from a corporate or public victim. For a program designed to repair the harm caused to personal victims of crime such "victimed" offenses clearly stand out as the most relevant. Using administrative data the following variables were constructed:

- Sequential number of conferences conducted (Practice-makes-perfect test)
- Cumulative number of conferences conducted (Expertise)
- Time (in weeks) between conferences (Skill maintenance test)

As soon as possible after treatment, offenders were interviewed about their experience. Among the battery of questions, there were items tapping the domains of procedural justice and legitimacy. Procedural justice is measured by a general scale and subscales examining different facets of the concept. Legitimacy is a single scale measure, based on the responses to three survey items. Cronbach's alpha is reported for each scale to indicate reliability. As a rule of thumb, alpha scores larger than are taken as an indication that the scale responses reliable.

<u>Scales</u>

*Procedural Justice* ($\alpha$ = .917):

- *Awareness of Process*
  - Understood what was going on at treatment
  - Understood what my rights were
- *Consistency, Fairness*
  - Treatment overall was fair

  - Treatment respected my rights
  - Offenders with same offense treated the same
  - Police were fair leading up to treatment
  - Police were fair at treatment
  - Police in Canberra enforce the law fairly
- *Correctability*
  - If police had facts wrong, able to correct
  - If police treated me unfairly, able to complain
  - Felt too intimidated at treatment to speak
- *Control*
  - Felt I had some control over the outcome
  - Had an opportunity to express my views
  - Had enough control over the way things ran
  - Treatment took account of what I said
  - Felt pushed around by others with power
  - Felt pushed into things I didn’t agree with
- *Impartiality*
  - Felt treated no better or worse than others
  - All sides had a fair chance to present views
  - Felt disadvantaged by age, income, sex, etc.
- *Ethicality*
  - Felt I could trust the police during treatment
  - I was made to confess to things I did not do

- *Respect*
  - People were polite to me at treatment
  - I was treated with respect at treatment
  - Police were rude when I was apprehended

The preceding item sets are summed to form a scale,[12] but are also treated individually to examine the extent to which experience may impact individual elements of procedural justice.

*Legitimacy* ($\alpha$ = .788):

- As a result of the conference, you have increased respect for the justice system
- As a result of the conference, you have increased respect for the law
- As a result of the conference, you have increased respect for the police

The descriptive statistics for the preceding variables are presented in Table 1.1. All scale and subscale means are positive, indicating general presence of the construct (at least in the eyes of the interview respondent). However, given that approximately 20 percent of respondents either generally disagreed or were neutral in response to questions on procedural justice, while nearly a third of respondents generally had reduced respect for various aspects of the CJS, we can assume general perceptions of procedural justice and legitimacy cannot be taken as given. We can further assume that there are RJ conference facilitators who were ineffective at altering the offenders' perceptions or were prone to having bad days.

---

[12] Items originally followed a Likert scale generally ranging from 1 to 5, "strongly disagree" to "strongly agree," respectively. The items were recoded with "disagreement" with the statement recoded with negative numbers and "agreement" coded positively. This recoding reflects a more intuitive sense of the scales. "Agreement" with the question would reflect the presence of the concept, while "disagreement" would indicate its absence.

## RESULTS

### Questions 1 & 2 - Practice makes perfect vs. Expertise: Procedural Justice & Legitimacy

Restorative justice, as a practice, should provide a forum for a more procedurally just process. Offenders have the opportunity to voice their opinions, have others treat them fairly and with respect, and have control over both the process and the outcome. Figures 1.3 through 1.10 examine the relationship between procedural justice (summed total and subscales) and facilitator experience, measured sequentially or as facilitators' cumulative conducted conferences. Lines represent the "learning curve," i.e. the change in the (standardized) measure of procedural justice as facilitators move from their first, second, etc. conference.[13] Dots represent the cumulative total experience of a facilitator or group of facilitators.[14,15]

In general, we can see "average" performance across both the accumulation of conference experience and the maximum number of conferences.[16] The process of gaining experience does not show a clear overall pattern, as it relates to procedural justice. This result is repeated when we look at procedural justice as whole and individual

---

[13] Conference experience is based on the total N of conferences conducted. There are facilitators who only conducted either drink-driving or juvenile/youth conferences, with a portion of all facilitators partaking in both. It would not be a fair comparison if we compared two facilitators on their fourth conference, for example, if one facilitator had completed an additional six drink-driving conferences, thus having ten completions. Although the dynamics of drink-driving conferences are different (see Strang et al. (1999) for details), nonetheless these conferences provide an opportunity for facilitators to gain experience.

[14] Therefore, a dot at 4 conferences would represent those facilitators who conducted four total conferences. Their contribution to the line on the figure would, at best, represent the scores from the first through fourth conferences.

[15] We note that a single conference facilitator has 51 total conferences. Conference frequencies above 25 represent the history of that facilitator's conference "career," in all subsequent practice-makes-perfect and initial expertise figures.

[16] Figures become more "noisy" towards the right side of the chart. As the number of conferences increases, the number of facilitators at those levels decreases.

subscales. We see undulation from above- to below-average procedural justice scores. An escalating pattern would indicate learning, while a declining pattern would possibly indicate fatigue. The lack of such patterns here would indicate that the facilitator's ability to achieve a procedurally just process is not a learned skill.

When we examine the influence of "expertise" on procedural justice, we see a similar pattern. Facilitators' total experience with conferencing does not indicate proficiency in delivering a procedurally just conference, with some notable exceptions. It appears that there is something different in those facilitators with 5, 10, and 11 total conferences. However, given the small size of these groups, formal statistical analysis is not possible. It seems to matter less how many conferences are performed, and may depend on individual, rather than group, characteristics. Some people may be better than others in conference delivery,[17] at least as it relates to delivering a procedurally just process. The sequence of conference delivery produced an almost serpentine pattern around the mean. Likewise, there was no pattern relating cumulative conference experience to procedural justice, i.e. there was nothing that distinguished ability with accumulation.[18] We next turn to the relationship between conference experience and legitimacy.

Restorative justice, as a generally procedurally just process, should provide an opportunity to enhance the legitimacy of the criminal justice system (CJS) in general, and the police in particular. Police officers with greater conference experience have had more

---

[17] It is unfortunate that we are unable to determine the reasons for facilitator drop-out. Of those facilitators with better than average scores, some could have "burned out," while others could have been reassigned (it is relatively common for AFP officers to serve on U.N. peacekeeping missions; e.g., East Timor.

[18] Ideally we would hope those who were proficient would "stick with it.". However, this is almost certainly not the case. Facilitators, regardless of ability, are going to end their conference careers at some point.

opportunities to enhance offenders' perceptions of the CJS. In a similar vein, a facilitator's ability to favorably represent the CJS could improve over time. In Figure 1.11 we examine these possibilities.

Similar to the results for procedural justice, it is apparent (visually) that there is no relationship between "learning" and perceived legitimacy. The accumulation of conferences is not related to an increase in perceived legitimacy. On the contrary, we see a similar pattern undulating above and below the mean, becoming more "noisy" as the sample size decreases at greater levels of experience. We see a similar pattern when we examine the overall experience. Facilitators with more overall experience are not consistently better (or worse) than facilitators with less experience.

In this test, instilling a sense of legitimacy is not something learned through practice, nor indicated by how many conferences one has conducted. Facilitators, as a whole, do not improve over time. Likewise, those who perform more conferences overall, do not enhance perceptions of legitimacy.[19] Although groups of facilitators consistently perform poorly, their experience with conferencing may matter less than their overall ability to instill these perceptions. It is possible the mechanism at work is the time between "practice" sessions. We will examine this in the next section.

**Question 3 - Impact of Skill Maintenance on Procedural Justice & Legitimacy**

Conferences are difficult to arrange at the best of times. The conference facilitator has to coordinate with offenders and victims, both of whom need to coordinate with their own supporters (at least within the context of RISE). On average, during the RISE

---

[19] A possible flaw in these analyses is the lack of baseline data on legitimacy. Thus those who did not change their opinion of the CJS, in general, may have already had a high opinion of the system. However, the ability to avoid reducing respect for various elements of the CJS is still important.

experimental period, it took 2 to 3 months to organize and successfully deliver a restorative justice conference.[20] As Luft et al. (1987) point out; successful outcomes can be related to the frequency with which one gets to practice. Conference facilitators who are able to organize successive conferences more quickly should have more opportunity to practice and maintain their skills than those who conduct conferences relatively rarely. However, stacking conferences (i.e. conducting conferences on the same day) could have a negative impact. We argue that clustering conferences within a relatively short period of time does not give a facilitator the opportunity to reflect on the conference and hone his or her skills.

In Figures 1.12 through 1.19 we examine the relationship between the time between successive conferences and procedural justice.[21] Visually no consistent pattern between the weeks between successive conference and procedural justice is revealed, on either the total scale or the individual subscales. It would appear, on average, those who quickly arrange and complete conferences do no better or worse than those with a longer timeframe between successive conferences. This could reflect the fact that the time between conferences is often out of the control of the facilitator,[22] or suggest that the ability to conduct a procedurally just process is not a skill that needs to be maintained.

We see a similar pattern in the relationship between legitimacy and the time between successive conferences (see Figure 1.20). The ability to enhance perceptions of legitimacy is not related to the time it takes to organize a conference. Similar to the

---

[20] Although the reasons for the "time to treatment" are often outside the facilitator's control, the time between successive conferences remains relevant as it affords one the opportunity to maintain one's skills.
[21] The average score obtained in relation to the initial conference is provided as a reference.
[22] Recall that a facilitator must coordinate the offender and victim to arrange a time amenable to everyone. In addition, offenders and victims must also coordinate with their supporters.

results for procedural justice, enhancing perceptions of legitimacy does not seem to be a skill that needs to be maintained.

It is apparent from the preceding paragraphs that the ability to conduct a procedurally fair process, arguably enhancing perceptions of legitimacy, is not related to the time between successive conferences. “Stacking” conferences does not incur a penalty, nor does spacing them out over a longer timeframe. Although some facilitators may organize conferences more or less efficiently than others, often this is out of their control. As such, evoking these two psychological mechanisms, procedural justice and legitimacy, may rely less on practice and more on initial ability.

**Question 4 - Procedural Justice & Legitimacy: Consistency or Development**

How consistently are procedural justice and legitimacy experienced in RJ conferences? Are those individuals who (overall) rank high on these factors consistently on top over their conference careers? Likewise, do some groups of facilitators improve or degrade over time? To answer the preceding questions, we defined four groups based on their quartile score on procedural justice & legitimacy, forming groups defined as “low,” “low-middle,” “high-middle” and “high” (see Table 1.2). Arguably, the two extreme quartiles should show the most stability over their conference careers, while the middle 50 percent could improve, degrade, or maintain their performance over time. Figure 1.21[23] demonstrates the relationship between (total) procedural justice and sequential conferences among different strata of conference facilitators.[24] Those facilitators in the top 25percent are almost always above the mean procedural justice score (as are the

---

[23] Gaps in the lines represent moments where the facilitators conducted drink-driving conferences.
[24] Conference frequency is truncated for ease of interpretation.

individuals in the third quartile), while those in the bottom 25 percent are consistently below the mean. The second quartile appears to be the most variable. However, it would appear that the ability to deliver an above average procedurally just restorative justice conference can be distinguished after approximately five conferences, in these data. This may be important for policy. If the goal is to make conferencing as procedurally just as possible, we could funnel facilitators who show proficiency at an early stage. Alternatively, those who fall short can be funneled out.

The pattern of perceptions of legitimacy is less obvious. Although the overall top 25 percent is consistently above the mean (and the bottom 25 percent consistently below), the ability to enhance legitimacy of the middle 50 percent is highly variable (see Figure 1.22). There are moments during the facilitator career where the third quartile group (on average) is the top performer. Although the bottom performers, in terms of legitimacy, drop out early, the top 25 percent of facilitators drop out relatively early too. Although we do not know baseline levels of CJS legitimacy in this sample, it is clear that those facilitators who effectively increase perceptions of legitimacy drop out after conducting a few conferences.

Facilitators differ in their ability to deliver a procedurally fair RJ conference and instill a sense of legitimacy in the CJS. This ability appears to come from within the facilitators. Those facilitators at the extremes of the distribution for either procedural justice or legitimacy (based on their overall score) maintained their performance across their conference career, in general. We do not see patterns of skill degradation or improvement over successive conferences. In the end, facilitators either “have it” or they do not. However, we should also examine whether there is a relationship between

procedural justice, legitimacy, and crimes against victims, at least within the context of conference facilitation.

**Question 5 - Impact of Procedural Justice & Legitimacy on Offending**

Tyler (1990) posited that (procedurally) fair processes would increase compliance with the law. The experience with "fair procedures leads offenders to view the law and legal authorities as legitimate, leading to enhanced commitment to obey the law" (Tyler et al., 2007, p.556). We would expect offenders who participated in RJ conferences with facilitators with a track record for enhancing perceptions of legitimacy or delivering procedurally just conferences should, on average, show net reductions in offending.

Based on offender interviews, we have grouped conference facilitators into quartile groups based on their overall procedural justice score across all conferences. Table 1.3 provides the change in "victimed" offending one year prior to random assignment to one year after. In general, the results are consistent with predictions, with one notable exception. On average, those offenders with RJ facilitators in the bottom 25 percent for procedural justice experienced net increases in offending, a result consistent with Tyler's theory. Given offenders could not choose their facilitator (and vice versa), this result is less likely to be due to selection.[25]

At first blush, contrary to prediction, those facilitators who show the most skill in delivering procedurally just RJ conferences overall have offenders who increase their offending. A closer examination of the data reveals this result is due to an outlier. There

[25] Examination of prior "victimed" offending yields similar frequencies for each quartile.

is another outlier influencing the offending change for facilitators in the second quartile.[26] When their offending is censored to the next highest increase in offending, there is an overall decrease in offending for the fourth quartile and the magnitude of the change nearly doubles for the second quartile. It would appear that the impact of procedural justice is similar to the effect of Vitamin C. On average, one does not need a lot to produce a beneficial result.[27]

We see a similar picture for the influence of legitimacy and offending in Table 1.4. Offenders with facilitators in the bottom 25 percent for legitimacy experienced net increases in offending. Although there is a net increase for the third quartile, examination of the data reveals both outliers had an RJ conference delivered by facilitators in this group. Accounting for their extreme scores, overall there is a net decrease. In general, those offenders who increased their respect for the CJS reduced their offending in the year following random assignment.

Instilling increased respect in the justice system or experiencing a procedurally just RJ conference results in reduced offending. When these features are absent, or below a certain average threshold, there is an increase. In the earlier analyses we discovered the ability to influence perceptions of procedural justice and legitimacy was not tied to the level of facilitator experience or the consistency of practice. We would argue these abilities are evident prior to conducting conferences. Some facilitators have it and some do not. Rather than identify conference facilitators who are exceptionally good (in terms

---

[26] These offenders are both Aboriginal and recent evidence suggests RJ conferencing may not be suitable for Aboriginal offenders (Sherman & Strang, 2007).

[27] Absent the Aboriginal experience, the net change in offending is similar for each of the upper three quartiles.

of procedural justice and legitimacy), we should identify those who are exceptionally bad, at least on average.

## DISCUSSION/LIMITATIONS/POLICY RECOMMENDATIONS

We examined the impact of facilitator experience in conference delivery on perceptions of procedural justice and legitimacy among offenders. Ultimately we examined how these perceptions were related to changes in offending. Diversionary conferencing, at least in the Australian Capital Territory, follows a basic initial template (Knight, 1999), but diverges from a planned route as conference participants interact with one another. The actors (offender(s), victim(s), and supporters) change from conference to conference. A conference facilitator can be as directive or "hands off" as he or she chooses. Because the RJ process allows both victims and offenders to contribute to the process and outcome within an environment of respect and fairness, RJ is viewed by both offenders (and victims) as procedurally just. Given this experience relies on the participation of a police officer (as a representative of the CJS), RJ as a procedurally just process generally increases the perceived legitimacy of the CJS within offenders. It was the variability in procedural justice and legitimacy in the RISE data that led to the present study.

The ability to deliver a procedurally just conference and enhance feelings of legitimacy was hypothesized to be based on a facilitators' experience with conferencing. Based on the medical literature, we hypothesized three routes to conferencing competence: initial expertise; learning; and skill maintenance. We hypothesized effective conference facilitators would perform the most conferences. Those facilitators who (overall) conducted more conferences would generate more perceptions of procedural

justice and legitimacy than those with less experience. Alternatively, we allowed for learning. Although some facilitators may be naturally "better" at RJ conference delivery, others may improve their skills over subsequent conferences. Finally, effective conference delivery may be based on how often a facilitator gets to practice. In the end, however, conference facilitation is not like surgery.

The initial expertise hypothesis may be valid, but not evident by the number of conferences "under one's belt." It is reasonably clear that some facilitators are better than others, but the ability to deliver a procedurally fair conference, arguably leading to increased legitimacy, is not based on the number of conferences one has completed. Alternatively, it was clear some facilitators were worse than others. It is possible some facilitators who showed promise dropped out (for whatever the reason), while others who were less capable continued. Unfortunately, we are not able to assess this possibility.

On average, facilitators who "stuck with it" did not gain competence in conference delivery relative to those who dropped out after comparatively few. If anything, perceptions of procedural justice and legitimacy are relatively stable over subsequent conferences. RJ conferences are dynamic processes. Because the actors change, each RJ conference brings new interactions and is less susceptible to routinization. Each conference offers a new experience and may provide fewer chances for perfecting skill at conference delivery. In essence, the conference facilitator has to relearn, or at least adapt to the new situation.

Thus, the novel conference experience may offer a partial explanation for the lack of support for skill maintenance. Given the dynamic nature of conferencing, we expected that conferences conducted closer in time would allow a facilitator more opportunities to

practice. However, the time (in weeks) between conferences did not impact the perceptions of procedural justice or legitimacy, at least from a visual/descriptive point of view. An alternative explanation is the time between subsequent conferences is often out of the facilitator's control. RJ conferences involve multiple people with different schedules and availabilities. Facilitators who can "practice" within a short timeframe may have been lucky. They may have put more time into organizing a conference than others, but it is also possible things just "worked out" for them.

Using a slightly different framework, we grouped facilitators based on their overall percentile score for either procedural justice or legitimacy.[28] We did this in order to ascertain whether perceptions of procedural justice and legitimacy were maintained over time, improved, or degraded. Our analyses demonstrated those in the extreme quartiles maintained their performance across their conference careers. The middle 50 percent were more variable. This variability may be due in part to the fact these facilitators conducted more conferences. We believe this is less likely to be true, because on average the top 25 percent almost always outperformed them and the bottom 25 percent almost always performed worse. Based on these data, the ability to increase respect for the justice system, or produce a procedurally fair conference, appears related to the individual's ability prior to engaging in his or her facilitator "career," rather than "on the job" experience with conference delivery. If the goal was to maximize the potential for procedurally fair conferences or increase respect for the justice system, we could make this determination relatively early.

---

[28] Recall these were analyzed separately because having a high percentile score for one did not necessarily relate to a high percentile score in the other.

Tyler et al. (2007) concluded that procedurally fair processes decreased future drink driving via the mechanism of legitimacy. Utilizing a different strategy, we examined the impact of the same mechanisms on personal victim crimes. After reducing the influence of two extreme cases, we determined the impact of procedural fairness and legitimacy was qualitative rather than quantitative. Provided a facilitator was not in the bottom 25 percent for either procedural justice or legitimacy, we witnessed reductions in personal victim crime from one year before to one year after random assignment. It is more important to have greater prevalence of procedural justice and legitimacy than to have a preponderance of these concepts. Similar to taking Vitamin C, one does not gain a greater benefit by taking more, yet taking some is certainly better than not taking any.

Based on the preceding analyses, we offer the following recommendation. It is apparent that RJ conference facilitation does not require "experts." Essentially any trained ACT police officer has the potential to run a conference well. However, it is clear that low-performers need to be identified and screened out. This is important from not only a crime reduction standpoint, but also a procedural fairness standpoint.

Given the potential links between procedural fairness, legitimacy, and crime, it is important to identify those facilitators who do not possess the skill set necessary to effectively evoke these processes. Alternatively, we need to understand the reasons facilitators end their conference careers. It is possible competent facilitators dropped out early while less competent facilitators continued to conduct RJ conferences. Whether these changes were driven by "burn out," overseas or local redeployment, or mere lack of commitment is an open question. We need to understand why facilitators continue or

leave RJ conferencing in order to recruit and maintain the best people for these positions.[29]

The present study examined the influence of experience on the ability of RJ conference facilitators to conduct procedurally fair conferences and enhance perceptions of legitimacy. Some caveats need to be addressed that may limit the conclusions. We only have data from offenders who agreed to an interview and received an RJ conference where the facilitator was known. Although over 70 percent of the offenders were interviewed, it is likely there were different experiences between those who were interviewed and those who were not. However, to counter this claim, we offer the variability of responses to our questions of interest. We clearly have offenders who had no problem discussing their dissatisfaction with their experience, albeit they were largely in the minority. In most cases, when a conference was delivered, we could ascertain the identity of the conference facilitator. However, in a few cases this was either not possible or we could only narrow the possibility to two facilitators. Finally we have no baseline measures for legitimacy. Among those individuals who did not experience an increase (or decrease) in respect for various aspects of the CJS, we could have either a ceiling or floor effect. That is to say, either their initial impression of various aspects of the CJS was so low it was unlikely to reduce further or so high that they could not demonstrate an increase. Although this criticism is fair, we would argue our results are important to the extent the conference did not lower respect for the CJS. Perception maintenance can be as important, in the sense that the facilitator has not made things worse.

---

[29] Or at the very least, we can keep the worst out.

We used a visual analysis of these data to determine patterns of experience. In the end, we discovered that conference expertise does not seem to be related to the quantity of procedural justice or legitimacy experienced. Procedurally just conferences, ideally leading to perceptions of legitimacy, appear to be related to the facilitators themselves and not the amount of experience, at least in the ACT. These perceptions, in turn, were related to lower offending in general. We suggest a better understanding of who is ineffective along these dimensions. Knowing who cannot consistently deliver a procedurally fair conference is more important than knowing who is most effective. We present a simpler solution for policy makers and program deliverers. Essentially any trained ACT police officer has the potential to lead a conference. This expands the pool of available facilitators. We merely need to monitor conferences in order to determine who is ineffective and should be screened out, while doing our best to retain those who can deliver a procedurally just RJ conference. In addition, future research on RJ should focus on the process in general, and the facilitator specifically, in order to increase our understanding of both “how” and “why” an RJ conference works.

Table 1.1 Descriptive Statistics for Outcome Variables

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Scale of PJ (25) | 159 | -43 | 49 | 20.53 | 14.406 |
| Awareness of the Process | 159 | -4 | 4 | 2.13 | 1.126 |
| Consistency & Fairness | 159 | -8 | 8 | 2.81 | 2.355 |
| Correctability | 159 | -6 | 6 | 1.37 | 2.315 |
| Control | 159 | -10 | 13 | 4.64 | 4.622 |
| Impartiality | 159 | -5 | 6 | 2.28 | 2.029 |
| Ethicality | 159 | -4 | 4 | 1.57 | 1.490 |
| Respect | 159 | -6 | 6 | 2.17 | 2.256 |
| Scale of Legitimacy (3) | 159 | -6 | 6 | 1.10 | 2.629 |

Table 1.2: Cross-tabulation of Average Procedural Justice & Legitimacy Quartile

| | | Quartiles of Legitimacy | | | | Total |
|---|---|---|---|---|---|---|
| | | 1st | 2nd | 3rd | 4th | |
| Quartiles of Procedural Justice | 1st | 20 | 1 | 2 | 0 | 23 |
| | 2nd | 1 | 54 | 32 | 1 | 88 |
| | 3rd | 2 | 4 | 8 | 8 | 22 |
| | 4th | 1 | 2 | 17 | 6 | 26 |
| Total | | 24 | 61 | 59 | 15 | 159 |

Table 1.3: Change in Personal Victim Crimes 1 Year Pre-/Post Random Assignment by Quartile of Average Procedural Justice

| Quartiles of Procedural Justice | Mean | N | Std. Deviation |
|---|---|---|---|
| 1st | .48 | 23 | 1.702 |
| 2nd | -.17[30] | 88 | 1.889 |
| 3rd | -.27 | 22 | .631 |
| 4th | .19[31] | 26 | 3.763 |
| Total | -.03 | 159 | 2.174 |

Table 1.4: Change in Personal Victim Crimes 1 Year Pre-/Post Random Assignment by Quartile of Average Legitimacy

| Quartiles of Legitimacy | Mean | N | Std. Deviation |
|---|---|---|---|
| 1st | .42 | 24 | 1.692 |
| 2nd | -.31 | 61 | 1.148 |
| 3rd | .17[32] | 59 | 3.152 |
| 4th | -.40 | 15 | .910 |
| Total | -.03 | 159 | 2.174 |

[30] The magnitude of the decrease would be larger, accounting for the influence of an outlier.
[31] The increase is due to an outlier. Deletion or censoring of this case reveals net decrease in average offending.
[32] Two facilitators who ranked in the third quartile (overall) for legitimacy conducted conferences with the two outliers. When their offending is censored, there is a net decrease in offending.

Figure 1.1: Number of conference facilitators per quarter

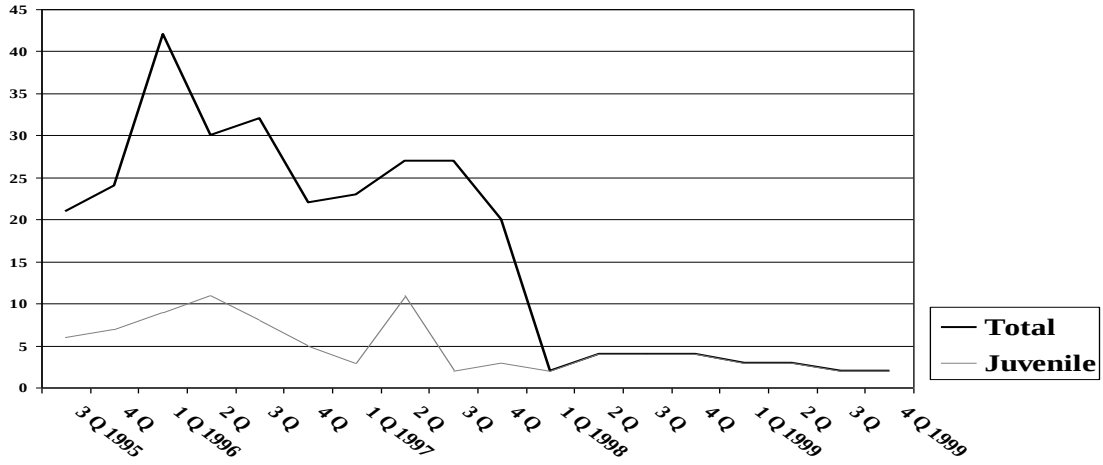



Figure 1.2: Average N of conducted conferences per facilitator

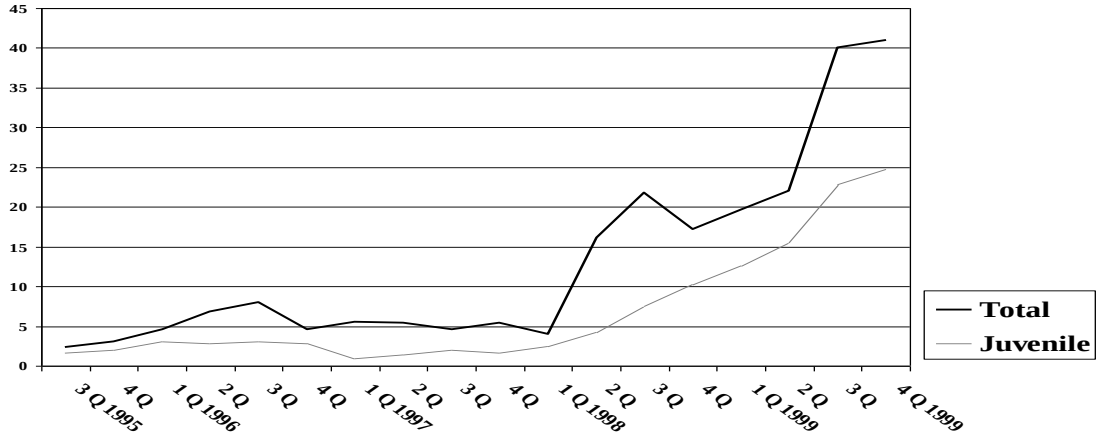


Figure 1.3: Procedural Justice by Type of Facilitator Experience

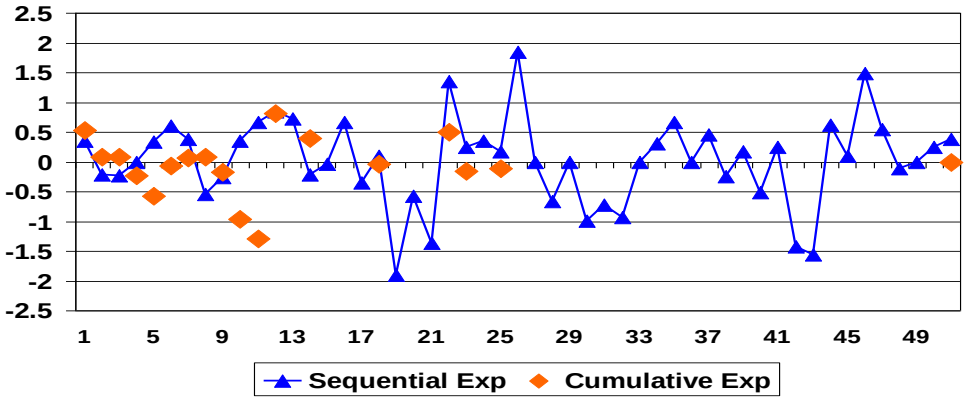



Figure 1.4: Awareness of the Process by Type of Facilitator Experience

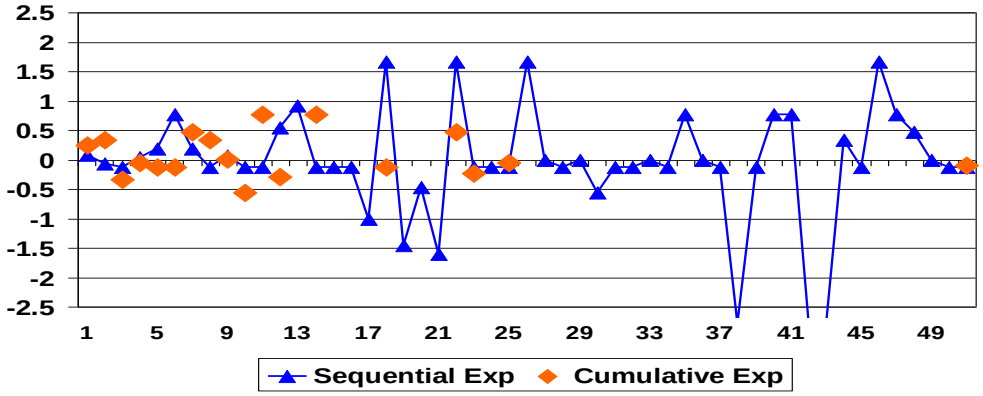



Figure 1.5: Consistency & Fairness by Type of Facilitator Experience

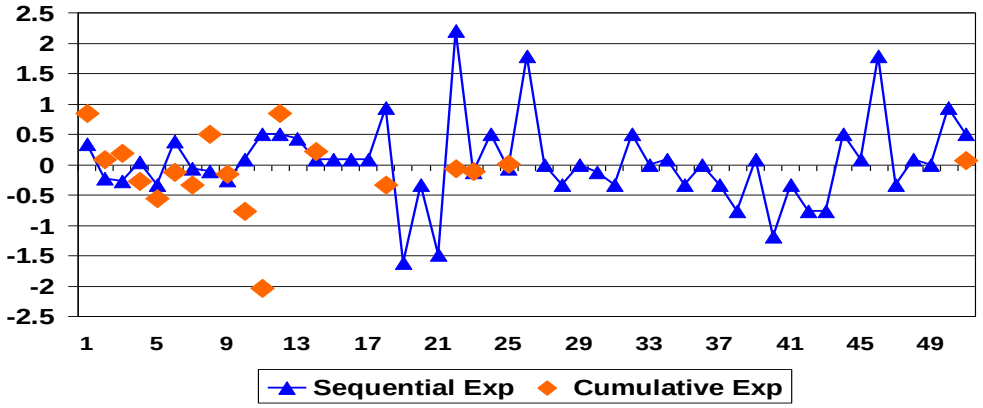


Figure 1.6: Correctability of the Process by Type of Facilitator Experience

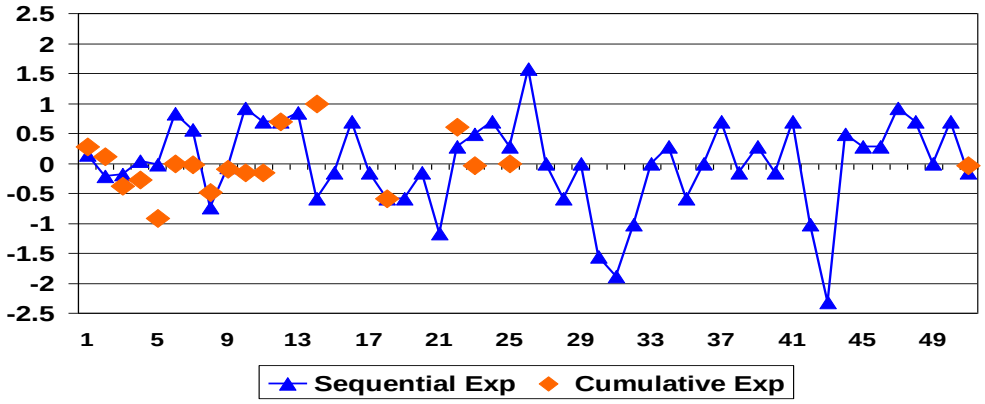



Figure 1.7: Control of the Process by Type of Facilitator Experience

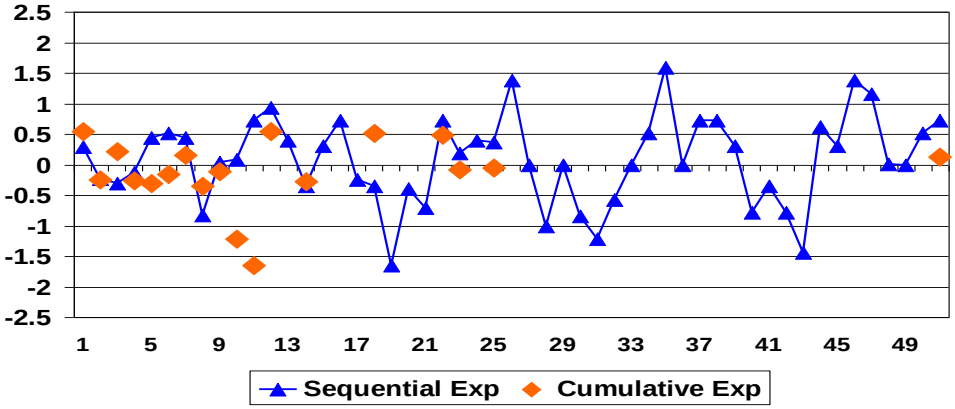



Figure 1.8: Impartiality of Conference Attendees by Type of Facilitator Experience

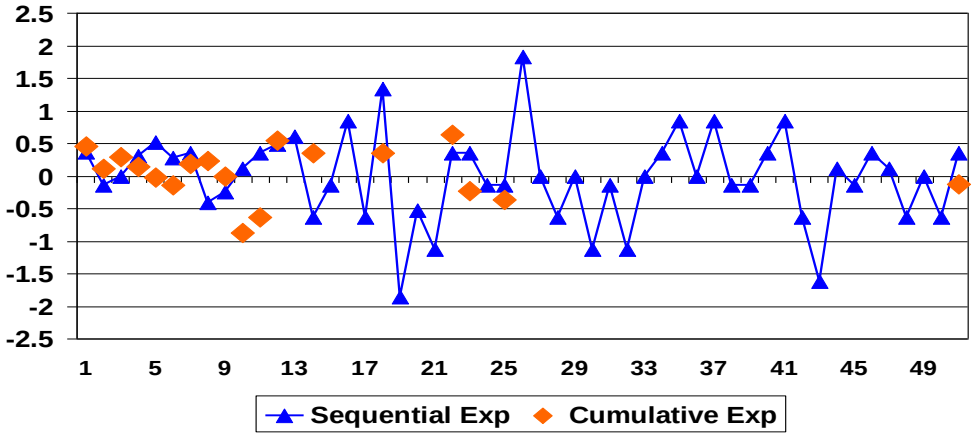


Figure 1.9: Ethicality of Conference Attendees by Type of Facilitator Experience

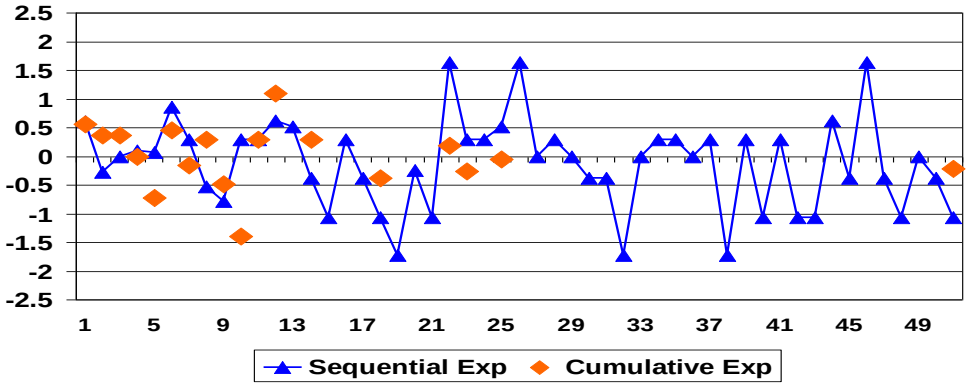



Figure 1.10: Respect by Type of Facilitator Experience

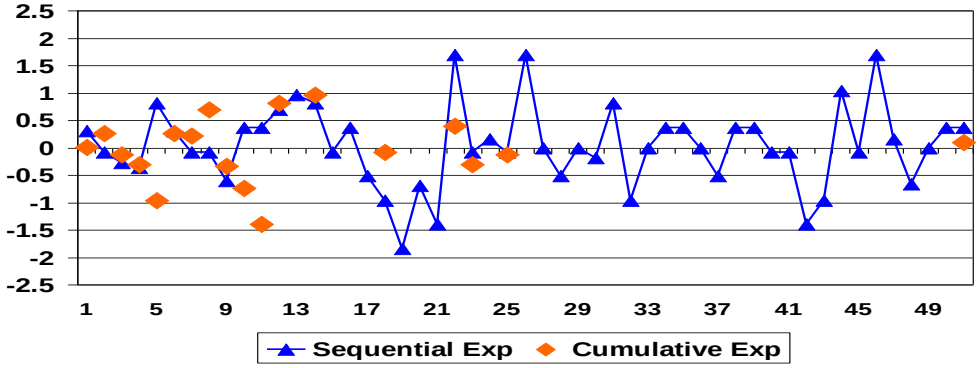



Figure 1.11: Legitimacy by Type of Facilitator Experience

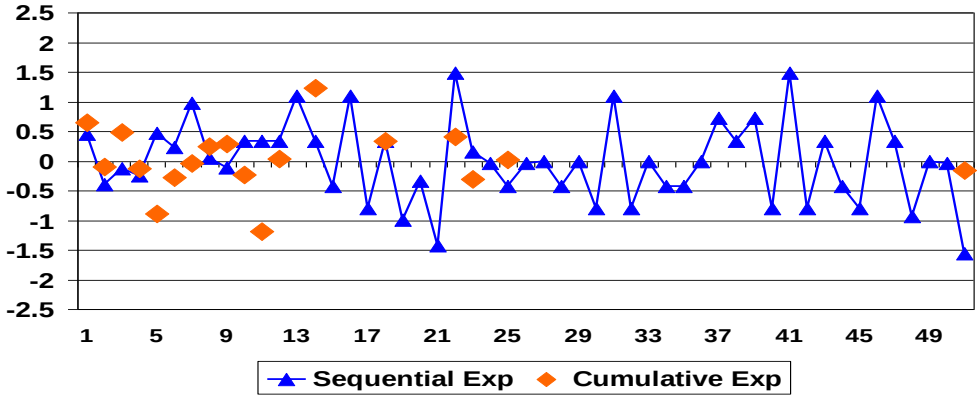


Figure 1.12: Procedural Justice by Weeks between Conferences

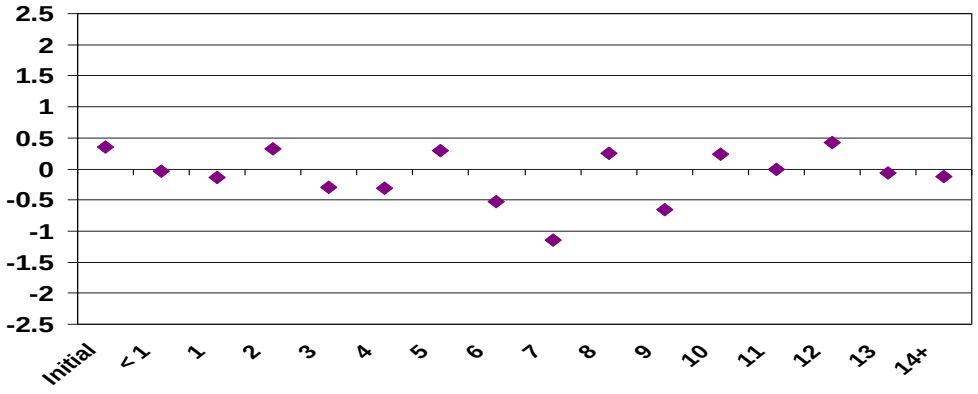



Figure 1.13: Awareness of the Process by Weeks between Conferences

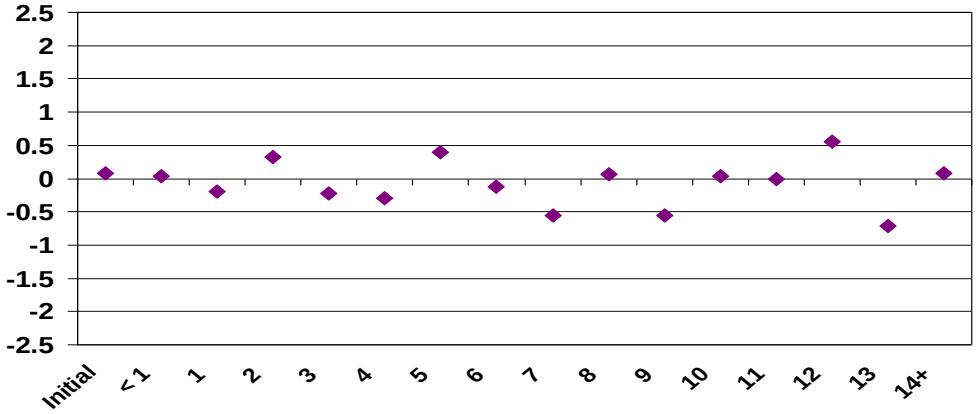



Figure 1.14: Consistency & Fairness by Weeks between Conferences

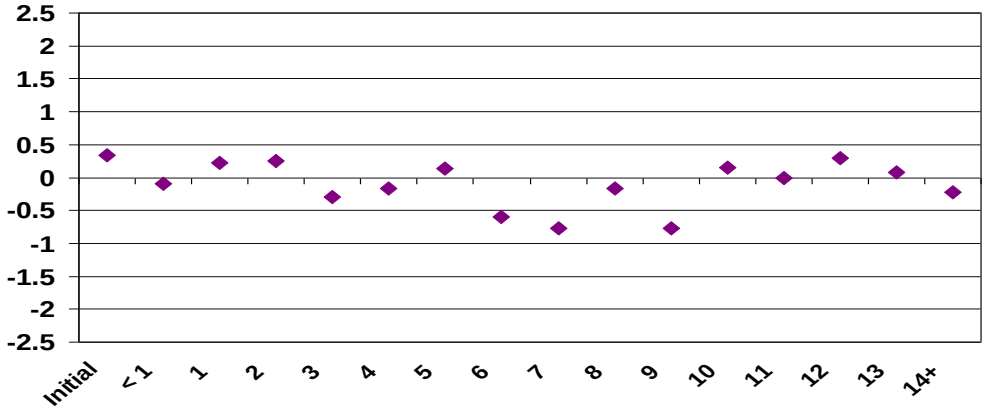


Figure 1.15: Correctability of the Process by Weeks between Conferences

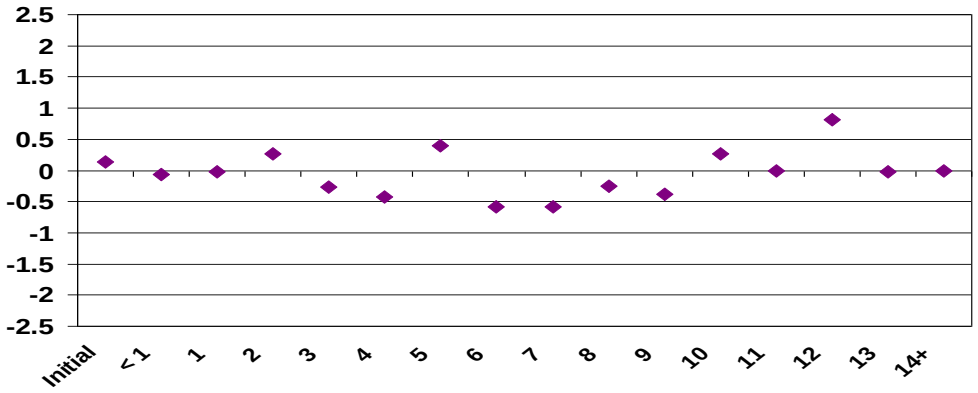



Figure 1.16: Control of the Process by Weeks between Conferences

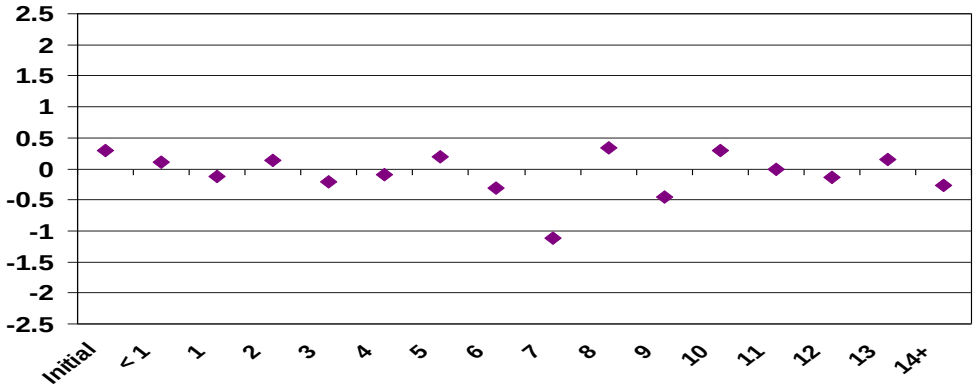



Figure 1.17: Impartiality of Conference Attendees by Weeks between Conferences

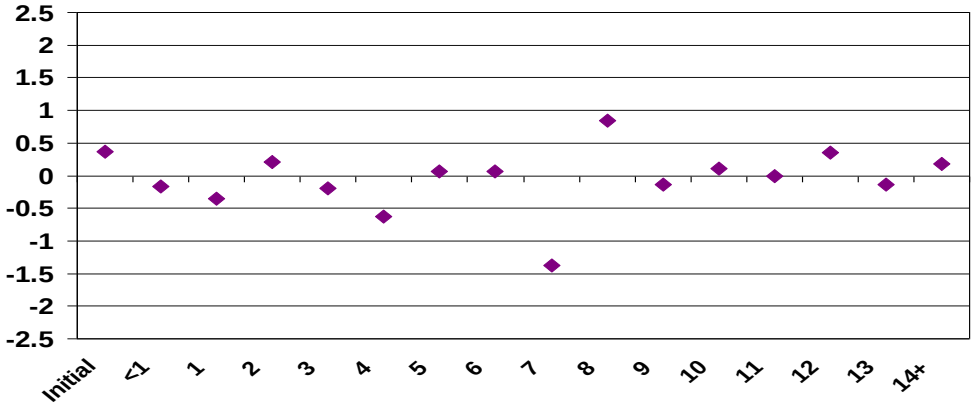


Figure 1.18: Ethicality of Conference Attendees by Weeks between Conferences

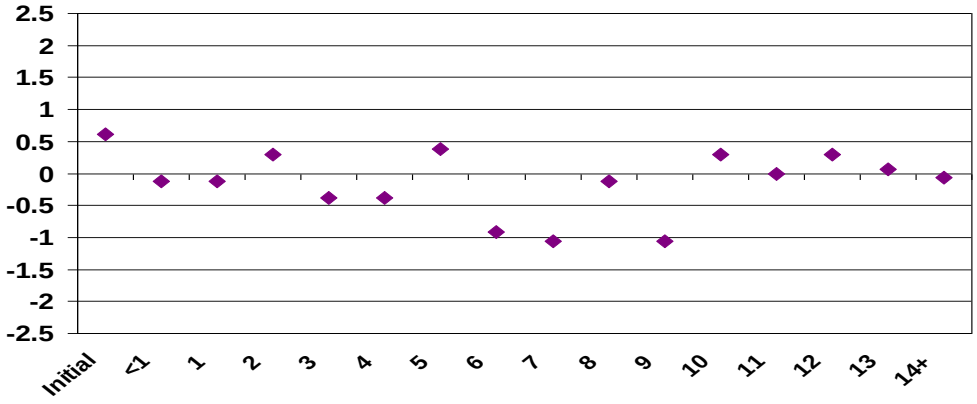



Figure 1.19: Respect by Weeks between Conferences

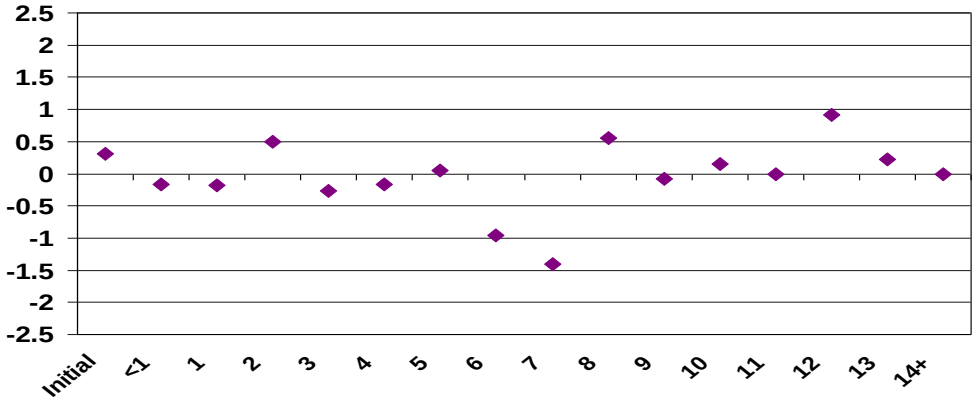



Figure 1.20: Legitimacy by Weeks between Conferences

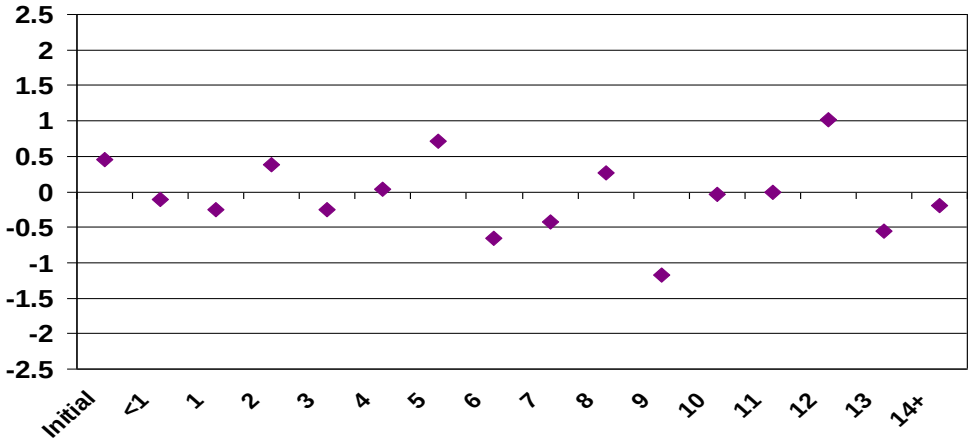


Figure 1.21: The Development of Procedural Justice over Sequential Conferences by Quartile of Average Procedural Justice.

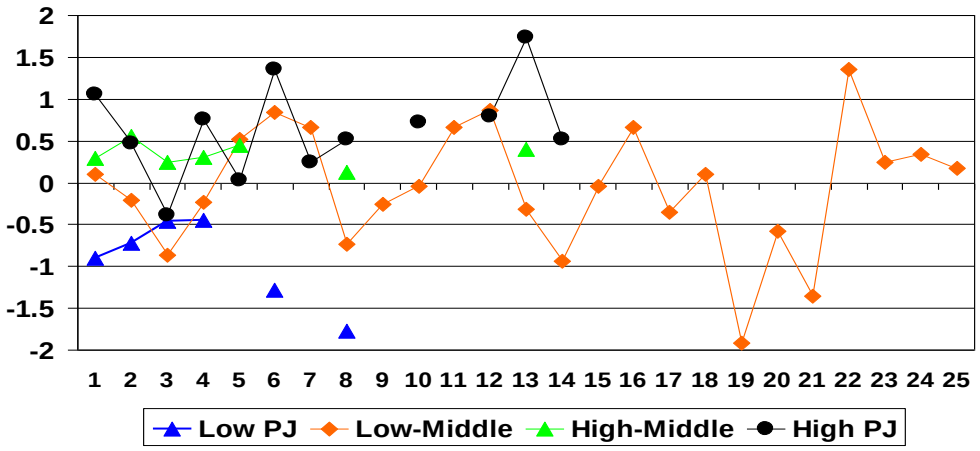
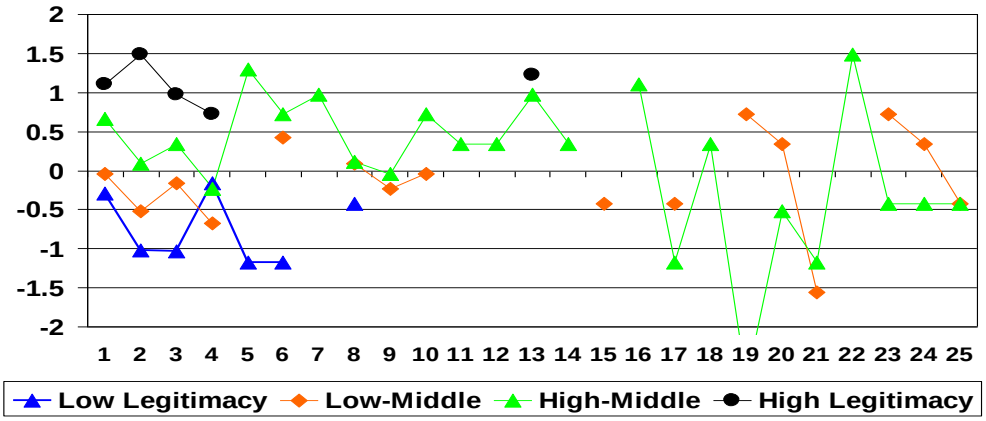



Figure 1.22: The Development of Legitimacy over Sequential Conferences by Quartile of Average Legitimacy



# CHAPTER 2: The Differential Impact of Restorative Justice on Offending Trajectories

## INTRODUCTION:

As a system of justice that focuses on repairing the harm caused by crime (rather than punishing offenders), restorative justice (RJ) offers an improvement over conventional justice, based on retributive models. RJ is a process by which the stakeholders (offender, victim, community, etc.) in a specific offense collectively handle the aftermath of the offense (Marshall, 1999). Offenders and victims, sometimes with supporters of each, meet in the presence of a facilitator to discuss a criminal event and decide what needs to happen to repair the harm caused. As a process it is both offender-focused and victim-focused. Theoretically, RJ provides a forum in which offenders and victims have a voice in the proceedings, unlike court. Generally, evaluations of RJ have focused on criminal behavior as the main outcome. Other outcomes have included procedural justice and legitimacy, and changes in victim quality of life. On paper, RJ programs would appear to be a better use of resources for policy makers.

Despite optimism for RJ, evaluations reveal mixed results for crime reduction (Sherman, Strang, & Woods, 2000; Sherman & Strang, 2007). While all parties (offenders and victims) are largely satisfied with the practice, RJ's effectiveness in crime reduction has been variable. In part, these varied results are based on evaluations of different crime types, social and racial "distances" between offender and victim, or actual

versus "symbolic" victim.[33] Face-to-face meetings have been subjected to more rigorous study. On average, at best RJ reduces crime relative to controls and at worst does equally as well. In some rare cases, RJ has been known to backfire.[34] The question "Does RJ work?" needs to be expanded to "For whom does RJ work and under what circumstances?" The process of RJ is meant to engender psychological principles of empathy among offenders *and* victims, while instilling a sense of reintegrative shame among offenders. If a RJ participant is not able to experience these emotional states, the effectiveness of RJ as a process could be dampened. As such, the ability of RJ practices could depend on the cognitive ability of those involved, which might be related to the age at which an offender participates in a RJ conference.

In addition, the effectiveness of RJ could depend on the developmental trajectory of the offenders involved. The experience of first-time offenders in RJ arguably would be different to the experience of one who historically has greater involvement in the criminal justice system (CJS). First-time offenders may benefit from the emotional intensity evident in RJ conferences, providing that process with greater "teeth" to facilitate desistance from crime. A similar process could operate with experienced offenders, whereby they gain a greater understanding of the harm they have caused, appreciate their voice in the process, and take steps to reduce their criminal behavior. Alternatively, the process could produce a defiant reaction where offending increases.[35] RJ could be viewed as unfair, or an overreaction to what the offender has defined as "no big deal." All in all, the RJ process is likely to produce variable effects depending on which offenders

---

[33] The RISE experiments included a juvenile shoplifting experiment where the security guard typically served as the "victim."
[34] See Sherman, et al. (2006)
[35] Sherman (1993)

participate. Understanding this variability is important for understanding the effects of RJ and what types of offenders it is most likely to benefit.

If we knew for which offenders RJ was likely to have the biggest benefit (or cause harm), using data available at the time of arrest, we could make informed decisions about who we sent down that CJS "stream." Whether RJ works best for first-time offenders, experienced offenders, or even "dabblers" (those offenders who sporadically offend), is an open question. An efficient use of resources would benefit from a greater understanding of how RJ "works" on different people with different criminal histories. To date, evaluations of RJ have looked at aggregate mean differences in offending. Within this aggregation, there may exist groups of offenders who have different offending careers relative to the others. Nagin and Land's (1993) semi-parametric group-based method of modeling developmental trajectories offers a way to disentangle the aggregate effects of RJ. In the present study we will examine whether there are distinct offending groups within the participants. Once established, we will assess the impact RJ has on official criminal offending relative to offenders assigned to traditional court within different clusters of individuals. It is likely RJ, as a more dynamic process, would impact future offending in different ways depending on an offender's developmental history of offending.

Trajectory modeling as a tool

"Group-based trajectory modeling is a specialized application of finite mixture modeling and is designed to identify clusters of individuals following similar progressions of some behavior or outcome over age or time" (Jones & Nagin, 2007, p.

542).[36] The technique is data-driven, and thus, does not rely on *a priori* definitions of offending groups. Trajectory modeling offers an opportunity to test developmental and life course theories derived without the use of this kind of analysis. In addition, the method has been used for exploratory study of the development of phenomena over time in the absence of theory. Of course, this method requires longitudinal data and more data provides richer detail. In the presence of detailed data over time, trajectory modeling has demonstrated utility. Through the use of trajectory modeling, especially in criminology, we now have a greater understanding of when antisocial (and criminal) behavior begins (onset), develops (whether continual or intermittent), and ends (desistance).[37]

Although trajectory modeling provides a useful tool for separating clusters of individuals with similar developmental trends, recent refinements of the method has yielded expanded utility. Specifically, we now can include "turning points" or interventions to determine how (life) events affect people according to their developmental pathway, both in terms of magnitude and direction.

The impact of certain life events, e.g. job or marriage, may vary depending on the age at which they occur. Loss of a job as a teenager can mean something different than when someone is at age 25 or 35. In addition, if an event is hypothesized to trigger a behavior, the impact may well depend on the developmental path of the offender experiencing the event. Nagin (2005) provides the example of a study skills program. If there is a group of students who already possess high quality study skills, the impact of

---

[36] Individual units are typically people, but Weisburd, Bushway, Lum and Yang (2004) demonstrated trajectory modeling to identify clusters of street blocks in Seattle.

[37] Less is known on desistance. Typically desistance refers to a lack of offending by the end of data collection, which may or may not reflect actual desistance as offenders may reoffend at a later date. However, Bushway, Thornberry and Krohn (2003) demonstrate a convincing modeling of desistance through Nagin and Land's method.

the program will likely be small. Likewise, a group of students who are already disengaged from school are unlikely to increase their studying regardless of the program. The biggest impact may be observed from those students following a more intermediary trajectory. In the absence of trajectory analysis, there is a default assumption of a homogeneous "treatment" effect. Through extensions of Nagin and Land's (1993) method, we may obtain a greater understanding of how life events may differentially impact groups of individuals.

Apel, Bushway, Brame, Haviland, Nagin, and Paternoster (2007), for example, used a nationally representative sample of youth to assess the impact of intensive school year work (at age 16) on both criminal offending and substance use conditional on developmental trajectory. They found intensive work during the school year had no impact on criminal offending, on average. However, for those individuals following a rising trajectory of criminal involvement, work reduced their involvement in criminal behavior. Laub, Nagin, and Sampson (1998) examined the impact of marriage on desistance in particular, and changes in criminal behavior in general. Consistent with notions of informal social control these investigators found early "good" marriages led to a growing preventative effect on subsequent crime. The experience of a turning point was not enough; rather the *quality* and *presence* of marriage produced the crime reduction effect.

Nagin, Paganini, Tremblay, and Vitaro (2003) examined the impact of grade retention on physical aggression. They examined both the experience of the event and its timing, i.e. when the grade retention occurred. They found for some groups of individuals the experience of grade retention had no impact on aggression. However, for some

groups of individuals, following a different trajectory, the impact of grade retention on aggression depended on when, not just if, it happened. Finally, Haviland and Nagin (2005) examined the impact of joining a gang at age 14 on violent delinquency. Joining a gang at 14 increased violence in all of their three identified trajectories. However, the increase was greatest among those at greatest risk for violence. If the average impact of the turning point across all cases was used, the aforementioned effects of turning points would have been underestimated. Thus, the reaction to an event was dependent on the developmental trajectory of the cases involved.

The above examples demonstrate the utility of trajectory group modeling in assessing the impact of life events. However, there may be a reciprocal relationship between the behavior in question, and the experience of the turning point. In particular, a developmental trajectory may have increased or decreased odds of experiencing particular events relative to other groups. Although propensity score matching has been applied to make individuals within developmental trajectories as similar as possible on observed characteristics, it is unknown how well matched they are on unobserved characteristics. The ideal way to ensure comparability between individuals on observed and unobserved characteristics is through random assignment.

The Impact of (experimental) intervention

Farrington (2006) reviews longitudinal-experimental studies in criminology. He notes two advantages of this research design: 1) we can gain a greater understanding of the “natural history” of criminal offending and 2) we can assess the long-term outcomes

of an experimental intervention.[38] Longitudinal-experimental studies are rare and longitudinal-experimental studies utilizing a trajectory framework are rarer still. Most rely on self-reported criminal (antisocial) behavior, leading Farrington (2006) to note, "It is surprising that no experimental researcher seems to have analyzed detailed criminal career data from official sources for several years before and after an intervention." (p. 135). Moreover, without allowing for different developmental pathways, treatment effects are assumed to be homogenous. Utilizing the trajectory method described by Nagin and Land (1993) would make it possible to detect differential treatment effects, both in magnitude and direction.

Lacourse, Cote, Nagin, Vitaro, Brendgen, and Tremblay (2002) present the first longitudinal-experimental study to stratify treatment effects based on developmental trajectory. They identify developmental trajectories based on self-reported anti-social behavior assessed in yearly surveys from ages 11 to 17. Ultimately they found youths in the intervention program were less likely than controls to follow high rate trajectories of antisocial behavior. There are some limitations to their study. They only had antisocial measures starting at age 11, two years after the intervention. However, research demonstrates the roots of antisocial behavior start much earlier. A more appropriate approach would have been to demonstrate developmental trajectories up to the moment of random assignment in order to assess the impact of the program. In addition, the two year lag between treatment and follow-up is less than ideal.

Muthen, Brown, Booil Jo, Khoo, Yang, Wang, Kellam, Carlin, and Liao (2002) assess the effectiveness of the "Good Behavior Game" (GBG) utilizing growth mixture

---

[38]In addition, the author would add that "treatment decay" can be assessed with this research method. This is the process whereby the treatment effect diminishes as the time from treatment increases.

modeling.[39] GBG is a universal intervention with the purpose of reducing aggressive behavior. They found that students who initially were the most aggressive gained the greatest benefit from the program. As a partial replication, van Lier, Vuijk and Crijnen (2005) extended the model to assess the impact of deviant peers on antisocial behavior. They also find that children who were initially the most aggressive had the largest reductions in antisocial behavior.

Although there have been a few interventions utilizing trajectory analysis, Farrington's comment has been unheeded. This puts the present study in context. We rely on an experimental design to assess the effects of RJ on subsequent criminal offending among different offending trajectories utilizing five years of offending data collected before and after random assignment.

## DATA AND METHOD

The current study relies on data from the Reintegrative Shaming Experiments (RISE) conducted between June 1995 and August 2000 in Canberra, Australia. The RISE project consists of 4 different experiments: adult drink-driving; juvenile personal property; juvenile shoplifting; and youth (under 29) violence. The latter 3 experiments will form the basis for the present analysis. According to Knight (1999), when a criminal suspect is in custody of the police, and, during interview, fully admits to the offense, an Australian Federal Police (AFP) officer has essentially three options for action:

- Formal caution
- Put the offender before the court

---

[39] Growth mixture modeling is an approach similar to that developed by Nagin & Land (1993) which also seeks to identify distinct clusters of individuals following similar developmental pathways.

- Refer the offender to the Diversionary Conference Team

With the cooperation of the AFP and during the study period from 1995 to 2000, an officer had a fourth option of referring the case to RISE. RISE researchers subjected the treatment decision to random assignment. Case referral was left to the discretion of the AFP officer. The matter had to be serious enough to warrant a court appearance, yet not so serious that the officer would have reservations about sending the case to a diversionary conference. In addition, the offender had to agree to attend the conference and the AFP officer had to agree to abide by the random assignment decision. A total of 513 offenders representing 173 juvenile property, 113 shoplifting, and 100 youth violence cases[40] were randomly assigned to either diversionary conference[41] or typical court processing. The cases within RISE do not represent the most serious offenders, but a mix of relatively low level offenders and first-timers. Diversionary conferences consisted of the offender and supporters, the victim and supporters, and facilitated by a trained AFP officer. The purpose of the conference was to address 3 key questions: 1) what happened; 2) who was harmed by the incident; and 3) what needs to be done to repair that harm (see Strang, Barnes, Braithwaite and Sherman (1999) for a more detailed description of the experiments).

The present study will utilize the offending histories of 249 personal property, 143 shoplifting, and 121 violent offenders randomly assigned to conventional justice (CJ) or restorative justice (RJ). Subjects will be treated as part of a single large experiment rather than 3 distinct experiments. Although the nature of the instant offense was

---

[40] The unit of assignment was the case, which could include a single offender or an offender and his or her co-offenders. This was done to avoid the problem of one offender receiving a different treatment than another involved in the same criminal event.

[41] "Diversionary Conferencing," represents the RJ option in the Australian Capital Territory.

different for each experiment, it does not necessarily characterize the offending history of the offender. Cases were randomly assigned between July 1995 and June 2000, with the last observed treatment in October 2000.

**Outcome Variable:**

Official arrest history was obtained for all subjects from the AFP PROMIS[42] (Police Real-time Online Management Information System) for the time period specified. The measure of offending will consist of crimes with a personal victim (i.e. residential burglary, robbery, assault, etc.).[43] Counts of arrests are summed per offender, per year for the 5 years before and after random assignment.[44]

**Data and Analytic Strategy:**

The present study utilizes an extension of the group-based modeling technique. As originally developed in Nagin et al (2003), as a means of analyzing the effects of turning point events, we analyze the impact of an experimental intervention. We utilize Proc Traj,[45] a procedural addition to SAS, for trajectory estimation. Similar to the approach noted by Bushway et al (2003), we focused on 3 choices for fitting the models: parametric form, functional form of the trajectory, and the number of trajectories. These

---

[42] Information within the PROMIS system encompasses police contacts independent of final official resolution. It has been argued that police records represent a truer representation of behavior than court records

[43] The focus on "victimed" offense is the most appropriate outcome of a program designed to restore the harm caused to victims of crime.

[44] This strategy departs from previous program evaluations utilizing trajectory modelling. Previous evaluations have included roughly homogenous samples treated at the same period of time. Cases were referred as they came to the attention of the AFP, and thus, were unlikely to be of a homogenous age. Given the outcome was the impact of RJ (regardless of age), it makes sense to base the time period around the date of random assignment. If there is an age specific element to the offending trajectories, we relied on Proc Traj to sort this out.

[45] Proc Traj files and documentation are available at the following URL: http://www.andrew.cmu.edu/user/bjones/. In addition, see Jones & Nagin (2007) for a list of method advancements and code.

choices were informed using both the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). Differences in these measures will be utilized to select the model that most closely fits the data.[46]

Proc Traj supports 3 distributions depending on the outcome to be modeled. With count data, as in the present study, the Poisson or Zero-Inflated Poisson (ZIP) models are available. The Poisson model distribution is the standard distribution initially utilized to estimate the frequency distribution of a phenomenon (in the present case, criminal offending) given an unobserved rate (Osgood & Rowe, 1994). The ZIP model allows for more non-offenders within a given time period than predicted by the Poisson model. Positive counts are (assumed) generated by a Poisson process, whereas zeroes are generated by a Poisson process and a (polynomial) logit model (Long, 1997). Overall, the offenders in the RISE sample are relatively young (median age = 16.08), thus many are likely to have fewer contacts with the police than an adult sample. Bearing this in mind, the ZIP model makes the most sense to account for the higher proportion of cases censored at zero.

The functional form of the trajectories should be flexible enough to allow for changes in offending over time. Proc Traj allows one to specify intercept-only to cubic

---

[46] Frequently the BIC is used as the measure of fit in trajectory analysis. However, BIC assumes there is a right model and it exists within the available data. Thus, there will be a set of regression coefficients that are non-zero and a set that are zero. AIC assumes that there is no clear distinction and all you want to do is eliminate the predictors that have smaller effects (which is rigorously defined). If you cannot make the case that you have in your data set all of the relevant predictors, you are pretty much back to the AIC by default (Berk, 2007, Personal Communication).

polynomials. The cubic function allows two changes in direction, i.e. an increase followed by a decrease, and vice versa. Given this flexibility, it is a logical first choice for modeling behavior over time.[47] The final consideration is the number of trajectories. Prior research on developmental trajectories has typically found three to six developmental pathways, depending on the outcome measure and number of follow-up periods. [48]

## RESULTS

### Selecting and describing the model

Initially we develop the model to determine the form and number of trajectories. At a later stage, we will introduce the impact of assigned treatment as a covariate. Table 2.1 lists the fit statistics for the estimated trajectory models. We tested all models with between 2 and 7 trajectory groups. In addition we tested both the linear and quadratic form of the logistic function. In the end, the 6 group model, with a quadratic logistic function, provided the best BIC and AIC value, thus indicating the best "fit" to the data. Initially each trajectory group pathway was estimated using a cubic functional form. Subsequent to model selection, we re-estimated the model setting the highest non-significant polynomial to zero. Table 2.2 provides the results for the trajectory model estimation. Figure 2.1 provides a visual depiction of the offending trajectories. Table 2.3 provides the posterior probabilities for group membership. Although not as high has some published studies, on average individuals assigned to a particular group had a relatively high probability of inclusion in that group, These probabilities range from .77

---

[47] Unless specified, Proc Traj uses cubics by default
[48] See Nagin et al. (2003) for a more detailed explanation of model development.

for Group 2 (High Drifters) to .91 for Low Blips. For individuals with relatively low posterior probabilities, the low probability represented possible inclusion in two groups with relatively similar offending pathways. This issue will be addressed later.

There are three low offending groups. Low Drifters, representing a quarter of the sample (24.5%), demonstrate a gradual rise followed by a gradual decline after random assignment. These individuals represent relatively low involvement in the traditional criminal justice process. Their offending never rises above 79 offenses per 100 offenders (which occurs in the year prior to random assignment). Throughout the ten years available, they begin and end their "career" at nearly the same level of offending. High Drifters, representing 2.4 percent of the sample, maintain a steady, albeit low, rate of offending throughout the study period. These individuals, on average, are arrested about once per year. Finally, there is the Little Blips group, which is essentially a group of negligible offenders. There is a spike in their offending at the time of random assignment; however, this would make sense given one must offend to enter the experiment. This group contains the majority of the first-time offenders in the sample. The spike at the time of random assignment likely represents an initial foray into criminal behavior.[49]

There are also three relatively high offending groups, although they differ in the shape of their offending trajectories. There are two groups that increase their offending around the time of random assignment. Low Plateaus begin to increase their offending in the year prior to random assignment and continues until they hit a peak of 3 offenses per year. This plateau is maintained for approximately 3 years before declining. Big Blips on the other hand, peak at nearly 3 offenses in the year prior to random assignment before

---

[49] Or this represents the first official contact. See Woods (2003) for a discussion of the differences and similarities between the RISE offenders official and self-reported criminal offending.

beginning a decline. Finally there is one group (Steep Risers) more involved in criminal offending. This group offends at a higher rate than the others. Their trajectory is marked by a steep rise, to a peak of 6 offenses per year in the year following random assignment, and then begins a decline. At the close of the time period covered in the follow-up period, this group's offending is lower than both Big Blips and Low Plateau groups.

Table 2.4 displays demographic characteristics of each offending group. Each group is majority male, with Little Blips, the non-offenders, having the greatest percentage of females (29%). Aboriginal offenders are over-represented in the higher offending groups (38%). Although many of the first time offenders (present in Little Blips) do not continue their official offending career (as evidenced by the flat, near zero offending trajectory), first time offenders in Low Plateau seem to initiate their offending career. The majority of the groups have little time between first arrest and their age at random assignment (roughly less than a year). However, two groups (High Drifters and Steep Risers) have long offending careers, albeit different pathways.

Although there are commonalities among the groups, there are differences that may relate to the impact of a restorative justice conference. We next estimate the model including randomly assigned treatment (either RJ or conventional justice (CJ)) as a covariate. In doing so, we will be able to see if RJ has a differential impact on offending history within trajectory groups. RJ may have no impact, increase, or reduce offending within a given offending pathway.[50]

---

[50] An alternative approach would be to look at the conditional probabilities of being in a trajectory given treatment assignment. Under this framework trajectories would not be defined by treatment. However, given the interests in assessing the direct treatment effect the covariate adjustment seemed the most prudent (MacDonald, personal communication, 2008).

**The impact of RJ on offending trajectories**

Figure 2.2 presents the trajectory model including randomly assigned treatment as a covariate.[51] The trajectories for each individual group are presented in Figures 2.3 through 2.9. At first glance, the experience of RJ is different depending on the group's developmental trajectory. For those groups who "dabbled" in criminal offending around the time of random assignment, RJ has no discernable effect on offending relative to traditional court processing. Within two of the other groups, those assigned to RJ significantly increased their offending relative to controls. Finally, within the two remaining groups, RJ assigned offenders experienced decreases in offending relative to controls. Depending on one's 'developmental' history of offending, the impact of RJ causes different effects.

We begin by describing the groups which experienced no difference in offending due to RJ, relative to the control condition of traditional court. The Little Blips had the majority of first-time offenders. It would appear that either response, RJ or CJ, was enough to end their offending career,[52,53] at least among this group. Nearly 87 percent of first-timers within the Little Blips never had a recorded offense within the 5 year follow-up period. Among all offenders in this group, the highest frequency following random assignment is 2 new offenses. This group essentially consists of individuals who 'tested the waters,' as it were, of criminal offending, experienced a criminal justice intervention (whether RJ or standard treatment) and ceased their official criminal offending.

---

[51] We follow the "Intention-to-Treat" (ITT) principle in these analyses. Regardless of the actual treatment delivered, we analyze the offenders as their treatment was intended. Thus, we examine a policy of RJ rather than the impact of the program itself.
[52] At least within the available data.
[53] $p \leq .970$

The High Blip is comprised of individuals who began their offending career in the year before random assignment. Regardless of assigned treatment, the offenders within this group continued their offending careers post random assignment, a career that is marked by a declining offending frequency as time passes from the date of random assignment.[54] These individuals were relatively young, with little 'career' time between their first arrest and random assignment (see Table 2.4). All individuals within this group, regardless of randomly assigned treatment, go on to offend, on average, at least once per year in the five years following random assignment. Given that these offenders were on average younger than any of the other groups, their offending pattern resembles something akin to adolescent-limited offending.[55]

Next we turn our attention to groups where RJ produced a decline in offending. Among those offenders in 'drifting' groups, those randomly assigned to RJ experienced declines in offending, relative to controls. Low Drifters represent very low level offenders, averaging less than one offense per offender in the time leading up to random assignment. Following random assignment, we see a 'steep' decline for offenders assigned to RJ.[56] This decline quickly levels off to near zero offending. Among those assigned to court, the path to near zero offending is more gradual. Given the low level of offending among this group in general, the declines experienced by the RJ sub-group is less noteworthy than if there had been greater 'gains.'

High Drifters represent offenders with near constant offending or about 2 official offenses per year, beginning and ending with, on average, one offense five years prior to

---

[54] $p \leq .736$.
[55] See Moffitt (1993).
[56] $p \leq .001$.

and following random assignment. However, among those assigned to RJ we would expect a 50 percent reduction in offending in the years following random assignment, as their offending drops from an average of two official offenses per year to less than one.[57] Among this group, assignment to RJ reduces offending. This reduction is maintained through the five years of available data. Offenders in this group, on average, are the oldest offenders with the longest 'career length' prior to random assignment. Although they continue to offend following random assignment, on average, assignment to RJ produces a consistent decline relative to standard court processing. Finally although these individuals do not represent the 'worst' offenders (in terms of offending frequency), the impact of RJ might be worth the effort as it produced a quick reduction in offending that is maintained.[58] However, the experience of RJ can produce a backfiring effect. Backfiring is evident in both of the groups with the highest offending frequency, either as "new" offenders, or among those who have a high rate of offending in the years prior to random assignment.

The Low Plateau group demonstrates a rising offending pattern in the years prior to random assignment. These offenders averaged a little over a year between their first official contact with the CJS to the date of random assignment. Their offending peaks in the year following random assignment; however, we would expect about a third less offending among those individuals randomly assigned to court, as evidenced by their trajectory post-random assignment. Among these individuals, assignment to court appears to halt a progression to greater offending.[59] Although this group, as a whole, has

---

[57] $p \leq .005$.
[58] At least over the 5 year follow-up period.
[59] $p \leq .005$.

a short offending career prior to random assignment, it is marked by increasing frequency of offending in the years prior to random assignment. Offenders exhibiting this pattern of offending would not do well in a RJ setting.

Steep Risers represent the group of offenders who are most engaged in criminal offending. In the years leading up to random assignment, their offending pattern is marked by an increasing offending frequency.[60] This offending pattern peaks in the year following random assignment. However, among those assigned to RJ, there is a steep increase in offending in the year following random assignment to RJ. It is apparent that the experience of RJ produces a defiant reaction among these individuals (Sherman, 1993). In general these individuals have a long offending career, averaging nearly three years between the age at first arrest and random assignment. Their offending also increases each year. These offenders also enter their careers at an early age, averaging 12 years old at the time of their first CJS contact. Entry at such a young age is an indicator of offending propensity and appears to be unaffected by assignment to RJ. From the available data, RJ is not an appropriate option for those with this pattern of offending. Although assignment to conventional justice does not end their offending careers, the pathway to reduced offending is quicker than for those assigned to RJ.

An intervention, such as RJ, is likely to have variable effects depending, among other things, on the characteristics of the offenders involved in that process. We have demonstrated the utility of using trajectory modeling for parsing out these variable effects. At least within the ACT, assignment to RJ does not produce any discernible effect relative to court assigned controls. Likewise, among those individuals who have a

---

[60] $p \leq .023$.

relatively short career-length, assignment to RJ produces no impact. However, among those individuals with careers marked by steady offending or among those with very low offending, assignment to RJ produces a decline in offending. What is most noteworthy is the impact of RJ on those who have greater involvement in the CJS. Among these individuals, assignment to RJ produces greater offending and appears to backfire. Relying on these trajectory group differences one might recommend against using RJ on individuals exhibiting these higher rate offending patterns. However, this is an incomplete story. Returning to Table 2.4, we see that most of the Aboriginal offenders (nearly two-thirds) were grouped in the highest offending groups. Is it possible that the experience of RJ among this demographic group drove the effects exhibited? Is the combination of greater involvement with the CJS and RJ detrimental among Aboriginal offenders?

**Racial differences in the experience of RJ**

Sherman, Strang, Barnes and Woods (2006) found Aboriginal offenders who participated in RJ were more likely to have a negative experience and increase their offending. Aboriginal offenders found RJ to be more emotionally intense, felt the outcomes were "tough," while feeling less forgiven and more stigmatized. Although, in general, Aboriginal s in the ACT felt RJ to be more procedurally fair, many increased their offending in the years following random assignment. For the above reasons, those offenders within the highest offending trajectory groups are likely to be the ones who drive the curve upwards after random assignment. There are fewer than 50 Aboriginal offenders within the three youth experiments. As a result, rather than conducting

trajectory analysis with the Aboriginal offenders, we removed them and analyzed the non-Aboriginal subsample.[61]

The offending trajectories for the non-Aboriginal sample are illustrated in Figure 2.9. A five group model produced the best fit to the data, based on both AIC and BIC measures. It is immediately apparent RJ has a different impact on the group exhibiting an increasing trajectory of offending pre-random assignment.[62] Rather than witness a spike in offending in the year following random assignment for those assigned to RJ with relatively high offending trajectories, we see the increase in offending is halted. Non-Aboriginal offenders assigned to court appear to increase their offending before leveling off and declining in later years. Among those who start offending early and have longer offending careers (prior to random assignment), RJ has a differential impact that is race dependent. Within the ACT, RJ puts a stop to rising offending among non-Aboriginal offenders, but increases the offending among Aboriginal offenders.

The Low Plateau group does not experience the backfiring effect of RJ. In the full sample, RJ increased offending among this group, whereas in the non-Aboriginal sample we see no difference in offending relative to court assigned offenders.[63] This offending group had the second highest proportion of Aboriginal offenders. It is likely the backfiring effect originally observed was due to Aboriginal offenders increasing their offending post-random assignment to RJ.

---

[61] Most of the subsample is of European descent; however, there are enough who claim Asian ancestry to make the term "Non-Aboriginal" more descriptive.
[62] $p \leq .053$.
[63] $p \leq .801$.

Again we see the beneficial impact of RJ among those offenders who dabble in offending in the years prior to random assignment.[64] Although these offenders had low offending frequencies prior to random assignment and represent a majority of first-time offenders, RJ reduces their offending relative to court in the years following random assignment. We see a relatively steep decline in offending relative to court. Court assigned offenders within this group take nearly four years post random assignment to match the offending levels of their RJ assigned counterparts.

Within the final two offending groups, RJ does no better nor worse than court assigned offenders. [65,66] These two groups included many first-time offenders or offenders who had recently begun their offending careers, as evident by the relatively short time between the first recorded offense and their age at random assignment (six months or less, on average, for both). There were larger proportions of female offenders in these two groups relative to the other offending groups. In general, these offenders began their criminal career later (after age 15), and it appears contact with either the CJS or RJ was enough to effectively end[67] their criminal career or reduce it. These offenders have a spike in offending in the year prior to random assignment, including many of whom had their very first recorded offenses get them into the RISE experiments.

One final thing to note is the lack of a backfiring effect of RJ, at least within the non-Aboriginal sample. At best, RJ seems to reduce offending among those with a longer offending career, characterized by a high frequency of offending, and among "dabblers." At worst, RJ does no harm relative to court among many first-time or low rate offenders.

---

[64] $p \leq .005$
[65] $p \leq .470$
[66] $p \leq .648$
[67] At least within the available data.

RJ as a policy does not evince a criminogenic effect among non-Aboriginal offenders in Canberra.

## DISCUSSION & CONCLUSIONS

Farrington (2006) noted a lack of longitudinal-experimental studies utilizing official offending as an outcome. Using data from the Reintegrative Shaming Experiments in Canberra, Australia, we have answered his call. We have modeled the offending trajectories of 513 youth offenders across 10 years using official arrest data from the Australian Federal Police in Canberra, Australia. We gleaned 6 offending trajectories from these data. In general, offenders (in the ACT) who were referred to the RISE experiments do not represent the most prolific offenders. However, the most criminally involved offending group averaged 6 (officially recorded) offenses per year and began their careers, on average, 3 years before random assignment. Although RISE included many first-time offenders, some of whom continued their offending careers post-random assignment, the experiments included many offenders who were well involved in the criminal justice system prior to RJ becoming an option. These different offending trajectories would be masked when examining the offending pattern of the full sample.

With the above fact in mind, we examined the impact of random assignment to RJ on the offending trajectory of the group, rather than focus on mean differences between RJ and court processing. We discovered that the impact of RJ on future offending was, in fact, variable depending on the offense trajectory of the offender involved. We discovered an even split among crime reduction, crime escalation, and null effects. Crime escalation effects were isolated to Aboriginal offenders with greater involvement in

criminal behavior prior to random assignment, while null findings and crime reductions were evident among lower level offenders. Further examination revealed the offending trajectories where there was a backfiring effect contained a larger proportion of Aboriginal offenders. While the small Aboriginal sample size limits our ability to draw firm conclusions, we have conclusive evidence that within the ACT, Aboriginal participation in RJ results in increased offending, relative to standard court processing (Sherman et al. 2006). When Aboriginal offenders were removed from the sample and the data reanalyzed, the escalation effects either reversed or disappeared. Among one group of offenders, RJ was no different than court; however, among those offenders most involved in official offending, participation in RJ brought about a stop in an escalating crime trajectory.

"Does RJ work?" The answer to this question depends on the history of the offenders assigned to an RJ process. We have demonstrated that the impact of RJ can vary depending on the offending trajectory of the offenders involved. RJ appears to decrease the escalation in offending for high-rate offenders and even shows reductions for those offenders who are dabbling in criminal behavior. Low rate offenders (who show small increases in offending in the years prior to random assignment) also reduced their offending relative to court-treated controls. In general, first-time offenders do no worse in RJ than through standard court processing. The findings have implications for policy. RJ shows the biggest benefit for those who have a long history of personal victim offenses, at least for non-Aboriginal offenders. It would behoove us to delve into the reasons why RJ backfired for individuals with an Aboriginal background. RJ could have produced a defiant reaction in Aboriginal offenders who have experienced institutionalized racism.

Alternatively RJ failed to provide the “teeth” to be an effective deterrent. The conference outcome agreed upon in the conference setting was not an effective deterrent to future criminal behavior. In addition, RJ conferences involving an Aboriginal offender invariably involved a non-Aboriginal victim. As such, the RJ process may not have created the desired empathy between offender and victim.

In a separate vein, future research needs to focus on the costs of RJ relative to the benefits. If we determine that RJ overall is more cost effective, it makes more sense to process juvenile offenders via this route. In the end, RJ is a procedurally fair process for both offenders and victims (Sherman et al, 1999; Strang, 2001). Although the offenders involved in the RISE experiments did not represent the “worst” offenders, we have evidence that RJ can alter the criminal offending trajectory of offenders who are significantly involved in offending, and also provides additional social benefits beyond crime reduction.

Table 2.1: Fit Statistics for 2-7 Trajectories of Personal Victim Offending

| Groups | Linear Logit Function | | |
|---|---|---|---|
| 2 | BIC = -3860.29 (N=5130) | BIC = -3847.62 (N=513) | AIC = -3824.30 |
| 3 | BIC = -3732.16 (N=5130) | BIC = -3713.74 (N=513) | AIC = -3679.81 |
| 4 | BIC = -3649.92 (N=5130) | BIC = -3625.75 (N=513) | AIC = -3581.22 |
| 5 | BIC = -3636.46 (N=5130) | BIC = -3606.52 (N=513) | AIC = -3551.40 |
| 6 | BIC = -3609.87 (N=5130) | BIC = -3574.18 (N=513) | AIC = -3508.45 |
| **7** | BIC = -3666.02 (N=5130) | BIC = -3624.57 (N=513) | AIC = -3548.25 |
| | Quadratic Logit Function | | |
| 2 | BIC = -3845.21 (N=5130) | BIC = -3831.40 (N=513) | AIC = -3805.95 |
| 3 | BIC = -3703.18 (N=5130) | BIC = -3683.61 (N=513) | AIC = -3647.57 |
| 4 | BIC = -3653.55 (N=5130) | BIC = -3628.22 (N=513) | AIC = -3581.57 |
| 5 | BIC = -3643.53 (N=5130) | BIC = -3612.44 (N=513) | AIC = -3555.20 |
| **6** | **BIC = -3598.01 (N=5130)** | **BIC = -3561.17 (N=513)** | **AIC = -3493.33** |
| **7** | BIC = -3631.86 (N=5130) | BIC = -3589.26 (N=513) | AIC = -3510.81 |

Table 2.2: Regression results for trajectory slope parameters

| Group | Parameter | Estimate | Error | t-value | Prob > \|t\| |
|---|---|---|---|---|---|
| 1 | Intercept | -0.27640 | 0.11665 | -2.369 | 0.018 |
| | Linear | -0.22004 | 0.06172 | -3.565 | 0.000 |
| | Quadratic | -0.09394 | 0.01645 | -5.710 | 0.000 |
| | Cubic | 0.01645 | 0.00444 | 3.706 | 0.000 |
| 2 | Intercept | 0.32743 | 0.15143 | 2.162 | 0.031 |
| 3 | Intercept | 0.25531 | 0.11611 | 2.199 | 0.028 |
| | Linear | -1.25211 | 0.12722 | -9.842 | 0.000 |
| | Quadratic | -0.91394 | 0.08640 | -10.579 | 0.000 |
| | Cubic | 0.21127 | 0.02006 | 10.533 | 0.000 |
| 4 | Intercept | 0.86520 | 0.16174 | 5.349 | 0.000 |
| | Linear | 0.46612 | 0.06982 | 6.676 | 0.000 |
| | Quadratic | -0.15768 | 0.03980 | -3.962 | 0.000 |
| | Cubic | 0.01534 | 0.00761 | 2.017 | 0.044 |
| 5 | Intercept | 1.17181 | 0.10599 | 11.056 | 0.000 |
| | Linear | -0.21380 | 0.07462 | -2.865 | 0.004 |
| | Quadratic | -0.26329 | 0.03820 | -6.893 | 0.000 |
| | Cubic | 0.06118 | 0.00736 | 8.312 | 0.000 |
| 6 | Intercept | 1.77069 | 0.07353 | 24.08 | 0.000 |
| | Linear | 0.13126 | 0.04122 | 3.185 | 0.002 |
| | Quadratic | -0.04285 | 0.00922 | -4.650 | 0.000 |
| | Cubic | -0.00747 | 0.00294 | -2.542 | 0.011 |

Table 2.3: Posterior probability group assignments for 6 group model

| Group | N | Mean Probabiliy | Standard Error | Minimum |
|---|---|---|---|---|
| **1** | 113 | .84 | .0145 | .40 |
| **2** | 11 | .77 | .0673 | .43 |
| **3** | 289 | .91 | .0052 | .50 |
| **4** | 24 | .88 | .0306 | .49 |
| **5** | 60 | .80 | .0003 | .34 |
| **6** | 16 | .91 | .0477 | .36 |

Table 2.4: Characteristics of Group members

| | **Group 1 (n=113)** | **Group 2 (n=11)** | **Group 3 (n=289)** | **Group 4 (n=24)** | **Group 5 (n=60)** | **Group 6 (n=16)** |
|---|---|---|---|---|---|---|
| **% Male** | 80 | 91 | 71 | 96 | 80 | 88 |
| **% Aboriginal** | 12 | 9 | 3 | 17 | 23 | 38 |
| **% First-time offender** | 34 | 0 | 65 | 33 | 13 | 0 |
| **Age at First Arrest** | 14.67 | 13.76 | 16.15 | 13.64 | 14.16 | 12.64 |
| **Age at RA** | 16.33 | 18.29 | 16.47 | 14.07 | 15.24 | 15.67 |

Figure 2.1: Predicted trajectories for the 6 group ZIP model

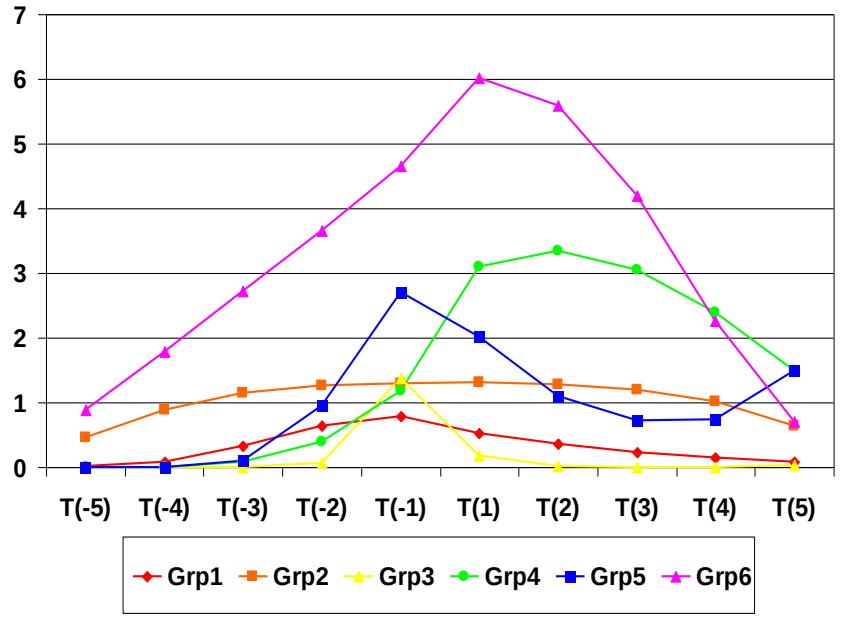



Figure 2.2: Impact of RJ on offending trajectories – 6 group model

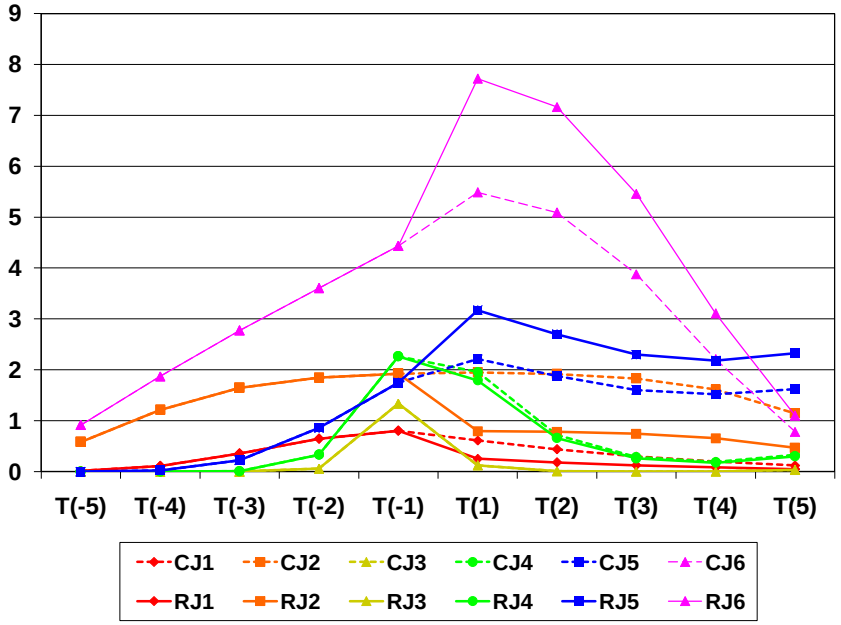


Figure 2.3: Group 1 (Low Drifters) – Impact of RJ on offending

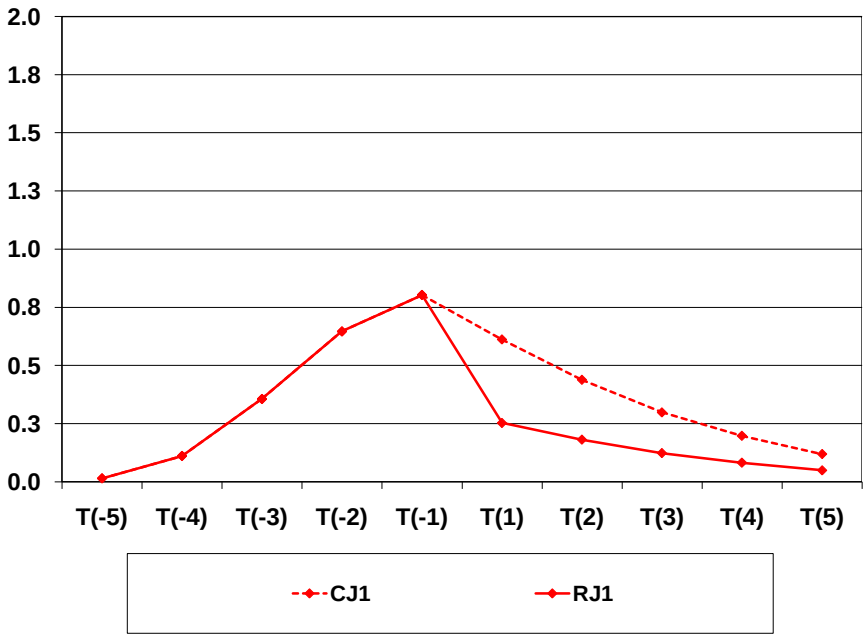



Figure 2.4: Group 2 (High Drifters) – Impact of RJ on offending

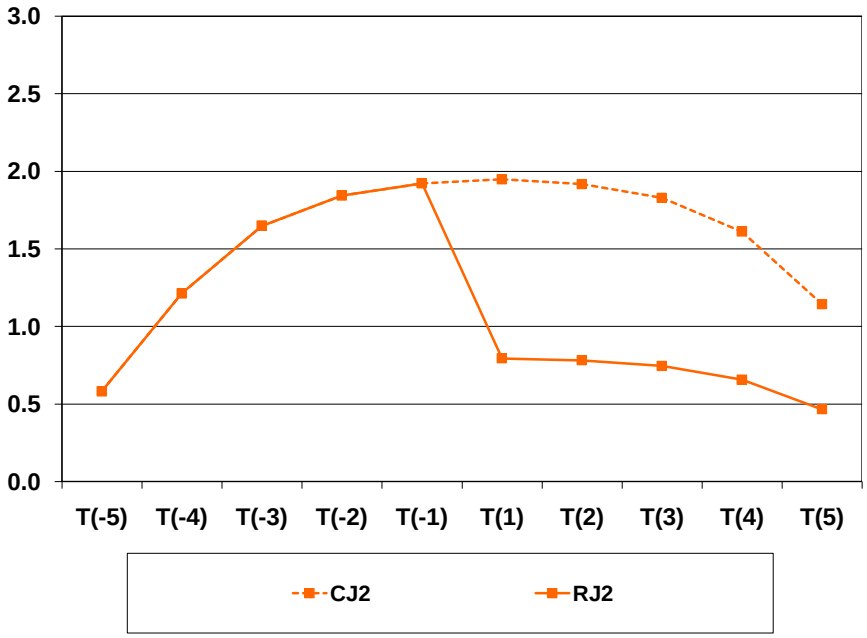


Figure 2.5: Group 3 (Little Blips) – Impact of RJ on offending

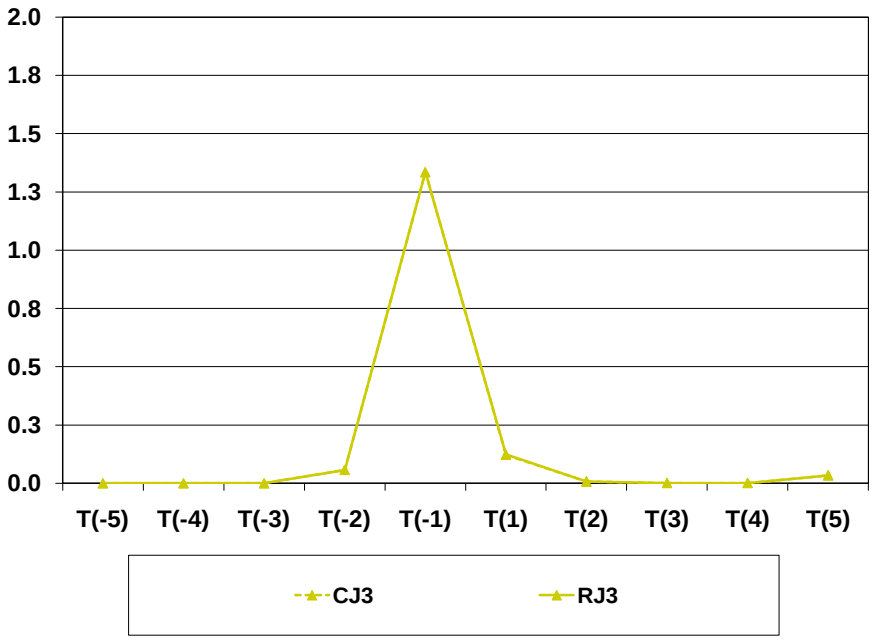



Figure 2.6: Group 4 (Big Blips) – Impact of RJ on offending

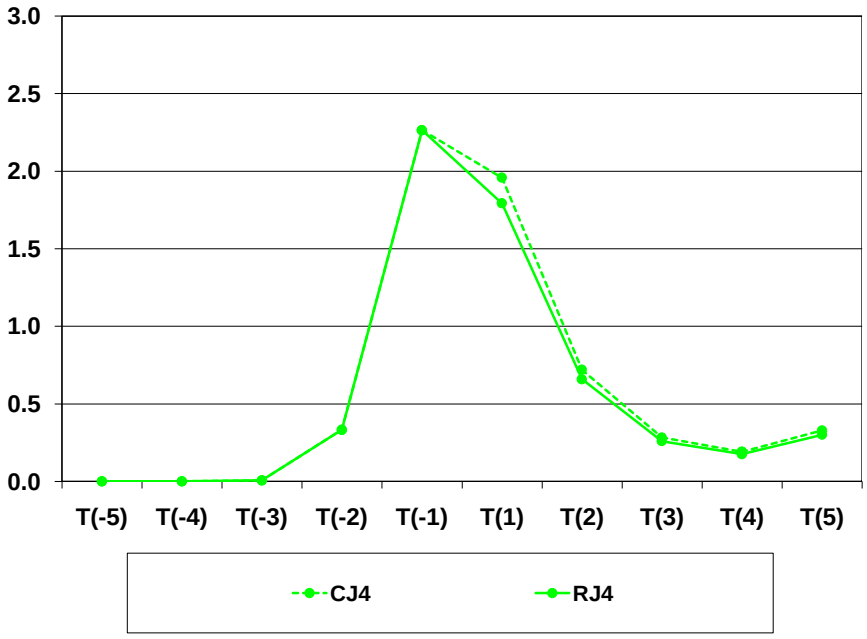


Figure 2.7: Group 5 (Low Plateau) – Impact of RJ on offending

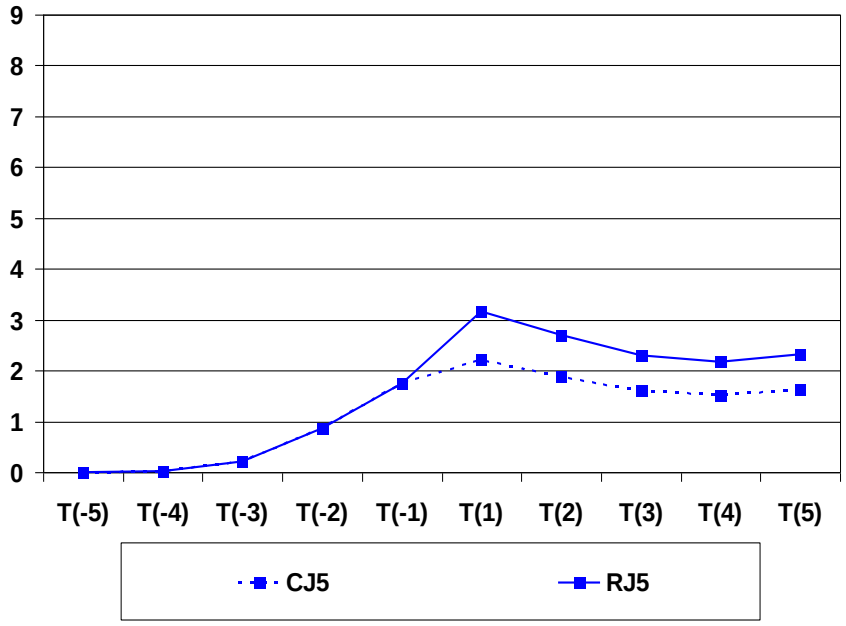



Figure 2.8: Group 6 (Steep Risers) – Impact of RJ on offending

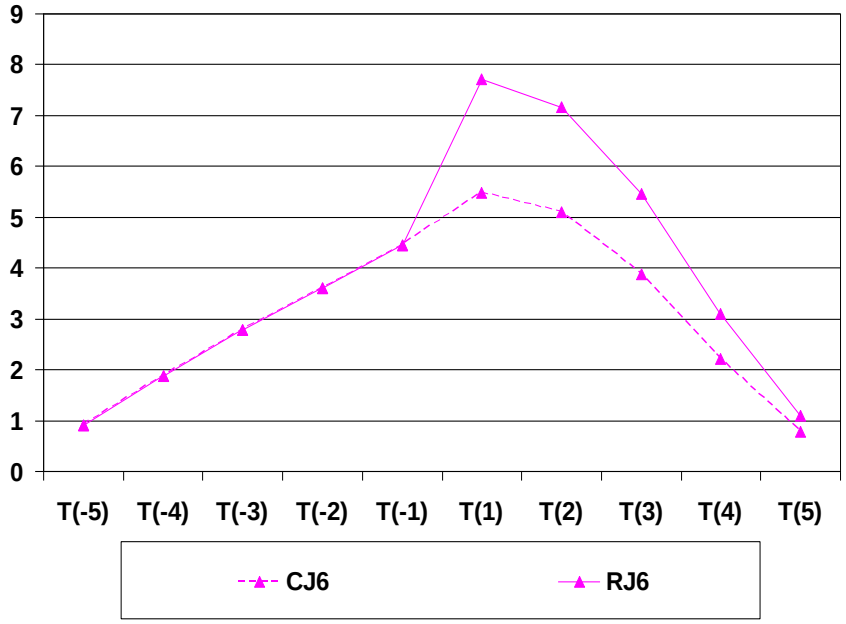
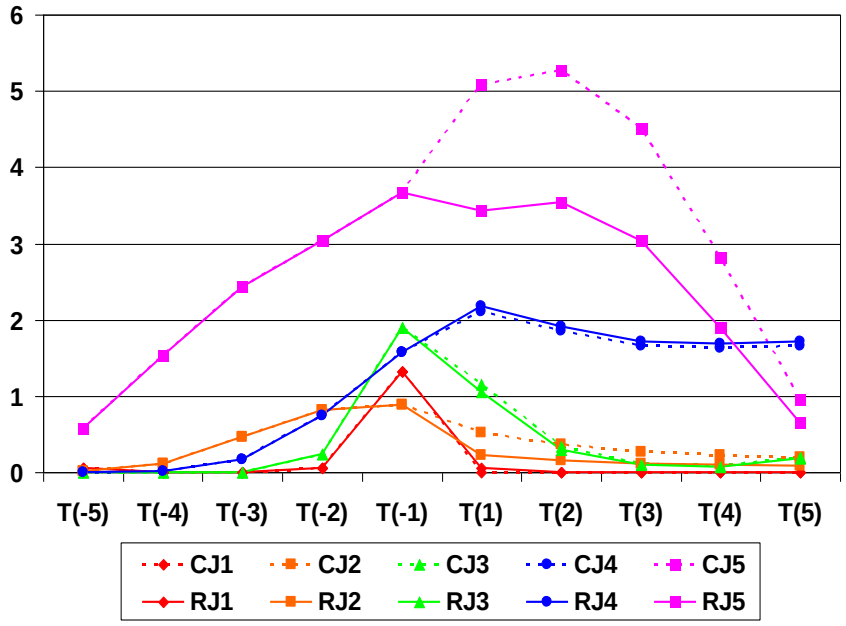


Figure 2.9: Non-Aboriginal subsample – Impact of RJ on offending



# CHAPTER 3: Characteristics of treatment non-delivery in the Reintegrative Shaming Experiments

## INTRODUCTION

Restorative Justice (RJ), as a system of justice, offers an improvement over conventional justice, based on retributive models, by focusing on repairing the harm caused by the crime. RJ is a process by which the stakeholders (offender, victim, community, etc.) in a specific offense collectively handle the aftermath of the offense. (Marshall, 1999). This broad definition covers a broad range of programs from face-to-face interactions between victim and offender to "shuttle communication" between parties through an intermediary. Arguably, face-to-face conferences are the most commonly evaluated type of RJ. Offenders and victims, sometimes with supporters of each, meet in the presence of a facilitator to discuss a criminal event and decide what needs to happen to repair the harm caused. The RJ process provides a better forum where offenders and victims have a voice in the proceedings, leading to "better outcomes" than the typical experience of people who go through the criminal justice system. However, better outcomes due to RJ are dependent on its successful delivery, i.e. the offenders and victims (and supporters) actually meet. Because RJ conferences require the attendance of multiple people who must coordinate their schedules, non-delivery of RJ conferences must be seen as a possible outcome. The present paper will focus on the use of randomized controlled trials to evaluate the effectiveness of RJ relative to conventional justice.

Evaluations of face-to-face RJ utilizing a randomized control trial design, depending on one's point of view, dually provide a test of policy and a test of practice.

The degree to which these related but different questions[68] can be answered largely depends on the degree of compliance with the randomly assigned treatment. Sherman and Strang (2004) describe the difference between these two goals within the context of interpreting the results from RCTs:

> The … problem revealed by the growing number of RCTs in crime prevention is the theory versus policy problem of defining what hypothesis has been tested: a theory, a policy, or both. Ideally, all RCTs try to test a theory of causation that underlies a policy or program that is randomly assigned to some units but not others. In practice, the test is rarely completed with all units randomly assigned to the program. This result forces a choice between analyzing data on the basis of randomly assigned "intentions to treat" (ITT) or on the basis of selectively occurring "treatment on the treated" (TOT), or both. ITT analysis tests the hypothesis that a policy of attempting to apply a program causes less crime. TOT analysis tests the hypothesis that the theory of applying a program to reduce crime is correct when the program is actually applied. (p. 578).

The extent to which an evaluation can answer either of these questions depends on the degree of compliance with the assignment to treatment. The test of policy provides a test of a system whereby everyone assigned, or mandated, to treatment does not necessarily comply with treatment. In essence, non-compliance[69] is built-in to the system. After all, not everyone takes the prescribed pill or attends the mandated counseling session. Implementation of treatment within the context of a randomized control trial (RCT) makes inferences about the effect of treatment more complex.

Statistical virtuosity or good study design?

In general, we know more about the existence (and correction or modeling) of non-compliance that we do about why it occurred in the first place. The fact that people

---

[68] In extreme cases these questions can be mutually exclusive.

[69] Non-compliance here is used in a general sense. In essence it refers to any case where the assigned treatment is not delivered. This can be due to refusals by study participants, errors by study personnel, or outright tampering with the random assignment sequence. Hollis and Campbell (1999) have noted, "… most types of deviations from protocol would continue to occur in routine practice…" (p. 671), but are open for some deviations to occur only in the trial setting.

do not always do what they told is often taken as given without an understanding of the factors that may influence non-compliance. Berk and Xu (2003) noted one of the problems with (randomized field) experiments was people often do not do what they are told,[70] leading them to question: "Why does it seem that the individuals who designed this experiment were caught by surprise?" (p. 319). If non-compliance with treatment should be expected, then researchers can plan ahead.

Berk, Smyth, and Sherman (1988) provide some helpful hints. They noted potential obstacles to random assignment should be discussed with study personnel. In addition, researchers should monitor the assignment of cases to treatment. If problems are identified, remedial measures should be taken. They advise the collection of data on the implementation of random assignment in order to inform statistical procedures to assess the impact of non-delivery of treatment. In the absence of these remedial measures post-hoc statistical corrections can be utilized. However, Berk and Xu (2003) state "[v]irtuoso statistical analyses can certainly help …, but they must not be seen as an inexpensive alternative to doing the study well to begin with" (p. 320).

The present paper focuses on the factors that influence non-delivery of RJ conferences. Any treatment modality is likely to encounter non-delivery. We investigate whether offender demographics are related to non-compliance with treatment, while also examining whether specific (logistic) characteristics related to the organization of a conference can influence non-delivery. Although there are statistical techniques to account for non-compliance with treatment, we argue the "trick" is to minimize the influence of factors associated with non-delivery. We will provide evidence for the

---

[70] We would extend that to many human endeavors.

factors associated with non-delivery while providing policy recommendations for those interested in pursuing a program of restorative justice who wish to maximize the treatment delivery.

## METHODS

### Data

The current study relies on data from the Reintegrative Shaming Experiments (RISE) conducted between June 1995 and August 2000 in Canberra Australia.[71] The RISE project consists of 4 different experiments: adult drink-driving; juvenile personal property; juvenile shoplifting; and youth (under 29) violence. The latter 3 experiments will form the basis for the present analysis. According to Knight (1999), when a criminal suspect is in custody of the police, and, during interview, fully admits to the offense, an Australian Federal Police (AFP) officer has three options for action:

- Formal caution
- Put the offender before the court
- Refer the offender to the Diversionary Conference Team

With the cooperation of the AFP and during the study period from 1995 to 2000, an officer had a fourth option of referring the case to RISE. RISE researchers subjected the treatment decision to random assignment. Case referral was left to the discretion of the AFP officer. The matter had to be serious enough to warrant a court appearance, yet not so serious that the officer would have reservations about sending the case to a

[71] See Sherman et al. (1999) for a more detailed description of the RISE experiments.

diversionary conference. Each case had to meet several eligibility criteria. Generally cases had to meet the following:

- they (and all co-offenders) had made full admissions about committing the offense
- they (and all co-offenders) had no outstanding warrants or bonds which would require them to attend court
- there was no reason to believe that they (or any co-offender or parent) would object to a conference
- they (and all co-offenders) lived in the ACT region
- the apprehending officer's sergeant approved the case being sent to RISE
- the apprehending officer agreed to accept the RISE recommendation (based on random assignment) for the case regardless of whether it was court or conference.

In addition to the above criteria, the Juvenile Property experiments (JPP & JPS) had to include at least one co-offender under 18 years of age. The Youth Violence experiment had an expanded age criterion. At least one co-offender had to be less than 30 years old.[72]

A total of 513 offenders representing 173 juvenile property, 113 shoplifting, and 100 youth violence cases[73] were randomly assigned to either diversionary conference[74] or typical court processing. The cases within RISE do not represent the most serious offenders, but a mix of relatively low level offenders and first-timers. Diversionary conferences consisted of the offender and offender supporters, the victim and victim

---

[72] Answers to these eligibility criteria were covered in the initial call to the RISE research team in an effort to minimize potential non-delivery of treatment by ensuring study participants were eligible and were more likely to comply with treatment.

[73] The unit of assignment was the case, which could include a single offender or an offender and his or her co-offenders. This was done to avoid the problem of one offender receiving a different treatment than another involved in the same criminal event.

[74] "Diversionary Conferencing," represents the RJ option in the Australian Capital Territory.

supporters, and facilitated by a trained AFP officer. The purpose of the conference was to address 3 key questions: 1) what happened; 2) who was harmed by the incident; and 3) what needs to be done to repair that harm (see Strang, Barnes, Braithwaite and Sherman (1999) for a more detailed description of the experiments).

The present study focuses on the 264 offenders[75] (representing 192 cases) from the three youth experiments who were randomly assigned to RJ. The sample consists of cases where the final treatment was known and could be delineated in the manner described below. Cases were randomly assigned between July 1995 and June 2000.

**Measures**

Outcome Variable:

The outcome in the present study is the final treatment for all youth cases randomly assigned to RJ. Although the majority of offenders received the randomly assigned treatment, in some cases treatment failed for one or more reasons:

- Offender-related reasons
    - Offender rejected the RJ conferencing, exercising his/her right to have the case processed through court.
    - Offender reoffended after random assignment and before an RJ conference was scheduled.
    - Offender withdrew his full admission to the offense.[76]

---

[75] For 2 offenders it was not possible to determine the reason for the treatment non-delivery and so those individuals are excluded from the analyses.

[76] RJ conferences were not meant to be fact-finding endeavors. Offenders were required to admit to the offense, or "decline to deny" the offense in order to facilitate a discussion of the incident, as a condition of eligibility for random assignment.

    - Offender failed to appear to the RJ conference or in some cases moved out of the territory.
- Administrative-related reasons
    - Victim failed to appear to the RJ conference.
    - Administrative error resulting in a RJ case being referred to court.
    - RJ conference never scheduled.
    - Police had "second thoughts" about the need for or appropriateness of a RJ conference.

In some cases if a treatment event "failed," the RJ facilitator attempted to reschedule another conference. In other cases the matter was either referred to the court, dropped by the police (no further action), or the offender was "cautioned."

We have divided the cases into three groups, based on the archival records of the treatment events: Treated as assigned (i.e. received a RJ conference), non-treatment for offender reasons (offender non-compliance), and non-treatment for logistical reasons (administrative failure).[77] Table 3.1 lists the summary statistics for the outcome and explanatory variables. Overall most offenders received the assigned treatment (74%), with the remaining 26 percent almost evenly divided between treatment failures for offender reasons (12%) and administrative reasons (14%).

<u>Explanatory variables:</u>

- Offender sex – The variable is coded as "1" for male; "0" for female.
- Attitude at the time of arrest – This was based on a subjective assessment by the referring officer and was originally coded as "Good," "Bad," or

[77] Administrative reasons go beyond the police and include victim refusals or no-shows as well.

"Indifferent." The variable has been recoded so that a value of "1" equals "Good Attitude."

- Season of the year – This variable is dummy coded for "Summer," "Fall," and "Winter" with "Spring" as the reference category.
- RISE Experiment – The RISE experiments were originally four separate experiments, with three experiments included in these analyses. The two property experiments (Juvenile Personal Property (JPP) and Juvenile Shoplifting (JPS)) are dummy coded with the Youth Violence (JVC) as the reference category.
- Offender Age
- Offender One Year Prior Offending – This variable includes all officially recorded "victimed"[78] offenses committed within one year of random assignment. This variable does not include the instant offense resulting in the offender's referral to the RISE experiment; therefore a value of "zero" represents first-time offenders.
- Career Length – This variable represents the time (in years) between an offender's first officially recorded crime and the date of random assignment.
- Weeks to Initial Treatment (WIT) – This variable represents the time (in weeks) between random assignment and the first treatment attempt, regardless of the number of attempts required for a case completion. Some

[78] "Victimed" offenses included person (assault, robbery, etc.) and property (burglary/theft, vandalism, etc.) offenses where there is a clear personal victim. These offenses are the most appropriate focus for a program intended to reduce harm.

cases required multiple attempts to either reschedule a RJ conference or refer the case for court processing.[79]

**RESULTS**

Table 3.2 presents the descriptive bivariate relationship between our explanatory variables and treatment non delivery, whether for offender-related or administrative-related reasons.[80] Male offenders are more likely to have a treatment failure than female offenders (nearly 30% vs. 13%, respectively). We can see that offenders with longer offending career lengths and who are older are slightly more likely to have a treatment failure, compared to offenders with shorter offending careers and who are younger (approximately 28% to 24%, respectively). The offenders attitude at the time of arrest appears to be related to treatment delivery, as well. Those offenders rated as having a "good" attitude were less likely to have a treatment failure than offenders rated as having a "bad" or "neutral" attitude (24% to 33%, respectively).

The season of the year in which the case was randomly assigned appears related to treatment non-delivery, specifically summer cases. Nearly forty percent of summer cases assigned to RJ experienced a treatment failure, with a quarter or less of cases assigned at other times of the year experiencing treatment failure. It is possible, given the number of people involved, and summer-time travel, that RJ conferences are more difficult to organize. The offense that got the offender referred to RISE appears related to treatment failure. If the offense was a personal property crime (JPP), the offender was more likely

---

[79] In cases where no treatment was scheduled or where there was "No Further Action" WIT is set to 56, representing a longer time period than the time to the longest completed conference attempt. When "No treatment" cases are excluded, the mean "weeks to initial treatment" reduces 12.4 weeks (with a median of 9 weeks).

[80] Continuous variable have been dichotomized around the median value.

to have treatment failure (34%) than if the offense had been either a personal victim offense (JVC, 21%) or shoplifting (JPS, 17%). Finally, the length of time it takes for the initial attempt at an RJ conference seems to be related to treatment failure. Cases taking more than ten weeks fail to have a completed conference than cases where the initial attempt took less time (40% to 13%, respectively). The bivariate examination of our explanatory variables reveals each is related to the chance of a RJ conference non-delivery. We include these variables in multivariate analysis to determine if these relationships hold after controlling for other characteristics.

**Analytic Strategy**

As our outcome variable is nominal with 3 categories, we will utilize multinomial logistic regression to estimate the effect of our explanatory variables on the final treatment. Rather than estimating separate binary logistic regressions for all possible comparisons, with multinomial logistic regression we conduct this process simultaneously.[81] We have ordered our outcome in such a way that each separate nominal logistic regression provides a comparison between each of our "failed" RJ outcomes with treatment as assigned (i.e. a RJ conference). As our purpose is exploring the factors associated with treatment failure, these comparisons make the most sense. To provide an illustration of the influence that the explanatory variables have on the outcome categories, we provide visualizations of the predicted probabilities across the levels of our predictors.

---

[81] As our outcome of interest is nominal, with more than two categories, standard logistic regression is not appropriate. See Long (1997) for a detailed description of multinomial logistic regression.

We present the results of our multinomial logistic regression in Table 3.3.[82] Initially we discuss the relationship between a successful[83] RJ conference and a treatment failure due to offender reasons. When controlling for our other variables, the offender characteristics (sex, age, attitude, offending career) are not related to the relative risk of a treatment non-delivery relative to a successful conference. Although males are over 4 times more likely not to have a successful conference, this relationship is non-significant after controlling for our other characteristics. Although having a good attitude (as rated by the arresting officer) appears to be related to a lower relative risk (18% lower) of treatment failure than having a bad/neutral attitude, this is not significant at conventional levels. An offender's offending frequency in the year prior to random assignment is only marginally significant. Each increase in an offense in the year prior to random assignment reduces the odds of a treatment failure by 41 percent. Turning our attention to non-offender related characteristics (weeks to initial treatment attempt, offense type, and season) we see the instant offense is not related to treatment failure, after controlling for the other variables. In addition, the season of the year in which the instant offense occurred does not affect the odds of a treatment failure due to offender related reasons. Finally we see that the length of time between random assignment and the initial treatment attempt increases the relative risk of a treatment failure.[84] Although the log of our linear term is negative, our quadratic term is significant and positive, indicating the

---

[82] Diagnostics for our model are presented in the Appendix.
[83] Here "success" is measured in terms of a completed RJ conference event, not to the quality of that conference or its impact on future outcomes.
[84] Recall if a conference had no treatment attempt recorded, the value for WIT was set to 56. Some might argue this would be the reason for the observed relationship between WIT and treatment failure. However, when the 32 cases (9 failing for offender related reasons; 23 failing for administrative-related reasons) are removed, the observed relationship persists leading us to conclude that the longer it takes to initially organize a conference is related to treatment non-delivery.

relationship between WIT and offender-related treatment failure is U-shaped. It appears that some time between offense and initial treatment is advantageous (in terms of increasing the odds of a successful conference); however, this relationship turns to an increasing risk of offender-related treatment failure as the weeks progress. We will return to this later with an illustration.

When we examine treatment non-delivery due to administrative-related reasons, we see a similar picture to the results we discussed in the previous paragraph. The number of weeks between random assignment and initial treatment attempt demonstrates a similar U-shaped relationship to treatment non-delivery.[85] Some lag between assignment and treatment increases the chance of a successful conference. This relationship turns as time progresses to an increasing risk of treatment failure for administrative reasons. Although the offending frequency in the year prior to random assignment is not significantly related to treatment non-delivery, the length of one's offending career is. Offenders with longer careers are more likely to fail for administrative reasons. Each one year increase in one's offending career increases the relative risk of treatment failure for administrative reasons by 28 percent. As for offender-related failure, none of our other offender characteristics (sex, age, "good" attitude, offending frequency) are significantly related to treatment non-delivery for administrative reasons. In addition, the season in which the offense (to gain entry into the experiment) was committed is not related to treatment non-delivery. Finally, the instant offense, i.e. the offender committed to get into RISE, was not significantly related to treatment non-delivery.

---

[85] See note 84.

**An Illustration**

Our regression coefficients (when exponentiated) give us the impact of each explanatory variable on the relative risk of each type of treatment failure to a successful RJ conference. However, for illustrative purposes we can determine the probability of our outcomes utilizing the following formula for each outcome[86]:

$$P = \frac{1}{1 + e^{-(a+bX)}}$$

We present our illustration in Figure 3.1.[87] Initially, we have set continuous measures (age, career length, and offending frequency in the year prior to random assignment) to their mean value, while setting categorical measures (season, sex, offense type, and attitude) to their modal response. We will adjust the wait time to initial treatment. For our illustration we focus on a male 16 year old personal property offender with a good attitude at the time of arrest. This "hypothetical" individual further committed the crime that referred him to the RISE experiments in the Fall and had a "wait time" of 17 and a half weeks until the initial treatment. We would predict a successful conference for this individual with a probability of .83, with a probability of administrative-related failure and offender-related failure of .04 and .13, respectively. If the wait time to initial treatment is increased to twenty four weeks, the predicted probability of successful treatment reduces to .75. The predicted probability of offender-related failure increases to .20; while the probability of administrative-related failure remains essentially unchanged.

---

[86] As the probabilities sum to one, when we know the probability of each outcome, we can determine the probability of our reference outcome (successful RJ conference).

[87] We should note that this converts the linear relationship observed in the original output to a non-linear relationship (Berk, personal communication). As a result the shape of the predicted relationship between an individual variable and the outcome will depend on the values of the other variables. Typically in these scenarios, the other variables are set to their mean values. This strategy presents some interpretational difficulties when the other variables are categorical. We will adopt a different strategy and note we only do this for illustrative purposes.

However, if the wait time to initial RJ conference attempt is reduced to four weeks, the predicted probability of a treatment success is about .73, while the predicted probability of an offender-related non-delivery decreases to .09. However, we see that the predicted probability of an administrative-related failure increases to nearly .18. In general, the probability of a completed RJ conference fluctuates as the length of time it takes to organize that initial RJ conference attempt changes. We expand on this in our discussion.

## DISCUSSION & CONCLUSIONS

The organization of a RJ conference is not an easy task. The facilitator has to coordinate the schedules of offender(s), victim(s) and their respective supporters. Even when a conference is organized, "interested parties" may fail to appear or show up intoxicated. A RJ facilitator may not perceive it to be worth the time and effort to reorganize another conference and either pass the case to the court or "caution" the offender. Difficulties aside, the longer it takes to organize the initial attempt, the less likely it is to reach a successful outcome.[88] That said, on the other hand, an initial conference attempt that is too soon is more likely to result in an administrative-related failure than if the initial attempt is later.

Twenty years ago, Berk et al. (1988) cautioned that random assignment was often less than ideal in the context of randomized field trials. Treatments can "fail" because of characteristics of the individual participants or structural characteristics of the treatment. Random assignment to treatment can help us analyze the impact of assigning individuals to participate in a treatment (i.e. the effect of the policy). In the case of perfect

---

[88] Recall this relationship held up even when we removed cases where there was not a treatment attempt.

compliance, random assignment also helps us answer what is the impact of the treatment on those treated.

Non-compliance does not invalidate the results of an experiment. In the case of serious non-compliance, ITT analysis provides less clear answers. Given that the problems of non-compliance should be well known, researchers should not be surprised when it occurs. Within the context of a randomized controlled trial, treatment assignments can be monitored and evidence of non-compliance can be addressed.

Berk et al. (1988) stated “prevention is always the best strategy” (p. 62). If random assignment is monitored, and obstacles to treatment integrity are discussed with key players in the research study, the need for complicated post-hoc statistical techniques is minimized. However, if these obstacles cannot be overcome, we do have statistical techniques available to make adjustments to our treatment impact estimates.

The RISE experiments had a set of eligibility criteria which were covered before cases were randomly assigned. Periodically cases were reviewed to establish whether the randomly assigned treatments were delivered. Even in the context of careful checks, we see that conferences are not delivered for a variety of reasons. Of the variety of variables available to us, few were involved in treatment failure for either offender non-compliance or administrative failure. Although the majority of cases within the RISE experiments received the assigned treatment, RISE researchers met with resistance to “fixing” the problem of non-treated cases.

Offender characteristics were less involved after controlling for structural/administrative characteristics. However, the length of one’s offending career was marginally related to treatment non-delivery. With the size of our sample, it would

be interesting to know if this result could be replicated in other settings. Those offenders with longer offending careers had an increasing probability of treatment non-delivery. In the future, if the goal is to maximize the delivery of treatment, the length of one's offending career could either serve as an exclusion or as an indicator that the particular case may require more effort.

We have demonstrated that the time between random assignment and initial treatment is a key window in which to achieve a successful treatment. Although a conference with multiple victims, offenders, and supporters may be more difficult to arrange, it is important for this process to begin as early as possible. An important lesson for policy makers and practitioners is "expediency is the key." If the goal is successful delivery of RJ conferences, we have provided evidence that a shorter lag between the crime and treatment, possibly with a "cooling off" period after the event increasing the chance that offenders and victims will participate, increases the chance of a successful conference delivery. Ensuring a successful conference will involve additional resources to increase the probability of an "early" conference, but in the end, time spent early will produce greater success.

Table 3.1: Descriptive Statistics for Analysis Variables:

| **Outcome** | | % | n | | |
|---|---|---|---|---|---|
| Final Treatment | Treated as Assigned | 73.9 | 195 | | |
| | Offender Non-compliance | 11.7 | 31 | | |
| | Administrative Factors | 14.4 | 38 | | |
| **Explanatory (Categorical)** | | | | | |
| | | % | n | | |
| Offender Sex | Male = 1 | 79.5 | 210 | | |
| Attitude @ Time of Arrest | Good Attitude = 1 | 72.4 | 181 | | |
| Season | Fall | 28.8 | 76 | | |
| | Winter | 27.3 | 72 | | |
| | Spring | 21.6 | 57 | | |
| | Summer | 22.3 | 59 | | |
| RISE Experiment | Juvenile Personal Property (JPP) | 47.3 | 125 | | |
| | Juvenile Shoplifting (JPS) | 29.2 | 77 | | |
| | (Youth Violence (JVC) – ref. cat.) | 23.5 | 62 | | |
| **Explanatory (Continuous)** | | Mean | Median | MIN | MAX |
| Offender Age | | 16.2 | 16.1 | 9.5 | 27.4 |
| Offending 1 yr Prior to RA | | 0.64 | 0 | 0 | 7 |
| Career Length (in years) | | 1.1 | 1.8 | 0 | 15 |
| Weeks to Initial Treatment | | 17.6 | 10.6 | 0 | 56 |

Table 3.2: Probability of Non-delivery (Offender related or Administrative) given sample characteristics

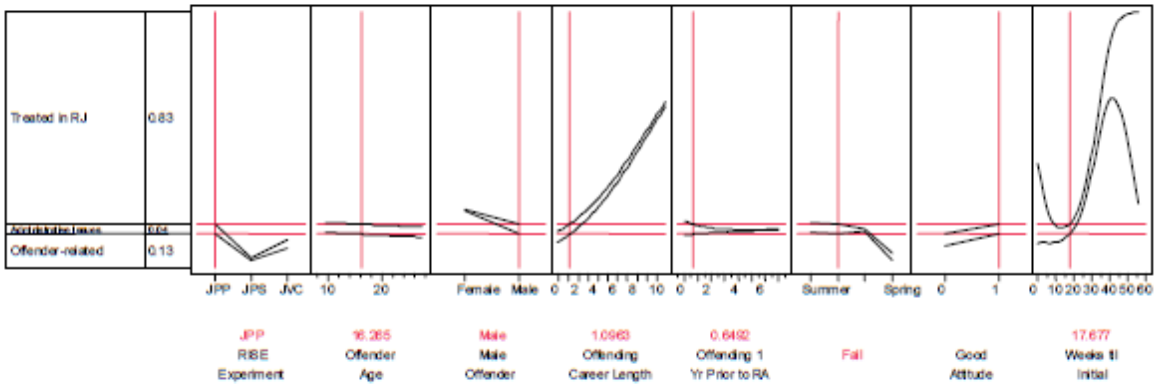
| **Explanatory Variable** | | **Percent Non-Delivery** |
|---|---|---|
| Offender Sex | Male | 29.5 |
| | Female | 13.0 |
| Attitude @ Time of Arrest | Good Attitude | 24.3 |
| | Bad or Neutral | 33.3 |
| Season | Fall | 25.0 |
| | Winter | 25.0 |
| | Spring | 15.8 |
| | Summer | 39.0 |
| RISE Experiment | Juvenile Personal Property (JPP) | 34.4 |
| | Juvenile Shoplifting (JPS) | 16.9 |
| | Youth Violence (JVC) | 21.0 |
| **Explanatory (Above Median Value)** | | **Percent Non-Delivery** |
| Offender Age | (≥ 16.1) | 28.0 |
| | (< 16.1) | 24.2 |
| Offending 1 yr Prior to RA (≥ 16.1) | No Priors | 25.3 |
| | 1 Prior | 22.4 |
| | 2 or More Priors | 36.1 |
| Career Length (in years) | (≥ 1.8) | 28.6 |
| | (< 1.8) | 23.7 |
| Weeks to Initial Treatment | (≥ 10.6) | 40.0 |
| | (< 10.6) | 12.9 |

Table 3.3: Multinomial regression results for non-treatment

| | | | | |
|---|---|---|---|---|
| | **Number of Observations = 248** | | | |
| | **Wald chi-square(24) = 69.97** | | | |
| | **Prob > chi-square = .0000** | | | |
| **Log pseudo likelihood = -108.0077** | **Pseudo R2 = .4343** | | | |
| **Term** | **Relative Risk Ratio** | **Robust Standard Error[89]** | **(z)** | **(p)** |
| **Offender-related Failure** | | | | |
| [Male-Female] | 4.53 | 6.21 | 1.10 | 0.270 |
| Offender Age | 1.00 | 0.10 | -0.01 | 0.993 |
| Offending 1 Yr Prior to RA | 0.59 | 0.18 | -1.76 | 0.079 |
| Good Attitude | 0.82 | 0.52 | -0.32 | 0.750 |
| Offending Career Length | 1.00 | 0.15 | -0.02 | 0.981 |
| Weeks to Initial Treatment Attempt (WIT) | 0.76 | 0.07 | -2.79 | 0.005 |
| WIT^2 | 1.01 | 0.00 | 3.93 | 0.000 |
| [Summer] | 1.78 | 1.96 | 0.52 | 0.603 |
| [Fall] | 1.77 | 1.77 | 0.57 | 0.570 |
| [Winter] | 0.51 | 0.55 | -0.63 | 0.531 |
| [JPP] | 1.37 | 1.26 | 0.34 | 0.733 |
| [JPS] | 0.36 | 0.31 | -1.17 | 0.243 |
| **Administrative-related Failure** | | | | |
| [Male-Female] | 0.55 | 0.43 | -0.76 | 0.445 |
| Offender Age | 0.99 | 0.11 | -0.08 | 0.939 |
| Offending 1 Yr Prior to RA | 1.02 | 0.18 | 0.10 | 0.921 |
| Good Attitude | 1.62 | 1.06 | 0.74 | 0.462 |
| Offending Career Length | 1.28 | 0.15 | 2.06 | 0.039 |
| Weeks to Initial Treatment Attempt (WIT) | 0.93 | 0.08 | -0.89 | 0.372 |
| WIT^2 | 1.00 | 0.00 | 2.37 | 0.018 |
| [Summer] | 5.17 | 6.31 | 1.34 | 0.179 |
| [Fall] | 4.93 | 5.65 | 1.39 | 0.164 |
| [Winter] | 5.19 | 5.28 | 1.62 | 0.106 |
| [JPP] | 1.83 | 1.47 | 0.75 | 0.455 |
| [JPS] | 0.34 | 0.32 | -1.16 | 0.247 |

[89] We have utilized robust standard errors using the cluster option in STATA 10. As the unit of assignment was the case rather than the offender the standard errors between cases may be independent, but within case they might not. Therefore we account for clustering within the RISE case.

Figure 3.1: Prediction Profiler from Multinomial Logistic Regression Results



# Appendices

## APPENDIX A: FIT STATISTICS FOR THE MULTINOMIAL LOGISTIC REGRESSION MODEL

| Fit Statistics for Multinomial Logistic Regression: | | | |
|---|---|---|---|
| Log-Likelihood Intercept Only: | -190.942 | Log-Likelihood Full Model: | -108.008 |
| D(222): | 216.015 | LR(24): | 165.869 |
| | | Prob > LR: | 0.000 |
| McFadden's R2: | 0.434 | McFadden's Adj R2: | 0.298 |
| Maximum Likelihood R2: | 0.488 | Cragg & Uhler's R2: | 0.621 |
| Count R2 | 0.827 | Adj Count R2: | 0.358 |
| AIC: | 1.081 | AIC*n | 268.015 |
| BIC: | -1007.966 | BIC*n | -33.546 |

## APPENDIX B: TEST OF THE INDEPENDENCE OF IRRELEVANCE ASSUMPTION (IIA)

Multinomial logistic regression models assume that the odds for each pair of outcomes do not depend on the inclusion of the other available outcomes.[90] Therefore, deletion of any outcome should not affect the odds of the remaining outcomes, which would indicate dependence within the outcomes. For example, the deletion of "Offender-related Failure" should not affect the odds ratio of "Administrative-related Failure" to RJ. Our test indicates we cannot reject the null hypothesis of independent alternatives.

| suest-based Hausman tests of IIA assumption | | | |
|---|---|---|---|
| Omitted Category | Chi-square | df | (p) |
| RJ | 9.96 | 13 | 0.697 |
| Offender-related Failure | 9.99 | 13 | 0.695 |
| Administrative-related Failure | 10.72 | 13 | 0.634 |

[90] Sarkisian, N. (2009). http://www.sarkisian.net/sc704/mlogit.pdf (Accessed August 24th, 2009)

## APPENDIX C: LIKELIHOOD-RATIO TEST FOR COMBINING ALTERNATIVES

We test whether we can combine the two non-delivery outcomes into one category. Specifically this LR Test examines all possible combinations, but for our purposes it makes sense to focus only on the two non-delivery options. The null hypothesis in these tests is that all coefficients associated with a given pair of outcomes are zero. If we fail to reject the null hypothesis, we have evidence the outcomes can be combined into one. In each of our tests we reject the null hypothesis, leading to the conclusions that coefficients associated with each pair of outcomes are not equal to zero. Therefore, it does not make sense to combine outcomes.

| LR Test for combining alternatives | | | |
|---|---|---|---|
| | Chi-squre | df | (p) |
| RJ vs. Offender-related Failure | 122.53 | 12 | 0.000 |
| RJ vs. Administrative-related Failure | 79.58 | 12 | 0.000 |
| Offender-related vs. Admin-related Failure | 30.49 | 12 | 0.002 |

## APPENDIX D: VARIANCE INFLATION FACTORS OF THE INCLUDED VARIABLES

Sarkisian (2009) recommends performing an OLS regression with the variables included in the multinomial logistic regression in order to assess multicollinearity among the variable by computing variance inflation factors (VIF). The rule of thumb is a VIF in excess of 4 would indicate a high degree of correlation between the independent variables. From the table below, we do not see a problem VIFs, with the exception of the linear and quadratic "Weeks to Initial Treatment" variable. However, these terms are naturally going to be correlated and are not seen as an issue.

| Variable | VIF |
|---|---|
| Weeks to Initial Treatment Attempt (WIT) | 20.74 |
| WIT^2 | 20.57 |
| [JPS] | 2.09 |
| [Winter] | 2.08 |
| [JPP] | 2.06 |
| [Fall] | 1.86 |
| [Summer] | 1.82 |
| Offender Age | 1.32 |
| [Male-Female] | 1.20 |
| Offending Career Length | 1.20 |
| Offending 1 Yr Prior to RA | 1.16 |
| Good Attitude | 1.11 |

# Bibliography:

Chapter 1:

Braithwaite, J. B. (1989). Crime, Shame and Reintegration. Cambridge: Cambridge University Press

Devereaux, P.J., M. Bhandari, M. Clarke, V. M. Montori, D. J. Cook, S. Yusuf, D. L. Sackett, C. S. Cina, S. D. Walter, B. Haynes, H. J. Schunemann, G. R. Norman & G. H. Guyatt (2005). "Need for expertise based randomized controlled trials." *British Medical Journal* , 330(7482): pp. 88-92.

Hayes, H. and Daly, K. (2003) 'Youth justice conferencing and reoffending', *Justice Quarterly* 20(4), pp. 725-764.

Hipple, N. K. (2003). Police officer facilitated versus civilian facilitated family group conferences: Does it matter who the facilitator is? Unpublished Doctoral dissertation. University of Indiana, Indianapolis.

Knight, J. (1999). "Having their say - The role of victims in diversionary conferencing." Paper presented at the Restoration for Victims of Crime: Contemporary Challenges conference. Melbourne, Australia.

Luft, H. S., J. P. Bunker, and A. C. Enthoven (1979). "Should operations be regionalized? The empirical relation between surgical volumes and mortality." *New England Journal of Medicine*, 301(25): 1364-1369.

Luft, H. S., S. S. Hunt, & S. C. Maerki (1987). "The volume-outcome relationship: Practice-makes-perfect or selective-referral patterns." *Health Services Research*, 22(2): 157-182.

Marshall, T. (1999). *Restorative Justice: An overview*. London: Home Office, Research, Statistics, & Development Directorate.

Sherman, L. (1993). "Defiance, deterrence, and irrelevance: A theory of the criminal sanction." *Journal of Research in Crime and Delinquency* 30(4): 445-473.

Sherman, L., H. Strang with G. Barnes, S. Bennett, C. Angel, D. Newbury-Birch, D. Woods, & Gill, C. (2007). "Restorative Justice: The evidence." London: The Smith Institute.

Sherman, L. W., H. Strang, H., & D. J. Woods (2000). "Recidivism in the Canberra Reintegrative Shaming Experiments (RISE)." http://www.aic.gov.au/rjustice/rise/recidivism/index.html.

Strang, H., G. C. Barnes, J. Braithwaite & L. W. Sherman (1999). "Experiments in restorative policing: a progress report on the Canberra Reintegrative Shaming Experiments (RISE)." http://aic.gov.au/rjustice/rise/progress/1999.html

Tyler, T. (1990). Why People Obey the Law. New Haven: Yale University Press.

Tyler, T., L. Sherman, H. Strang, G. Barnes & D. Woods. (2007). "Reintegrative shaming, procedural justice and recidivism: The Engagement of offenders' psychological mechanisms in the Canberra RISE Drinking-and-Driving Experiment." *Law & Society Review*, 41(3), 553-586.

Chapter 2:

Apel, R., R. Paternoster, S. Bushway, & R. Brame (2006). "A job isn't just a job: The differential impact of formal versus informal work on adolescent problem behavior." *Crime and Delinquency*, 52(2):333-369

Bushway, S.; T. P. Thornberry & M. Krohn (2003). “Desistance as a developmental process: A comparison of static and dynamic approaches.” *Journal of Quantitative Criminology*, 19(2):129-153.

Farrington, D. P. (2006). “Key longitudinal-experimental studies in criminology.” *Journal of Experimental Criminology*, 2: 121-141.

Haviland, A. M. & D. S. Nagin (2005). “Causal inferences with group based trajectory models.” *Psychometrika*, 70(3): 1-22.

Jones, B. L & D. S. Nagin (2007). “Advances in group-based trajectory modeling and an SAS procedure for estimating them.” *Sociological Methods and Research*, 35(4): pp. 542-571.

Knight, J. (1999). “Having their say - The role of victims in diversionary conferencing.” Paper presented at the Restoration for Victims of Crime: Contemporary Challenges conference. Melbourne, Australia

Lacourse, E., S. Cote, D. S. Nagin, F. Vitaro, M. Brendgen, & R. E. Tremblay (2002). “A longitudinal-experimental approach to testing theories of antisocial behavior development.” *Development and Psychopathology*, 14: 909-924.

Laub, J. H., D.S. Nagin, & R. J. Sampson (1998). “Trajectories of change in criminal offending: Good marriages and the desistance process.” *American Sociological Review* 63(2): 225-238.

Moffitt, T. E. (1993). “Adolescent-Limited and Life-Course Persistent antisocial behavior: A developmental taxonomy.” *Psychological Review*, 100(4): 674-701.

Muthen, B., C. H. Brown, K. M. Booil Jo, S. Khoo, C. Yang, C. Wang, S. G. Kellam, J. B. Carlin, & J. Liao (2002). "General growth mixture modeling for randomized preventive interventions." *Biostatistics*, 3(4): 459-475.

Nagin, D. S. (2005).Group-Based Modeling of Development. Harvard University Press: Boston.

Nagin, D. S. & K. C. Land (1993). "Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model." *Criminology*, 31(3): 327-362.

Nagin, D. S., L. Paganini, R. E. Tremblay & F. Vitaro (2003). "Life course turning points: The effect of grade retention on physical aggression." *Development and Psychopathology*, 15: pp. 343-361.

Sherman, L. (1993). "Defiance, deterrence and irrelevance." *Journal of Research in Crime and Delinquency*, 30, pp. 445-473.

Sherman, L., H. Strang with G. Barnes, S. Bennett, C. Angel, D. Newbury-Birch, D. Woods, & Gill, C. (2007). "Restorative Justice: The evidence." The Smith Institute.

Sherman, L. W., H. Strang, G. Barnes, & D. J. Woods (2006). "Race and Restorative Justice: Systematic observations and intention to treat results in the Canberra RISE experiments." Panel presentation at the 65th annual meeting of the American Society of Criminology.

Sherman, L. W., H. Strang, & D. J. Woods (2000). "Recidivism in the Canberra Reintegrative Shaming Experiments (RISE)." http://www.aic.gov.au/rjustice/rise/recidivism/index.html.

Strang, H. (2003). Repair of Revenge: Victims and Restorative Justice. Oxford University Press: USA.

Strang, H., G. C. Barnes, J. Braithwaite & L. W. Sherman (1999). "Experiments in restorative policing: a progress report on the Canberra Reintegrative Shaming Experiments (RISE)." http://aic.gov.au/rjustice/rise/progress/1999.html

Woods, D. J. (2003). "Self-reported vs. official offending in the Reintegrative Shaming Experiments (RISE). (Master's Thesis, University of Maryland).

Chapter 3:

Berk, R. A. & H. Xu. (2003). "Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City: comment" *Journal of the American Statistical Association*, 98(462), pp. 318 – 320.

Berk, R. A., G. K. Smyth, & L. W. Sherman (1988). "When random assignment fails: Some lessons from the Minneapolis Spouse Abuse experiment." *Journal of Quantitative Criminology*, 4(3), pp. 209 – 223.

Knight, J. (1999). "Having their say - The role of victims in diversionary conferencing." Paper presented at the Restoration for Victims of Crime: Contemporary Challenges conference. Melbourne, Australia.

Long, J. Scott. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oakes, CA: Sage.

Marshall, T. (1999). *Restorative Justice: An overview*. London: Home Office, Research, Statistics, & Development Directorate.

Sarkisian, N. (2009). "Topics in multivariate statistics: Multinomial logit." http://www.sarkisian.net/sc704/mlogit.pdf (Accessed August 24th, 2009)

Sherman, L. W. & H. Strang. (2004). "Verdicts or inventions? Interpreting results from randomized controlled experiments in criminology." *American Behavioral Scientist*, 47(5), pp. 575 – 607.

Sherman, L., H. Strang with G. Barnes, S. Bennett, C. Angel, D. Newbury-Birch, D. Woods, & Gill, C. (2007). "Restorative Justice: The evidence." The Smith Institute.

Strang, H., G. C. Barnes, J. Braithwaite & L. W. Sherman (1999). "Experiments in restorative policing: a progress report on the Canberra Reintegrative Shaming Experiments (RISE)." http://aic.gov.au/rjustice/rise/progress/1999.html