

Protocol for Efficacy Trials- Crim-PORT

The following protocol follows the format recommended by Crim-PORT 1.0: *Criminological Protocol for Operating Randomized Trials* (@ 2009 by Lawrence W. Sherman and Heather Strang). Crim-PORT has been used by the Cambridge Institute of Criminology Registries of Randomized Trials, which is in the process of being transferred to the American Society of Criminology's Division of Experimental Criminology. As of 12th November 2018, this protocol is dually registered at the Jerry Lee Centre of Experimental Criminology at the Institute of Criminology, University of Cambridge, and at the Departments of Psychology and of Anthropology at Aarhus University, Denmark.

1. Name of Experiment

Konfliktråd Impact Project (KIP)

2. Principal Investigators

1st Co-Principal Investigator: Sarah van Mastrigt, Ph.D., Aarhus University

2nd Co-Principal Investigator: Christian B. N. Gade, Ph.D., Aarhus University

3rd Co-Principal Investigator: Heather Strang, Ph.D., Cambridge Centre for Evidence-Based Policing

4th Co-Principal Investigator: Lawrence W. Sherman, Ph.D., Cambridge Centre for Evidence-Based Policing

3. General Hypothesis

Restorative justice conferences (RJC), generally conducted with supporters of both victim and offenders present, will be a more effective form of "Konfliktråd" (KR) in Denmark than current victim-offender mediation practices.

4. Specific Hypotheses

- H1: The conduct of RJC will differ from the conduct of the current mediation method of KR on a number of theoretically important dimensions.
- H2: Relative to current mediations, RJC will reduce the prevalence, frequency, seriousness, and harms of repeat offending.
- H3: Relative to the current method of mediations, RJC will produce better outcomes for victims on a number of theoretically important dimensions.

5. Organizational Framework

Dual Partnership: Operating agency delivers treatments with independent research organization providing random assignment, data collection, analysis, training, tracking feedback and support.

- Name of Operating Agency: Danish National Police, National Crime Prevention Centre (NFC)
- Name of Research Organizations: Aarhus University Departments of Psychology and of Anthropology and Cambridge Centre for Evidence-Based Policing Ltd.

6. Unit of Analysis

Eligible cases of crime which feature at least one identifiable offender and one personal victim.

7. Eligibility Criteria

A. Criteria for Inclusion

- One or more offender(s) in the case must be willing to accept responsibility for having harmed one or more of the victim(s) in the case.
- Both the offender(s) and victim(s) must understand what will happen in an RJC as well as in a KR mediation.
- Both the offender(s) and victim(s) are willing to accept random assignment to either RJC or KR mediation.
- The offence in the case must be a violation of the criminal codes on violence, threats, robbery, burglary, theft, and vandalism, or enumerated on a detailed list of other eligible criminal code offences available from the Aarhus University Department of Psychology at vanmastrigt@psy.au.dk
- The offence is the kind of case traditionally defined as eligible for KR under current practices and law in Denmark.

B. Criteria for Exclusion

- Either a victim or an offender is challenged by mental health issues to the extent that an RJC or mediation might cause a violent reaction.
- Any matter that is not a violation of the criminal code.
- Any criminal code matter that constitutes or relates to domestic violence, murder, attempted murder, manslaughter, sexual offences, evictions, violations of restraining orders, and all offences without a personal victim.

8. Pipeline: Recruitment or Extraction of Cases

A. Where will cases come from? Referrals of potentially eligible cases can come from any of the following sources:

- Police officers or staff who refer a case to the District KR Coordinator, including patrol officers and investigators.
- Victims or offenders who independently contact the Police to request a KR process.
- Relatives or friends of offenders or victims.
- Professionals such as teachers, social workers, and heads of institutions.
- District KR Coordinators themselves, based on review of daily crime reports and other information systems available to police.

B. Who will obtain them? The KR Coordinators in each Police District are responsible for identifying the eligible cases.

C. How will they be identified? See A and B above.

D. How will each case be screened for eligibility? Once a case has been identified as eligible by the District KR Coordinator, and consent from victim and offender has been obtained (see below), the case will be forwarded to the site managers or Co-Principal Investigators working at Aarhus University. The latter research personnel will perform a final check for eligibility.

E. Who will register the case identifiers prior to random assignment? The site managers or Co-Principal Investigators working at Aarhus University.

F. What social relationships must be maintained to keep cases coming? It is essential that positive and close relations among Coordinators, mediators/facilitators and site managers be maintained. The goal is to build a team that takes shared pride in sustaining the case flow needed to provide an adequate sample size for the experiment.

9. Has a Phase I (no-control, “dry-run”) Test of the Pipeline and Treatment Process Been Conducted?

Yes, in each of the seven test districts:

- All RJC facilitators have conducted at least one observed RJC prior to the start of random assignment.
- A total of 25 “practice” RJC were observed, including at least one case in every district.

10. Timing

Cases come into the experiment by a trickle-flow process, one case at a time.

11. Random Assignment

A. How is random assignment sequence to be generated? Each random assignment of a case to an RJC facilitator or a KR mediator will be generated by a customized randomization module in REDCap, as follows:

- The assignment will be generated by a random numbers case-treatment generator program in the secure, password-protected REDCap system at Aarhus University on a 50% allocation to each treatment arm, in block random assignment for each of the 7 participating police districts.

B. Who is entitled to issue random assignments of treatments? Site managers and Co-Principal Investigators working at Aarhus University only.

C. How will random assignments be recorded in relation to case registration? The REDCap system will instantaneously record the assigned treatment as soon as it is generated, making it part of the same case record that will track whether the treatment is delivered as intended and other information on outputs and outcomes, all to be entered by the site managers.

12. Treatment and Comparison Elements

A. Experimental or Primary Treatment (Restorative Justice Conferences)

1. What elements must happen?

Element A: Each “full” RJC should be conducted with at least one victim, one offender, and one or more “supporters” of each victim or offender. If the full RJC standards are not met, the RJC will proceed as if the RJC standards had been met, but the details of how the fidelity to the RJC theory was incomplete will be recorded by a site manager or other responsible observer, who will generally be present at all randomly assigned RJCs.

Element B: Each RJC should produce a restorative plan for what the offender will do to try to make up for the harm caused, or to prevent any further crime in the future. If this standard is not met, the RJC will proceed as if the standard had been met, but the fact that fidelity to the RJC theory was violated will be recorded by a site manager or other responsible observer, who will generally be present at all randomly assigned RJCs.

Other Elements: The site manager must record in the REDCap observation record:

- Whether at least one offender has apologized for causing harm to at least one victim
- The number of minutes the parties were in the conference room, from start to finish.
- Whether there was any exposure of the discussion to people not invited to the meeting, who may have been listening through open doors or walking through the meeting space.

- Other elements of the KR events as specified in the observation instruments, a list of which is available from Aarhus University at vanmastrigt@psy.au.dk.

2. What elements must *not* happen?

Element A: The RJC Facilitator should:

- not offer opinions
- not dominate the discussion

At the end of each RJC event, the site manager or other responsible observer will make a global rating of whether the RJC has met the overall standards of the model (requiring a rating of 3 or higher on a 5-point scale).

B. Control Treatment (Mediation as usual in Denmark)

1. What elements must happen?

Element A: All KR mediations should have at least one offender and at least one victim present with a mediator. If this standard is not met, the mediation will proceed as if the standard had been met, but the fact that fidelity to the mediation theory was violated will be recorded by the site manager or other responsible observer, who will generally be present at all randomly assigned KR mediations.

Other elements: The site manager must record in the REDCap observation record:

- Whether at least one offender has apologized for causing harm to at least one victim.
- The number of minutes the parties were in the conference room, from start to finish.
- Whether there was any exposure of the discussion to people not invited to the meeting, who may have been listening through open doors or walking through the meeting space.
- Other elements of the KR event as specified in the observation instruments, a list of which is available from Aarhus University at vanmastrigt@psy.au.dk.

At the end of each KR mediation, the site manager or other responsible observer will make a global rating of the extent to which the mediation observed was similar to the key elements of an RJC (on a scale from 1-5).

2. What elements must not happen?

Element A: KR mediations are not restricted beyond the prohibitions and requirements of Danish law.

13. Measuring and Managing Treatments

A. Measuring

- 1. How will treatments be measured?** By observation and feedback from facilitators and mediators in unobserved cases.
- 2. Who will measure them?** Site managers or other responsible observers.
- 3. How will data be collected?** By site managers who enter the data in REDCap.
- 4. How will data be stored?** On a secure Aarhus University server.
- 5. Will data be audited?** Occasionally.
- 6. If audited, who will do it?** Co-Principal Investigators working at Aarhus University, who will also observe RJC's and KR mediations.
- 7. How will data collection reliability be estimated?** Audits.
- 8. Will data collection vary by treatment type?** No.

B. Managing

- 1. Who will see the treatment measurement data?** Site managers and Co-principal investigators working at Aarhus University, and de-identified data for other co-principal investigators.
- 2. How often will treatment measures be circulated to co-principal investigators working at Aarhus University?** Monthly.
- 3. If treatment integrity is challenged, whose responsibility is correction?** Co-principal investigators will meet to determine whether any issue is serious enough to take to the project governance board, or can be handled in less formal ways.

14. Measuring and Monitoring Outcomes

A. Measuring

- 1. How will outcomes be measured?**
 - H1: By Observations of KR events
 - H2: By accessing official police records
 - H3: By 4-week and 6-month follow-up surveys of victims & offenders
- 2. Who will gather the data for these measures?** Site managers or co-Principal investigators.

3. **How will data be collected?** For H2, criminal records will be searched for a 24-month period after the date of random assignment. Exact process to be negotiated. A list of ID numbers and an electronic sweep of the police files may be possible, and would be more efficient than a case-by-case extraction of data. For H1 and H3, the data will be generated by research team members.
4. **How will data be stored?** In the REDCap case files, and in secure registry systems.
5. **Will data be audited?** Yes.
6. **If audited, who will do it?** A Co-Principal Investigator at Aarhus University will run a second sweep to confirm the data in a 10% sample of cases.
7. **How will data collection reliability be estimated?** From the second sweep compared to the first.
8. **Will data collection vary by treatment type?** No.

B. Monitoring

1. **How often will outcome data be monitored?** For H1 and H3, the monitoring will occur at least twice a year. For H2, the monitoring of 12-month interim effects will begin 18 months after the launch of random assignment.
2. **Who will see the outcome monitoring data?** Two Co-Principal Investigators at Aarhus University.
3. **When will outcome measures be circulated to key leaders?** When adequate power has been attained to detect any effect.
4. **If the experiment finds early significant differences in H2 or H3, what procedure is to be followed?** Any significant difference in key outcomes will be presented to the governance board as soon as it is discovered. There may be cause to stop the experiment and expedite a complete cessation of the RJC, or of the KR mediations with a rollout of the RJC. This will be a decision by the National Police based on full analysis by the research team.

15. Analysis Plan

A. Which measures are primary for each of the three hypotheses?

1. **Which output measure is considered to be the primary indicator of a difference between the delivery of treatments in the experimental and comparison group (H1)?** The observers' global assessment on a scale of 1 to 5 indicating the similarity of the KR events as delivered to the RJC model as trained.
2. **Which outcome measure is considered to be the primary indicator of impact of the two treatments on offenders (H2)?** The mean Danish Crime Harm Index value

(see Andersen and Muller-Johnson 2018) for repeat offending per offender, across cases in each of the two treatment groups.

- 3. Which outcome measure is to be the primary indicator of impact of the two treatments on victims (H3)?** The PTSD-8 (see Hansen et al 2010) scores for post-traumatic stress symptoms.
 - 4. Other measures** Additional outcomes, such as victim and offenders satisfaction, will also be measured and compared for each treatment.
- B. What is the minimum sample size to be used to analyze outcomes?** Fifty cases per group for pooled analyses.
- C. Will all analyses employ an intention-to-treat framework?** Yes.
- D. What is the threshold below which the percent Treatment-as-Delivered would be so low as to bar any analysis of outcomes?** 60% of RJC's with "full" delivery as defined above.
- E. Who will do the data analysis?** van Mastrigt with Sherman.
- F. What statistic will be used to estimate effect size?** Cohen's H or D.
- G. What statistic will be used to calculate P values?** T tests.
- H. What is the magnitude of effect needed for a $P = .05$ difference to have an 80% chance of detection with the projected sample size (optional but recommended calculation of power curve) for the primary outcome measure?** See discussion below.

16. Statistical Power

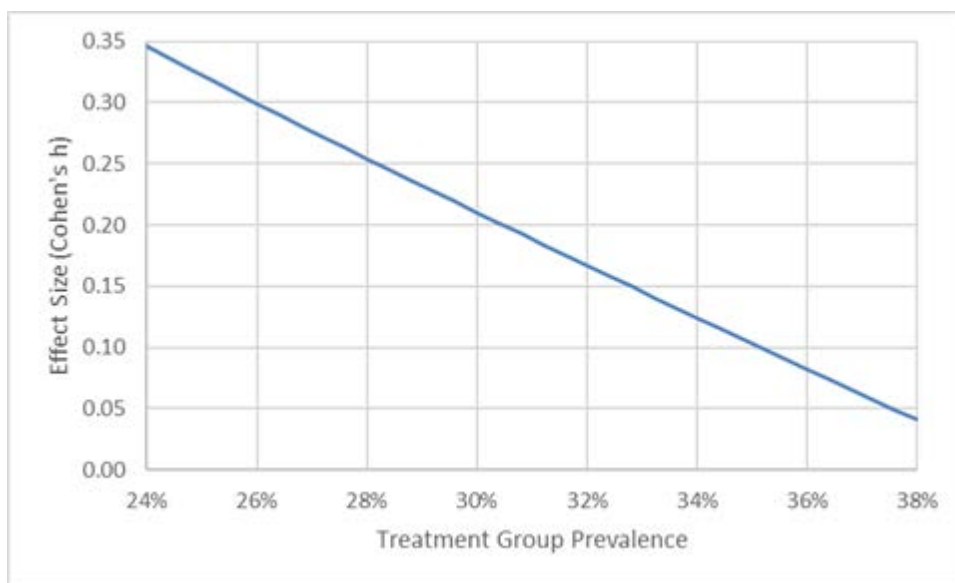
The basic plan for maximizing statistical power is to analyze the pattern of differences between RJC and mediation across the 7 participating police Districts using a meta-analytic forest graph to estimate the standardized mean difference for each hypothesis. This will be done for testing all three basic hypotheses.

As a sensitivity test, we also propose to pool all the cases across all 7 districts into the two treatment groups for calculating all three crime outcome measures for the mean offender value in each case: prevalence and frequency of repeat offending per case in a two-year follow-up period, as well as the mean Danish CHI value per event-attending offender per case based on charges filed against them (regardless of convictions).

For a power analysis of a test of effects on the prevalence of repeat offending, we begin with a modification of the dichotomous approach to allow for multiple offenders. Rather than coding each case as 0 or 1 (1 = one or more new charge in the next two years by any offender included in the KR event), we will code each case on the basis of the percentage of all offenders who have one or more new charge. If one does and one does not, the value coded would be 0.5. For three offenders, of which one has two new charges in two years, the value coded would be 0.33.

For the purpose of the following power analysis, however, we take the simplest case in which we assume (based on Kyvsgaard, 2016) that the two-year prevalence base rate of reoffending in our particular subset of all offence types included in her report is roughly 40%, meaning that 40% of those assigned to the control group will reoffend during the follow-up period with any violation of the criminal code. We also anticipate that members of the treatment group will reoffend during this same period. While we can't know exactly what rate will be produced by the treatment group, we can assume a range of possible values. These range from a rather modest reduction in prevalence (38%) to a much more substantial, but still relatively conservative estimate of 24%.

Depending on how strongly the treatment group responds to the experimental intervention, these possible reductions in reoffending correspond to a range of different effect sizes, as shown in the graph below:

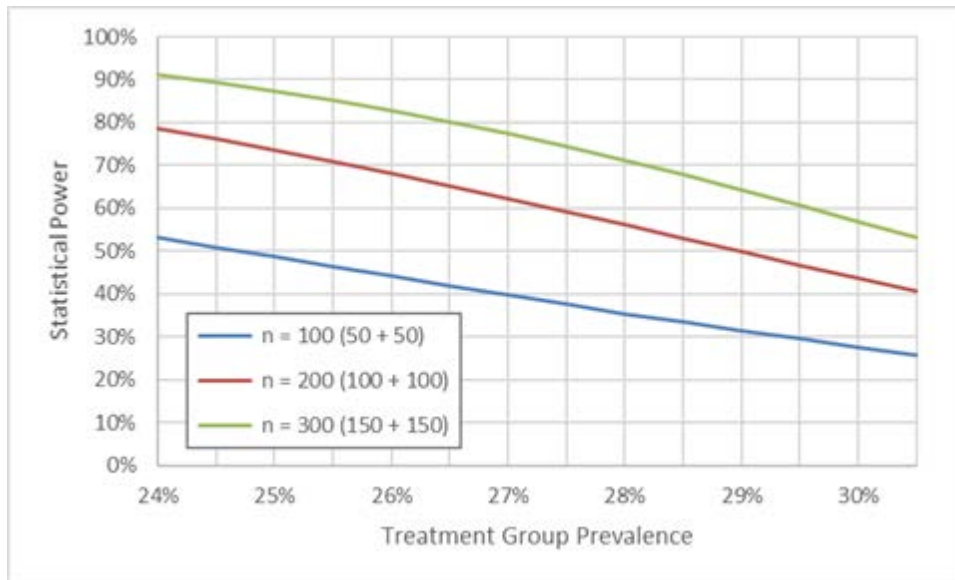


None of the possible effect sizes shown here are especially large (which is a good thing—since even small effects can have enormous benefit multiplied over thousands of cases). Cohen (1988) defines an effect size of 0.20 as “small”, while a value of 0.50 would qualify as only “moderate” in size. Even under the most advantageous circumstances presented here, the possible effect size remains rather modest, and reaches a value of just 0.35. While this value may seem rather small, it is in the normal range (or even well above) observed in many criminological experiments. For example, the overall Cohen’s D in the Campbell Collaboration Systematic Review of the Effects of Restorative Justice Conferences (Strang et al 2013) was only .15, in comparison to the no-treatment controls. The current efficacy trials may demonstrate even less of a difference as we are comparing two kinds of meetings, as distinct from meetings vs. no meetings.

Thus for the purposes of calculating statistical power, we follow what is generally considered prudent - to plan on effect size that is “small”, at around 0.20, while also examining what might happen if the experimental treatment turns out to be even more potent than expected. At the same time, there is little point in planning for effects any smaller than this value, since they would lead to desired sample sizes so large as to be unrealistic. Our analysis

will therefore assume a recidivism base rate of 40% in the control group, and a rate of between 24% ($h = 0.345$) and 31% ($h = 0.199$) in the treatment group.

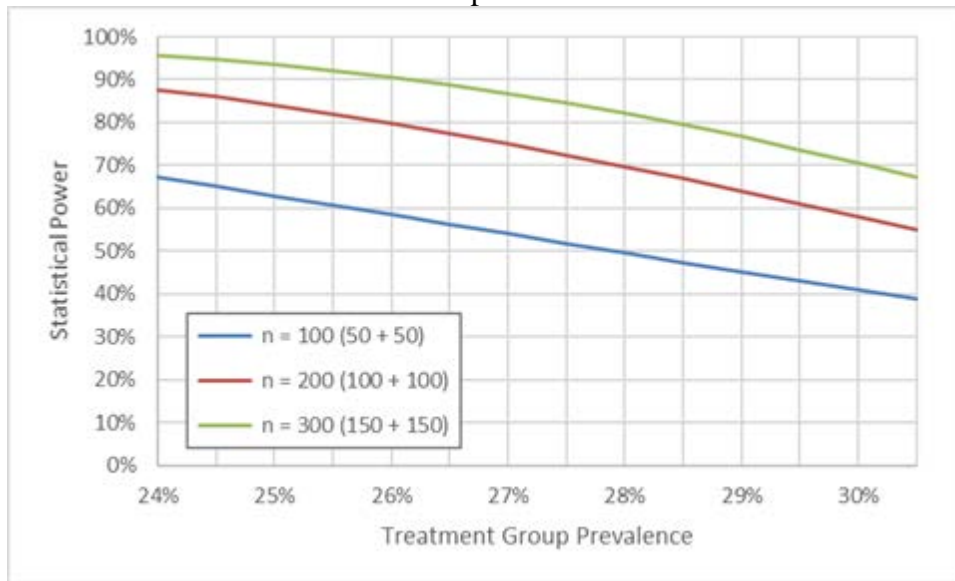
With this range of effects, the following estimates of statistical power are produced when one assumes a total sample of 100 (50 treatment + 50 control), 200 (100 + 100), and 300 (150 + 150), and when a one-tailed test is used with a $p < 0.05$ limit applied to define statistical significance:



These calculations mean that if we can obtain 300 cases across the seven block-randomized district experiments, and combine them into a single analysis of 300 cases, we can hit 80% power as long as the treatment group prevalence is as low as 26.5%, and the control group prevalence is at least 40%.

The goal for most statistical power analyses is to reach a power of at least 80%, meaning that a study will correctly identify a given treatment effect as statistically significant in at least four out of every five potential samples. In order to reach this level, the sample size must be greater than 200 offenders, even under the most advantageous assumptions. With a sample of 300 offenders, the desired level of statistical power applies to a much wider range of possible treatment effect sizes, although the treatment will still need to be capable of reducing the reoffending prevalence from its base rate of 40% to a rate of 26.5% ($h = 0.288$). To detect effects smaller than this, an even larger sample size would be required.

Statistical Power with One Tail and $p < 0.10$:



In conclusion, the proposed trials have enough power to detect about a one-third reduction in repeat offending prevalence if we can hit 300 cases and employ a one-tailed test. The latter decision can be justified by the strong evidence that RJC's reduce repeat offending, and the lack of any evidence to that effect for the KR mediations.

17. References

- Andersen, H.A., & Mueller-Johnsen, K. (2018). The Danish Crime Harm Index: How it works and why it matters. *Cambridge Journal of Evidence Based Policing*, 2(1-2), 52-69.
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.
- Hansen, M., Andersen, T.E., Armour, C., Elklit, A., Palic, S., & Mackrill, T. (2010). PTSD-8: A short PTSD inventory. *Clinical Practice & Epidemiology in Mental Health*, 6, 101-108.
- Kyvsgaard, B. (2016). *Evaluering af konfliktråd*. København: Justitsministeriets Forskningskontor.
- Strang, H., Sherman, L.W., Mayo-Wilson, E., Woods, D.J., & Ariel, B. (2013). *Restorative justice conferencing (RJC) using face-to-face meetings of offenders and victims: Effects on offender recidivism and victim satisfaction. A systematic review*. Oslo: The Campbell Collaboration.