

# The effects of custodial vs non-custodial sanctions on reoffending: Lessons from a systematic review

Martin Killias and Patrice Villettaz  
University of Lausanne (Switzerland)

Based on a systematic review of some 23 (out of 300 originally located) studies, it is concluded that most studies show lower rates of re-offending following a non-custodial compared to a custodial sanction. However, this outcome may be biased because, in most quasi-experiments of this kind, subjects with the worst prospects of rehabilitation are likely to be sent to prison. In a meta-analysis limited to five randomised controlled trials and one natural experiment, it is concluded that custodial and non-custodial sanctions do not differ significantly in terms of re-offending. Lessons from this review include, in view of future evaluations, the need to increase randomised controlled trials, to use broader measures of re-offending and rehabilitation, to look at long-term effects, and to deal with possible Hawthorn effects. Equal rates of re-offending do not mean that «nothing works» or «nothing matters», but that criminal justice policies should not be based on the belief that short-term confinement will be damaging.

*Los efectos de las penas privativas vs no privativas de libertad en la reincidencia: lecciones de una revisión sistemática.* En una revisión sistemática de 23 estudios se concluye que la mayoría de las investigaciones muestra una tasa más baja en la reincidencia de los delincuentes condenados a cumplir una condena alternativa a la prisión, comparados con los que son condenados a penas de reclusión. No obstante, esta estimación puede estar sesgada, debido al hecho de que los sujetos que tienen el peor pronóstico de rehabilitación son los que suelen ingresar en prisión. En un meta-análisis que realizamos de cinco estudios que incluían diseños experimentales con asignación al azar, más un experimento natural, obtuvimos el resultado de que las condenas de reclusión y las penas alternativas no difieren de modo significativo en cuanto a la reincidencia. El artículo a continuación deriva algunas conclusiones de la investigación, entre las cuales se citan las siguientes: la necesidad de aumentar el número de experimentos con asignación al azar, emplear medidas más inclusivas de la reincidencia y la rehabilitación, y estudiar los efectos de las sanciones a más largo plazo. Que se obtengan tasas iguales de reincidencia no significa que «nada funciona», sino que la política criminal no debería asumir que, necesariamente, la reclusión breve en la cárcel tendrá efectos negativos en el interno.

The authors have just finalised a systematic review on the effects on re-offending of custodial and non-custodial («alternative») sanctions (Villettaz, Killias, & Zoder, 2006). Since the mid-19<sup>th</sup> century, it was common knowledge, if not a dogma, that short-term imprisonment is «damaging» because, in the words first coined by Bonneville de Marsangy and later copied by von Liszt and many others, incarceration for shorter periods does not last long enough to «cure» criminal propensities (seen as a kind of a disease), but still too long to avoid first-time offenders to be exposed to the risk of contamination by hard-core criminals (Kuhn, 2000). Based on this quasi-medical theory of criminal contamination, von Liszt and many others ever since have

stimulated the development of sanctions that do not imply custody, such as suspended sentences, probation, fines and later community work and electronic monitoring. Simultaneously and rather ironically, the same movement has also stimulated long-term incarceration and even incapacitation for offenders considered as sufficiently «sick» to warrant long-term «cures» in confinement. All these trends have been stimulated by the idea to offer better «alternatives» to custody, i.e. to reduce re-offending through more efficient «alternative» sanctions. Many such programs have been evaluated over the last decades world-wide, usually with results that confirmed the superiority of non-custodial over custodial sanctions.

Keeping these backgrounds in mind, the *Campbell Collaboration Crime and Justice Group* invited the authors to start a systematic review of the evidence on whether or not custodial or non-custodial sanctions are more effective in preventing re-offending. This essay will give a resume of the methods and results of this meta-analysis. We shall conclude with an overview on how future evaluations of new sanctions and programmes could be made more convincing.

### Method

#### *What is custodial? What is non-custodial?*

According to the protocol, the systematic review was to include world-wide published or unpublished studies conducted between 1961 and 2002 where some form of «non-custodial» sanction has been compared, in terms of re-offending, with «custodial» sanctions. For pragmatic reasons, sanctions were considered «custodial» whenever they were imposed by a judge, as a form of punishment (with or without «treatment»), thus excluding pre-trial detention and arrest policies (e.g. in connection with domestic violence), and whenever they implied some form of confinement in a closed residential setting. Sanctions were considered «non-custodial» whenever they did not imply deprivation of liberty. According to our pragmatic definition, «boot camps» or sentences implying deprivation of liberty in a closed therapeutic setting have been considered as «custodial», whereas community work, electronic monitoring, financial or suspended sanctions (probation) have been considered as «alternative» or «non-custodial». Given the large international differences in crime policies, but also changes since the 1960ies in character and typologies of prisons and other sanctions, we are aware of some difficulties in this respect. Practically, however, we guess that the number of studies on sanctions at the margin of what one might consider «custodial» or not was not particularly large. More serious was the heterogeneity of «alternative» sanctions, since this concept included sanctions as different as community service, probation, fines and electronic monitoring. The only common denominator was that all these sanctions did not involve deprivation of liberty. We are aware that this is a questionable category. On the other hand, it is certainly not irrelevant for assessing the claim that confinement is damaging *per se*. Finally, the duration of custody and «alternative» sanctions can vary considerably, whereas virtually all studies included compared any «alternative» to rather short prison terms. Again, this is not a problem in a context where the century-old claim has always been that *short-term* confinement is damaging. But we recognise that one may legitimately question whether longer prison terms are equally (or more or less) damaging (a question addressed by the meta-analysis by Smith, Goggin, & Gendreau, 2002). To be eligible, studies had also to offer some results on re-offending, however measured.

#### *The quality of studies*

A far more difficult choice was to select studies according to methodological quality. Originally, we had in mind to consider only randomised or natural experiments and quasi-experiments where important intervening variables had been controlled for that went beyond what one usually finds in criminal records (such as number and kind of previous convictions and offences, plus age and gender); in particular, we thought that a certain number of studies might have considered also education, employment record and drug or alcohol abuse history. To our surprise, we rapidly realised that there were only four randomised controlled trials, one natural experiment (van der Werff 1979, 1981), but, despite hundreds of studies, almost no quasi-experiments where these additional requirements had been satisfied. This led us to abstract in full some 50 studies (out of 3,000 screened abstracts and some

300 studies that seemed, at first sight, relevant to our review), but to select only 23 for our systematic review.

This selection process raised a difficult choice between *internal* and *external* validity. Meta-analyses and systematic reviews are designed to overcome limitations to external validity of studies whose scope has been limited in time and space, as is the case of virtually all randomised controlled trials. On the other hand, the internal validity of most quasi-experiments is highly questionable in this field, since judges or correctional officers who have to decide whether a defendant is to serve his sentence in prison or in any «alternative» arrangement are regularly told, either by law or internal guidelines, to consider the defendant's risk of re-offending or, as often phrased, his «need» for treatment. In assessing such risks, they typically pay attention to many individual characteristics that go far beyond what they may find in a criminal record, namely his employment record, the family situation, his substance abuse history and many other variables that are typically hard to measure, including intuition. These factors are, later, again typically related to recidivism. In evaluations where, as usual, subjects with the worst prediction records were systematically sent to prison, while all others were eligible for an «alternative», re-offending will, almost by necessity, be more frequent among those serving their sentence in confinement. As observed by Walker, Farrington & Tucker (1981) 25 years ago, such differences (in favour of non-custodial sanctions) typically vanish the more independent variables are being controlled. Since such controls rarely include variables beyond what can be found in criminal records, as mentioned, much variance in re-offending regularly remains unexplained, often beyond 70 and up to 80 percent, a fact that points to a wide array of uncontrolled factors in recidivism. In a situation like this, including in a meta-analysis quasi-experimental studies whose outcomes tend to be systematically biased in one direction will likely reproduce the same bias. If systematic reviews increase external validity, they should obviously not achieve this at the costs of internal validity. Therefore, it has been decided to restrict the meta-analysis to four randomised controlled experiments and the one natural experiment whose internal validity seems beyond doubt.

### Results

If all the 23 studies with at least 4 control variables were «counted», the result was clearly favourable to non-custodial sanctions. In 11 out of 27 possible comparisons, re-offending was significantly more frequent after a custodial sanction, whereas the opposite outcome was observed only in 2 cases. Thus, the «vote count» method clearly favoured «alternative» sanctions over imprisonment.

In the case of the meta-analysis, however, no significant mean standardised effect size has been observed, although the trend has been slightly in favour of «alternative» sanctions again.

Only one among the selected studies (Killias, Aebi, & Ribeaud, 2000) contained information on rehabilitation beyond re-convictions, such as employment or substance abuse history or family constellations. It did not confirm claims that short custodial sanctions negatively affected subjects' integration at work or in the family.

This result does not necessarily mean that «nothing works» or that «nothing matters». It simply means that custodial and non-custodial sanctions do not differ very much in their effects on re-

offending, and that claims about «damaging» effects of short-term confinement may have been overstated. For policy-makers, this result is rather good news in the sense that they can choose between custodial or non-custodial sanctions based on other considerations. For example, «alternative» sanctions may be less expensive or, as in the case of community work, more beneficial to the community in other respects. On the other hand, custodial sanctions may have symbolic functions (in terms of higher «equity» and «equality») that are not necessarily involved in «softer» responses, a fact that may have its bearing in connection with domestic violence or traffic offences.

Finally, it may be important to keep in mind that all this holds for short-term imprisonment, i.e. sentences of a few weeks or months. Long-term deprivation of liberty may have effects that go far beyond what has been observed in the selected studies.

### Discussion

Our systematic review of studies having compared re-offending rates following custodial and non-custodial (“alternative”) sanctions has allowed to identify a number of shortcomings that might be relatively easy to overcome in the future, no matter in what country an «alternative» program is to be evaluated. Such lessons will be paid due attention to in the following discussion.

#### *Increasing the number of randomised controlled trials*

Looking at evaluations in Europe in the field of criminal justice, the first and most obvious conclusion concerns the lack of controlled experiments that, for the time being, remain rare exceptions, often promoted by dedicated researchers or policy makers who, for whatever reasons, are committed to «objective» results rather than to «proof» that their policies worked. The result of this situation is the impossibility to draw firm conclusions about the effects of custodial vs. non-custodial sanctions, despite hundreds of evaluations conducted world-wide over this question (Smith, Goggin, & Gendreau, 2002; Villettaz, Killias, & Zoder, 2006). Therefore, randomisation should become a far more acceptable, if not the standard option for policy makers who mandate evaluations of any new forms of treatment or sanction. The obstacles that are routinely invoked are far less absolute than often claimed (see a similar problem about present counter-terrorism intervention in Lum, Kennedy & Sherley, 2008). Once the number of randomised experiments will increase, researchers and policy makers will probably learn how to overcome legal and ethical obstacles in acceptable ways —everything is, in the end, a question of *how* rather than of *whether* controlled trials can be conducted.

It is common to invoke ethical reservations whenever a randomised controlled experiment is being envisaged. Such arguments seem to be quite odd as long as no evidence has shown that «new» sanctions or programs produce better results than traditional ones, or that they are at least not damaging. No one encourages pharmaceutical firms to sell new «promising» products before adequate testing through randomised controlled trials. Why should a new correctional program be «sold» to participants as long as its effects have not been adequately tested, simply because a few correctional specialists argue more or less convincingly that it may most likely be beneficial? If, as usual,

new programs are first implemented on an «experimental» basis rather than large-scale, the argument that refusing the new alternative to certain subjects (those in the control group) is unethical, sounds strange as long as their fate is to serve the standard program: implicitly, such an argument implies that most convicts are subject to «damaging» treatment —an obviously unethical option if we knew that this is the case.

Our experience with correctional services, convicted offenders participating in new programs and policy-makers has been that random assignment has many advantages also for staff and decision-makers operating in the field, despite some evidence in the literature to the contrary (Erez, 1985). Therefore, random assignment may often be easier to justify than any kind of choice on the grounds of personal characteristics, merits or institutional constraints, particularly if the number of candidates exceeds the planned capacity of the experimental group (Weisburd, 2000).

Serious practical problems are likely to arise whenever practitioners are highly committed to a program, as they should be (Petrosino & Soydan 2005), and if random assignment is rigidly applied without due consideration of practical concerns. There are often programs that are designed only for the treatment of individuals having certain characteristics. In such cases, it is good practice to screen subjects to assess their eligibility, as for example in the case of the *Sozialtherapeutische Anstalten* in Germany (Ortmann 1994, 2000). Even if this offers practitioners the possibility to eliminate subjects not suitable from the treatment group, they may have strong reservations whenever a «particularly needy» subject is being assigned to the control group (Little, Kogan, Bullock, & van der Laan, 2004). In order to defend the trial’s integrity, an excellent method, already suggested by Wilkins (1969), is to admit a certain (pre-fixed) number of subjects before any randomisation takes place. In the case of the Swiss community service vs. prison experiment, social workers were allowed to admit up to 25 percent of subjects before (i.e. without) randomisation. These subjects were, of course, not comparable to the two randomised groups. For this reason, they were kept identifiable and analysed separately. In the end, this procedure reduced temptations to «cheat» during randomisation (and, thus, increased the experiment’s integrity), increased acceptability of randomisation among practitioners, and prevented covert opposition to the trial, as in the case of the Kingswood Training School described by Cornish (1987).

In some countries, it may be wise to remove legal obstacles by appropriate legislative actions. For example, the Swiss parliament adopted, in 1971, an amendment to the penal code allowing the Government to introduce, on an experimental basis, i.e. for a limited number of offenders and for a limited period of time, innovative sanctions and correctional arrangements beyond those provided for by the penal code. Under this law, offenders who are offered the chance of serving their term in an «innovative» program may, at any time, refuse and claim to be treated «according to the law» (i.e. serving their term in prison, as a rule); however, no one is entitled to claim to become part of an experiment that is, by essence, limited in scope. Although this law removes legal obstacles precluding randomisation among those who volunteer and are eligible for an «experimental» sanction or program, it does not impose randomised controlled trials either, and it lasted more than 20 years before such a design has been first adopted. In other countries, too, new sanctions are, as a rule, introduced as a temporary and more or less «experimental»

arrangement. Therefore, controlled experiments should not be less feasible under such conditions.

#### *Looking at long-term effects*

No less deplorable than the lack of randomised trials is the fact that, whatever the method adopted, follow-up periods in evaluations rarely extend beyond two years, presumably because policy makers need rapid «feed-backs». However, randomised experiments make longer observation periods far more feasible. If subjects were, originally, randomly assigned to different conditions, their development over their entire life-span can be studied without undue investment in time and resources. Quasi-experiments, however, will never allow assessing long-term outcomes that were not anticipated, such as unforeseen health problems. If, for example, subjects in the treatment group suffer later in their lives more often than those in the control group of cardiovascular problems, as observed in the case of the Cambridge Somerville experiment (McCord 1990), it would, without randomisation, not be possible to rule out that, from the onset, candidates with more vulnerable health had been assigned disproportionately to the treatment group. Probably due to the unpopularity of randomisation, studies conducted so far in Europe never have extended to significant parts of subjects' later biographies. In European countries where population mobility (particularly across national and language barriers) remains relatively modest, long-term studies could be particularly fascinating given the availability of many data in official records over extended periods of time. This is the case in many European countries, although the best known example certainly is Denmark where huge databases covering entire biographies can be matched. Brennan & Mednick (1994) have used such records to study re-convictions in an entire birth cohort (born at Copenhagen between 1944 and 1947). Currently, the data of the Swiss community service vs. prison experiment are being re-analysed at our institute in the light of subjects biographies ten years later.

#### *Beyond official records as measures of «success» or «failure»*

Despite alternative (and presumably more valid) measures of re-offending (such as self-reports), most studies do not include measures of re-offending beyond re-arrest or re-conviction. Given the strong correlation between offending and victimisation, one might also validly consider, in evaluation research, a combination of self-report and victimisation questionnaires in order to assess effectiveness of programs. If a program is indeed successful at reducing offending rates, one should also be able to identify such an effect through reduced rates of victimisation. This is not trivial, since questions on victimisation often provoke less resistance than self-report instruments. A methodological study on some 500 subjects (Aebi, 2006) has documented a reasonably high validity of all three methods to identify program effects on the prevalence of offending.

#### *Looking at relative improvement*

In most studies, re-offending has been measured through the prevalence of post-intervention re-convictions or re-arrests. Left alone that questionnaires of self-reported delinquency and/or victimisation were rarely used, the simple prevalence (“yes/no”)

of arrests or convictions after an intervention may mask important variations in the frequency of offending (“incidence rates”) and relative improvement following different sanctions (Little, Kogan, Bullock, & van der Laan, 2004). Depending on the population studied, convictions are not necessarily frequent and may, eventually, not allow observing sufficient variance in order to discover any sanction (or intervention) effect, especially if the sample is not large. This is particularly true under the continental sentencing system where one global sentence is imposed for all (new) offences the defendant has been found guilty of, rather than one sentence for each verdict, as under the Anglo-American system. If re-arrest data are used, this problem is less serious because police contacts are more frequent than re-convictions, one court appearance being eventually related to several new offences known by the police. However, survey data systematically allow to observe far higher rates of re-offending than any official measures (Aebi, 2006).

Some studies have shown that most offenders reduce offending rates after whatever type of intervention (Empey & Lubeck 1971; Killias, Aebi, & Ribeaud, 2000). Thus, the relevant question may be to what extent they improve differently by type of sanction. Therefore, it would be urgent to look in future studies at rates of improvement (or reductions in offending) rather than merely at «recidivism» as such. This is particularly true if samples are not very large, if «failure» rates are not very high (or not very different across groups) and if, as not unusual in such situations, subjects' pre-intervention offending rates were, despite randomisation, higher in one group compared to the other. By comparing pre- with post-intervention incidence rates (e.g., number of offences known to the police during two years before and after the intervention), statistical power can be increased under such circumstances. Increasing the sample size is not always possible for practical reasons, or produces adverse side effects, such as reducing the «dosage» of treatment (Weisburd, Petrosino, & Mason 1993).

#### *Looking at rehabilitation beyond re-offending*

In studies comparing custodial and non-custodial sanctions, lower re-offending rates among those sentenced to an «alternative» sanction were, whenever observed, usually attributed to the fact that these offenders were not separated from their work and family life and may have had, therefore, better opportunities to integrate after having served their sentence. However, the evidence is extremely limited in this respect (Lamb & Goertzel 1974, Killias, Aebi, & Ribeaud 2000) and does not necessarily confirm this assumption, since almost all studies focus on re-offending (Israel & Chui, 2006). Given the often extremely short duration of custodial sentences compared to «alternative» sanctions under European law, it seems unlikely that any lasting «prisonisation» effect may have been produced. In the case of randomised controlled trials, it would be easy to conduct later follow-up studies including, beyond measures of re-offending, any kind of indicators of social integration, as they can routinely be found, for example, in the files of income revenue services. The files of such services routinely collect data on family disruption, unemployment, welfare payments, debts, revenues and resources. Such data would be highly relevant in assessing any negative long-term effects on integration of custodial compared to «alternative» sanctions, or of any other types of programs. Such data are

currently being used in two randomised experiments in Switzerland. They are also an attractive alternative to data collected through interviews, given the usual difficulties in locating and motivating subjects many years after their correctional experience. As the few available examples of long-term evaluations show, one important advantage of controlled trials is precisely the possibility to consider later outcomes in areas that no one had anticipated to be causally related to the intervention at stake.

#### *Hawthorn or placebo effects and double blind trials*

To the extent that, in randomised controlled trials, lower re-offending rates have been observed after «alternative» compared to custodial sanctions, it should not be ruled out that this outcome could be the result of a Hawthorn or placebo effect. Indeed, persons convicted to a custodial sanction who get the «chance» to serve it under the form of an «alternative», i.e. usually without having to go to prison, or who are placed in a «special» treatment unit within confinement, are offered, in some way, a second (often unexpected) chance which, in turn, may favourably affect their attitudes (as observed by Killias, Aebi, & Ribeaud 2000). A placebo effect has also been envisaged by the authors of the ISSP trial in England, where the experimental group showed moderately lower overall re-offending rates (Little, Kogan, Bullock, & van der Laan, 2004). Subjects in an experimental group may, whenever treatment means avoiding prison or any other unwelcome experience, typically develop the feeling of having been treated «better than expected» —or with more «fairness». As experiments on cooperation between unrelated individuals (Fehr & Rockenbach, 2003) have shown, sanctions perceived as fair do not affect subjects' willingness to cooperate, whereas sanctions resented as unjust or unfair destroy altruistic cooperation almost completely. That sanctions perceived as «fair» (in practice, this probably equals «better than expected») increase willingness to cooperate, matches similar results on reduced rates of re-offending as a result of attitude change. Such outcomes have been observed in studies on attitudes influenced by cognitive-behavioural treatment (Henning & Frueh, 1996; Vennard, Hedderman, & Sugg, 1997) or by «fair» procedures (Paternoster, Bachman, Brame, & Sherman, 1997).

In the medical field, the obvious answer would be to organize double-blind trials. For obvious reasons, double-blind experiments

are not feasible in the field of criminal justice where the nature of «treatment» cannot be concealed to subjects. It is surprising, however, that the possibility of such effects has, so far, found very little attention in the criminal justice literature. One way of assessing possible Hawthorn (or placebo) effects might be to look at subjects' development over longer periods of time, although this way may not necessarily allow distinguishing such effects from vanishing treatment outcomes. If, however, innovative «alternatives» —irrespective of what they are or what they imply— consistently produce better outcomes than more traditional sanctions, it may be fair interpreting such outcomes as a Hawthorn (or placebo) effect.

#### *Concluding remark: the role of evaluations in policy making*

If —as often observed in randomised controlled trials and again in case of the present meta-analysis— rates of re-offending are similar no matter what intervention subjects were assigned to, many think the money invested in the experiment had been wasted. Such a view is inappropriate, since an inconclusive outcome does not imply that «nothing works» or «nothing matters». Rather, similar outcomes after carefully evaluated interventions allow researchers and policy makers to validly conclude that effects of all options compared are similar. For policy makers, such an outcome means that the choice between programs can be based on considerations beyond their effectiveness, such as relative costs, availability of resources, fairness and equity to offenders and victims, consistency in sentencing, and popularity among defendants and the public. In this sense, striving for evidence-based crime prevention policies, as advocated by the Campbell Collaboration Crime and Justice Group (Welsh & Farrington, 2005), means that policy-makers will know better what they can expect in adopting solutions or policies whose effects have been documented through meta-analyses based on carefully designed experiments. One immediate policy implication may be that short-term prison sentences should no longer be replaced by whatever «alternative» in the name of their «damaging» effects on re-offending. Such policies have been conducted throughout Europe over several decades, with the unanticipated consequence that short prison terms have often been replaced by longer sentences. The result often have been dramatically increasing incarceration rates. Portugal and Spain are particularly worrisome examples of this trend (Kuhn, 2000).

#### References

- Aebi, M.F. (2006). *Comment mesurer la délinquance?* Paris: Armand Colin.
- Brennan, P.A., & Mednick S.A. (1994). Learning theory approach to the deterrence of criminal recidivism. *Journal of Abnormal Psychology* 103(3), 430-440.
- Cornish, D. (1987). Evaluating residential treatments for delinquents: A cautionary tale. In K. Hurrelmann, F.-X. Kaufmann & F. Lösel (Eds.): *Social intervention: Potential and constraints*, New York/Berlin: de Gruyter, 333-345.
- Empey, L.T., & Lubeck St.G. (1971). *The Silverlake Experiment*, Aldine, Chicago.
- Erez, E. (1985). Random assignment. The least fair of them all: Prisoners' attitudes towards various criteria of selection. *Criminology*, 23(2), 365-379.
- Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422, 137-140.
- Henning, K.R., & Frueh, B.C. (1996). Cognitive-behavioural treatment of incarcerated offenders. *Criminal Justice and Behavior*, 23, 523-541.
- Israel, M., & Chui, W.H. (2006). If 'something works' is the answer, what is the question? Supporting pluralist evaluation in community corrections in the United Kingdom. *European Journal of Criminology*, 3, 181-200.
- Killias, M., Aebi, M.F., & Ribeaud, D. (2000). Does community service rehabilitate better than short-term imprisonment? Results of a controlled experiment. *The Howard J. of Criminal Justice*, 39, 40-57.
- Kuhn, A. (2000). *Détenus: Combine? Pourquoi? Que faire?* Bern (Switzerland): Haupt.

- Lamb, R.R., & Goertzel, V. (1974). Ellsworth House: A Community Alternative to Jail. *American Journal of Psychiatry*, 131(1), 64-68.
- Little, M., Kogan, J., Bullock, R., & Van der Laan, P. (2004). ISSP: An experiment in multi-systemic responses in persistent young offenders known to children's services. *British J. of Criminology*, 44(2), 225-240.
- Lum, C., Kennedy, L.W., & Sherley, A. (2008). Is counter-terrorism policy evidence-based? What works, what harms, and what is unknown. *Psicothema*, 20(1), 35-42.
- McCord, J. (1990). Crime in moral and social contexts. *Criminology*, 28(1), 1-26.
- Ortmann, R. (1994). Zur Evaluation der Sozialtherapie. Ergebnisse einer experimentellen Längsschnittstudie zu Justizvollzugsanstalten des Landes Nordrhein-Westfalen. *Zeitschrift für die gesamte Strafrechtswissenschaft*, 106, 782-821.
- Ortmann, R. (2000). The effectiveness of social therapy in prison. A randomised experiment. *Crime and Delinquency*, 46(2), 214-232.
- Paternoster, R., Bachman, R., Brame, R., & Sherman, L.W. (1997). Do fair procedures matter? The effect of procedural justice on spouse assault. *Law & Society Review*, 31, 163-204.
- Perneger, T.V., Giner, F., Del Río, M., & Mino, A. (1998). Randomised trial of heroin maintenance programme for addicts who fail in conventional drug treatment. *British Medical Journal*, 3(17), 13-18.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism. Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1(4), 435-450.
- Smith, P., Goggin, C., & Gendreau, P. (2002). *Effets de l'incarcération et des sanctions intermédiaires sur la récidive: Effets généraux et différences individuelles*, Ottawa: Solicitor General of Canada 2002.
- Van der Werff, N. (1979). *Speziale preventie*. The Hague: WODC.
- Van der Werff, N. (1981). Recidivism and special deterrence. *British Journal of Criminology* 21(2), 136-147.
- Vennard, D.J., Hedderman, C., & Sugg, D. (1997). *Changing offenders' attitudes and behaviour: What works?* London: Home Office (Research Findings n° 61).
- Villetzaz, P., & Killias, M. (2005). *Les arrêts domiciliaires sous surveillance électronique: Une sanction «expérimentale»*, Lausanne: Institute of Criminology and Penal Law of the University of Lausanne.
- Villetzaz, P., Killias, M., & Zoder, I. (2006). *The effects of custodial vs. non-custodial sanctions on re-offending. A systematic review of the state of knowledge*, Campbell Collaboration Crime and Justice Group (under review).
- Walker, N., Farrington, D.P., & Tucker, G. (1981). Reconviction rates of adult males after different sentences. *British Journal of Criminology*, 21(4), 357-360.
- Weisburd, D. (2000). Randomized experiments in criminal justice: Prospects and problems. *Crime and Delinquency*, 46(2), 181-193.
- Weisburd, D., Petrosino, A., & Mason, G. (1993). Design sensitivity in criminal justice experiments. *Crime and Justice*, 17, 337-379.
- Welsh, B.C., & Farrington, D.P. (2005). Evidence-based crime prevention: Conclusions and directions for a safer society, *Canadian Journal of Criminology*, 47(2), 337-354.
- Wilkins, L.T. (1969). *Evaluation of Penal Measures*. New York: Random House.