



Kent Academic Repository

Losel, F, Koehler, J, Hamilton, L, Humphreys, D and Akoensi, Thomas D (2012)
Strengthening Transnational Approaches to Reducing Reoffending Final Report.
University of Cambridge

Downloaded from

<https://kar.kent.ac.uk/58675/> The University of Kent's Academic Repository KAR

The version of record is available from

http://www.starr-probation.org/uploaded_files/Rep%20STARR%20ENG.pdf

This document version

Publisher pdf

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Strengthening Transnational Approaches to Reducing Reoffending

Final Report



**Institute of Criminology
University of Cambridge**

Research team: *Principal Investigator*
Friedrich A. Lösel

Research Assistants
Johann A. Koehler (06/10-08/11)
Leah Hamilton (02/11-08/11)
David K. Humphreys (10/09-01/11)
Thomas D. Akoensi (11/09-02/11)

Table of Contents

	Page
1 Introduction	3
2 Systematic Reviews	4
2.1 Juvenile and Young Offenders	4
2.2 Drug Abusing Offenders	6
2.3 Domestic Violence Perpetrators	8
3 Questionnaire Surveys	10
4 Policy Implications	16
5 Appendices	
Appendix A: A systematic review and meta-analysis on the effects of young offender treatment programmes in Europe	
Appendix B: A systematic review and meta-analysis of the effectiveness of drug treatment programmes to reduce reoffending in Europe	
Appendix C: A systematic review of the effectiveness of domestic violence perpetrator programmes to reduce repeat abuse in Europe	
Appendix D: Treatment programmes to reduce reoffending throughout Europe: Three surveys on current practice	
Appendix E: Bibliography	

1. Introduction

The STARR Cambridge Team has produced six research products, comprising three systematic reviews and three questionnaire surveys. These products have focused on interventions to reduce reoffending among three offence categories, respectively: young offenders, domestic violence perpetrators, and substance abusing offenders. The questionnaire survey also gathered additional information on interventions to reduce reoffending among alcohol abusing offenders.

The three systematic reviews investigated the current state of European evidence on programmes to reduce reoffending in each of the three primary offence categories under STARR's purview. The questionnaire surveys clarified our understanding of what is currently practiced throughout Europe in reoffending programmes. We summarise the key conclusions from each of our primary research outputs below. Readers seeking further elaboration of the methods, processes, and outcomes are referred to the corresponding Appendices.

In addition to conducting primary research, the Cambridge Team has made various other contributions to the STARR project: 1. support of the project partners in the design and execution of a series of focus groups on obstacles to programme evaluation; 2. provision of research guidance and expertise in the execution of the STARR project partners' pilot projects in Bulgaria, Hungary, and France; 3. assistance in the organisation of the STARR seminars and conferences; 4. lectures at all STARR conferences; and 5. facilitation of group workshops at three STARR conferences and seminars. Furthermore, we have bolstered STARR's international presence by disseminating selected findings in invited plenary lectures at a number of conferences; e.g., at the International Prisons and Corrections Conference at Ghent, Belgium (2010), International Conference of Psychology and Law at

Miami, USA (2011), and the Summit of the Chinese Criminological Society at Hangzhou, China (2011).

2. Systematic Reviews

2.1 Young Offenders

Although the knowledge of what works in young offender treatment has substantially increased, the vast majority of evaluations have been carried out in North America. In contrast there is a clear lack of systematic knowledge on what works with this offender group across Europe. Therefore, this systematic review specifically addresses European evidence in reducing reoffending with young offenders. A summary of key points from our systematic review can be found below. Readers seeking further elaboration of the review's methods, processes, and outcomes are referred to Appendix A: *A systematic review and meta-analysis on the effects of young offender treatment programmes in Europe*.

Method

In order to locate unpublished and published studies, we searched online computerised databases as well as meta-analytic and systematic review publications dealing with juvenile offending and reoffending treatment programmes. We also contacted academics and experts in an effort to locate studies that might not have been accessed by the more conventional strategies. We searched for well-designed control group studies that had been conducted within Europe, and had been written in any language in common use throughout the continent.

Our electronic database search yielded 26,989 discrete titles. We supplemented this with our hand-searches of relevant bibliographies and questionnaire data gathered from

European experts from various sources. After detailed scrutiny of the titles, we located 22 studies conducted in Europe between 1980 and 2009 that targeted treatment of young offenders between the ages of 16 and 25. This yielded 25 separate controlled comparisons that met our eligibility criteria and addressed a total of nearly 8,000 young offenders. Two thirds of these evaluations came from the United Kingdom; only four other European countries had carried out at least one controlled evaluation of correctional programmes within our target group.

We coded studies according to important treatment and methodological characteristics, such as the type, setting and intervention of the programme, and the key features of the evaluation research design.

Results

Although there was a large variation in the outcomes of different studies, the mean effect size across all studies showed a significantly positive effect (Odds Ratio = 1.343; $p < 0.05$). This value corresponds to a recidivism rate of approximately 46% for the treatment group and 54% for the control group, what indicates a reduction of approximately 15%.

Behavioural and cognitive-behavioural treatments were more effective than other types of treatments. The effectiveness of the various non-behavioural treatment types was lower and reasonably homogeneous. Purely punitive and deterrence-based measures showed no or even a slightly *criminogenic* effect. There was a positive relationship between effectiveness of treatment and adherence to the principles of addressing the offenders' risk level, criminogenic needs, and responsivity (RNR model). Programmes that adhered to the RNR approach showed an average of approximately 18% less reoffending than the untreated control group. There was a negative relationship between effectiveness of treatment and methodological quality of evaluation design. In addition, studies with large samples,

evaluations of routine practice, low implementation quality, and custodial programmes showed a tendency to smaller effects. The average effect in the UK studies was smaller than in studies on the continent; however, this difference disappeared after controlling for programme type. There was no difference between voluntary and mandatory treatment.

Conclusion

Our systematic review and meta-analysis provides the first comprehensive overview of European evidence of what works in young offender rehabilitation programmes. Overall the results are encouraging and corroborate much of the evidence that has been provided from North America, although our findings are somewhat more moderate. They support basic principles of appropriate offender treatment, the particular relevance of behavioural and cognitive-behavioural approaches and the importance of methodological study characteristics. The overall small number of sound evaluations and their concentration in the UK and a few continental countries suggests that we need much more implementation and evaluation of best practice programmes across the whole of Europe.

2.2 Drug Abusing Offenders

Various programmes to treat substance abuse have been shown to reduce reoffending, but again this knowledge is derived primarily from North America. We therefore conducted a systematic review to investigate the impact of European interventions for drug abusing offenders on reoffending. Although a summary of key points from our systematic review can be found below, we refer readers seeking further elaboration of the review's methods, processes, and outcomes to *Appendix B: A systematic review of the effectiveness of drug treatment programmes to reduce reoffending in Europe*.

Method

In order to locate unpublished and published studies, we searched online computerised databases as well as meta-analytic and systematic review publications dealing with substance abuse treatment programmes. We also contacted academics and experts in an effort to locate studies that might not have been accessed by the more conventional strategies. We searched for studies that had been conducted within Europe, and had been written in any language in common use throughout the continent.

Our electronic database search yielded 37,473 discrete titles. We supplemented this with our hand-searches of relevant bibliographies and questionnaire data gathered from European experts from various sources. After detailed scrutiny of the titles, we located only 13 studies conducted in Europe that targeted treatment of substance abusing offenders and had good methodological quality. This yielded 15 separate controlled comparisons meeting our eligibility criteria. Eight of these evaluations were carried out in the United Kingdom and five continental countries contributed the remaining studies.

Results

All the evaluations applied treatments to opiate-dependent offenders who had previously rejected a course of treatment of some kind. The majority of evaluations addressed pharmacological substitution programmes (often combined with psychosocial measures). Only three studies investigated primarily non-pharmacological programmes. The interventions were mostly applied in the community and to voluntary participants. Although many of these studies had sound evaluation designs they were subject to various threats to validity (e.g. high numbers of dropouts).

The mean effect size was $r = 0.19$, $p < 0.001$. This value corresponds to a recidivism rate of roughly 40.5% in the experimental group and 58.5% in the control group, which

indicates a reduction of approximately 30%. Primarily pharmacological substitution-based treatments revealed a more positive effect than other types of treatment. Studies conducted within the United Kingdom reported more modest effects than those conducted elsewhere; however, this seemed to be due to different programme types.

Conclusion

The results of our systematic review suggest that drug treatment programmes in Europe have a substantial positive effect on reducing reoffending. The evidence included in this review applied a stringent level of methodological rigour to minimise the threats to internal validity that are endemic to previous reviews on the topic. However, the study sample consisted mainly of evaluations of pharmacological substitution programmes conducted in the United Kingdom, using samples of opiate-addicted offenders who had previously failed treatment. There is a clear need for a large-scale programme of implementation and systematic evaluation of a broader range of treatment programmes for more diverse populations of drug abusing offenders and across the whole of Europe.

2.3 Domestic Violence Perpetrators

Although some treatment programmes for domestic violent perpetrators show a significant reduction in reoffending, this research stems nearly exclusively from North America. Therefore, we carried out a systematic review of European evaluations of domestic violence perpetrators. The key points from our analysis are presented below. Readers seeking further elaboration of the review's methods, processes, and outcomes are referred to *Appendix C: A systematic review of the effectiveness of domestic violence perpetrator programmes to reduce repeat abuse in Europe*.

Method

In order to locate unpublished and published studies, we searched online computerised databases as well as meta-analytic and systematic review publications dealing with domestic violence perpetrator programmes. We also contacted academics and experts in an effort to locate studies that might not have been accessed by the more conventional strategies. We searched for studies that had been conducted within Europe, and had been written in any language in common use throughout the continent.

Our electronic database search yielded 10,446 discrete titles. We supplemented this with our hand-searches of relevant bibliographies and questionnaire data gathered from European experts from various sources. Our search for evaluations in Europe of domestic violence perpetrator programmes revealed that the general standard of methodological rigour in this field is considerably lower than what was found in our other systematic reviews. We therefore relaxed the constraints of our eligibility criteria for this review. After detailed scrutiny of the reports, we located 11 studies conducted in Europe that targeted treatment at domestic violence perpetrators and contained at least some kind of systematic evaluation.

Results

The 11 studies came from six European countries and the majority of these studies were carried out in the last decade. In total the evaluations contained 1,413 domestic violence perpetrators. Typically these studies used single group pre-post measurements of a range of outcomes. Only one study contained a comparison group, but this was not equivalent to the treatment group. The length of the follow-up period before outcome measurements were taken varied between 0 and 12 months after the programme concluded. All studies suffered from high dropout rates (between 8%-73%). Outcome measures were mainly collected by

means of self-report questionnaires administered to the perpetrator, and this data was rarely triangulated with other sources.

All studies evaluated treatments that incorporated a combination of psychodynamic/psycho-educational and cognitive-behavioural approaches. In all studies, programme completers reported reductions in domestic violence, psychological problems or other outcome measures from the beginning of the programme compared to the end. There were also a number of desirable effects at follow up, for instance 5 to 30 percentage point reductions in violence.

Conclusion

In comparison to the evaluations in our other systematic reviews, the studies on domestic violence perpetrators are of much lower methodological quality. Although *prima facie* improvements were observed in all of the evaluations, the threats to validity do not allow us to conclude that these studies have a positive causal effect on domestic violence reoffending. In particular, one cannot rule out that desirable outcomes are due to selection effects and natural desistance patterns. This underscores the necessity of comparison group research designs in this field. Furthermore, in the absence of long-follow-up periods and the triangulation of multiple outcome measurement sources, it is likely that the evaluations did not capture the full picture of post-intervention domestic violence. For these reasons we need more and better research on the effectiveness of domestic violence perpetrator programmes in Europe.

3. Questionnaire Surveys

The STARR Cambridge team has administered three surveys to investigate current practice in reducing reoffending throughout the European Union. We summarize the key points from our questionnaire survey below and refer readers seeking further elaboration of the method, process, and outcomes to Appendix D: *Programmes to reduce reoffending throughout Europe: Three surveys on current practice*.

Method

We developed three questionnaires that contained multiple choice and semi-structured questions, and gathered information on the following themes: the type and theory of correctional treatment programmes; participant and staff selection and characteristics; programme organisation; programme effectiveness and administration; and general information on the infrastructure of rehabilitation programming in the respective country. A pilot study was used to evaluate questionnaire design. We approached four main sources of respondents: ministries of justice, expert contacts, pan-European organisations, and community sector searches. We supplemented these contact networks with correspondences that emerged from the STARR conferences and seminars.

A separate questionnaire was designed and distributed for programmes for young offenders, drug abusing offenders, and domestic violence perpetrators. To avoid overload, resistance and confusion among our European partners the questionnaires were distributed sequentially: we distributed the young offender questionnaire in July 2010, the domestic violence questionnaire in October 2010, and the substance abuse questionnaire in February 2011. Many contact persons and institutions returned their questionnaires in one to two months. However, it was also often necessary to contact some programmes repeatedly by both phone and email over multiple months. We made provisions for respondents to fill out

the questionnaire in any language when it became clear that English would be an obstacle to participation. Due to these persistent efforts to have representative data, we received information on at least some programmes from all EU countries. All questionnaire responses were completed and returned by June 30th, 2011.

Results

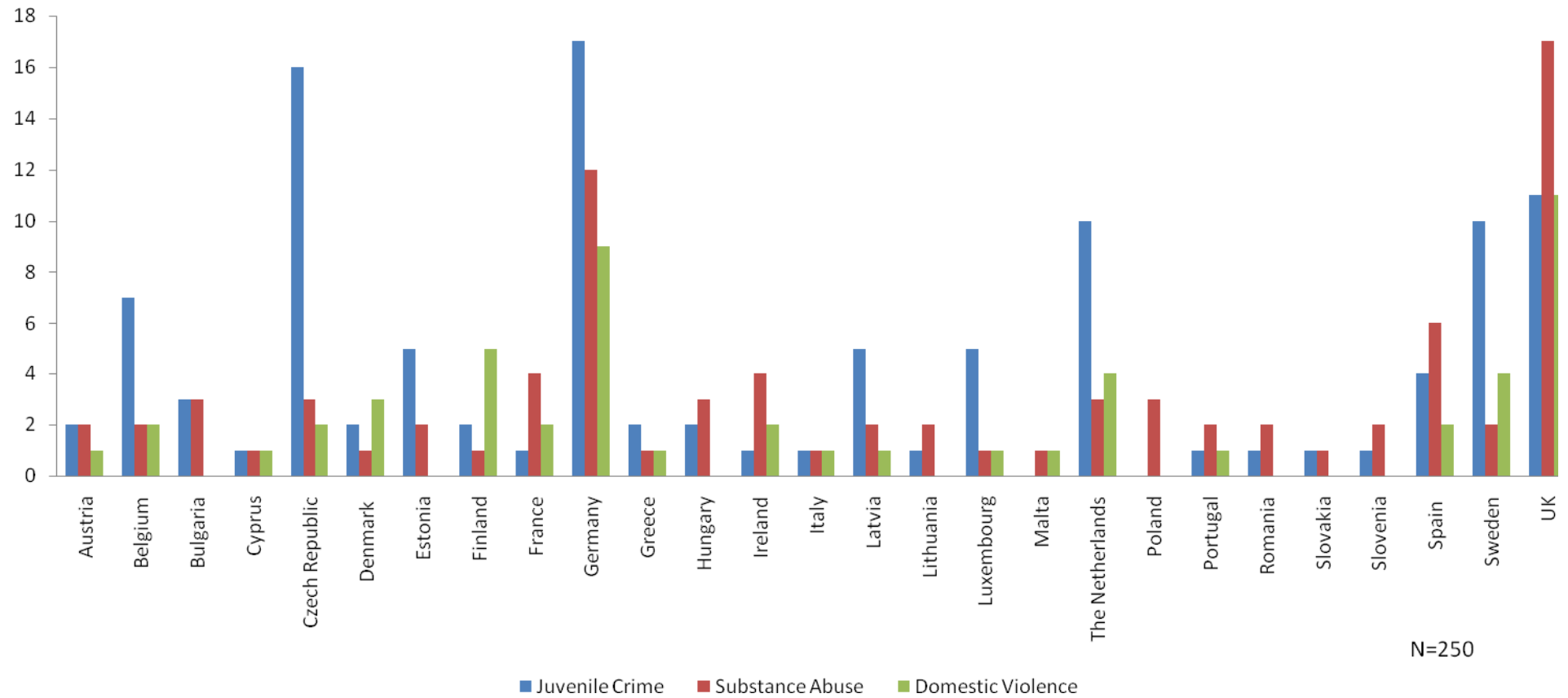
Country Representation

We collected 250 responses from programmes in all 27 EU countries: 112 were young offender programmes, 84 were substance abuse programmes and 54 were domestic violence programmes (see Figure 3.1). The most prolific responses in all three categories originated from Northwest Europe. In the young offender survey we located programmes from 25 of the 26 countries that had established reoffending interventions. In the substance abuse survey we collected questionnaires from all 27 countries. We located four countries that had not yet developed any domestic violence perpetrator interventions, and we collected questionnaires from 19 out of the 23 remaining countries. Domestic violence perpetrator programmes proved to be the most difficult to locate. This was due to their diverse and often NGO-based organisation, relatively low connections to government bodies and often limited contacts between programmes within a given country.

Treatment Modality

The prevailing treatment modality was cognitive/behavioural in all three surveys. However, many programmes also incorporated other theoretical perspectives in their treatment method such as counselling or psycho-dynamic therapy. In all three categories, roughly a third of programmes were transferred from another country.

Figure 3.1: *Country representation of questionnaire responses*



Young offender treatment programmes are delivered in community and custody settings to relatively equal degrees. These programmes tend to be highly structured, and tend to use risk assessment to focus their treatment delivery to the risk level of the offender. However, many risk assessments are not based on systematic tools and most countries lack differentiated programmes to accommodate variance in risk level.

Programmes for drug and alcohol abusing offenders tend to be highly structured, are just as likely to be administered in the community as in custody settings, and the majority are cognitive/behavioural in approach. There is a relative dearth of pharmacological maintenance programmes throughout the EU that are designated to deal, in part, with the reduction of reoffending. Programmes use risk assessment more frequently than young offender or domestic violence programmes; however, programme adaptation to risk level remains weak.

Domestic violence perpetrator programmes are generally structured at least to some extent, and the majority are administered in the community. All the programmes in our survey use a mixture of cognitive/behavioural and psycho-dynamic/psycho-educational approaches. Most programmes use risk assessment to focus their treatment delivery to the risk level of the offender. This risk assessment is frequently based on professional judgment rather than risk assessment tools; therefore, programmes delivery is rarely influenced by systematic assessment.

Evaluation

Although our survey revealed a broad range of programmes that are delivered to the three offender groups across Europe, they mostly lack local evaluation. Only 26 of the 250 programmes (10%) are based on an evaluation that is in accordance with standards of scientific rigour that would suffice to ascertain effectiveness. Based on respondents' answers,

the lack of incorporation of evidence-based methods into routine practice is due to four main reasons: a lack of affinity for systematic evaluation, a lack of knowledge about the methods of evaluation, a lack of institutional support, and a lack of resources.

Conclusion

It is in accordance with the mainly North-American ‘what works’ research that most ‘best practice’ treatment programmes for young offenders in Europe adopt a cognitive/behavioural approach. In contrast to their substantial effects in our systematic review, pharmacological substitution programmes for drug abusing offenders are the least prevalent treatment type in Europe. Although psychosocial programmes are more frequently used in this field, their effectiveness has not yet been established with sufficient validity. The field of domestic violence perpetrator programmes is least developed across Europe. Our survey suggests that this may be due to the more diverse and less government-driven organisation of such programmes.

In general, systematic evaluation is still under-appreciated and under-used among programmes to reduce reoffending throughout Europe. The fact that the majority of programmes tend to be structured and funded by government administrations, would make them suitable for much more controlled evaluation of their effectiveness. Therefore, systematic strategies to promote evidence-based practice across Europe should be implemented.

4. Policy Implications

- Although we found rather limited and not wholly consistent knowledge of what works in European reoffending programmes, there is substantial evidence for effective types of approaches in the fields of young offender and drug abusing offender treatment. Such programmes should be implemented with priority across Europe.
- Many programmes to reduce reoffending in Europe are not yet based on any systematic evidence. Therefore, such programmes should be evaluated in order to understand how well they are currently operating and how best to refine their treatment effectiveness.
- Often programmes are transferred from other countries without any local evaluation of their effectiveness. Although our findings suggest that basic characteristics of effective programmes may be transferrable across countries, this needs to be investigated in those large parts of Europe where we currently do not have any systematic strategies for programme evaluation.
- Our research revealed deficits in risk assessment, programme implementation and other process characteristics. Therefore, we need to improve routine practice in these fields that would increase treatment quality, rates of programme completion and other positive treatment characteristics.
- The lack of systematic evidence on the effects of correctional programmes in the majority of European countries requires clearer standards of good practice. As a consequence, standards that are suitable for a broad range of countries should be developed and transferred across Europe.

- The development of more evidence-based approaches to reducing reoffending in major parts of Europe is at least partially due to deficits in funding, ideological scepticism, insufficient human resources, and a lack of expertise in conducting sound evaluation. Therefore, we need increased efforts in resource allocation, training and dissemination to overcome such obstacles for rational and cost-effective crime policies.

A Systematic Review and Meta-Analysis on the Effects of Young Offender Treatment Programmes in Europe*

Johann A. Koehler¹ Friedrich Lösel^{1,2}

Thomas D. Akoensi¹, and David K. Humphreys¹

¹ Institute of Criminology

University of Cambridge, UK

² Institute of Psychology

University of Erlangen-Nuremberg, Germany

* This study was carried out within the European cooperation project “Strengthening Transnational Approaches to Reducing Reoffending”. The project was funded by the European Union. We thank our partners from the UK Ministry of Justice, London Probation Trust, European Organization for Probation, and the Ministries of Justice of Bulgaria, France and Hungary for their cooperation. We also thank primary study authors for providing us with data to assist in the meta-analysis.

Abstract

Most primary studies and systematic reviews on what works in offender rehabilitation concentrate on North America. To investigate how far these findings can be generalised across countries and cultures, this article presents a meta-analysis on the effectiveness of correctional programmes for young offenders in Europe. A literature search of approximately 27,000 titles revealed 25 controlled evaluations that fulfilled the eligibility criteria (e.g. equivalent treatment and control groups). In total, the studies contained 7,940 offenders with a mean age of 17.9 years. The outcomes ranged widely from an $OR = 0.58$ to 6.99 , the mean effect was significant and in favour of treatment ($OR = 1.34$). Behavioural and cognitive-behavioural treatment ranked above average ($OR = 1.73$), whereas purely deterrent and supervisory interventions revealed a slightly negative outcome ($OR = 0.85$). Programmes that were conducted in accordance with the Risk-Need-Responsivity principles revealed the strongest mean effect ($OR = 1.90$), which indicates a reduction of 30% in reoffending against a baseline of 50%. Studies of community treatment, with small samples, high programme fidelity, and conducted as part of a demonstration project had larger effects; high methodological rigour was related to slightly smaller outcomes. Overall, most findings agreed with North-American meta-analyses. However, two thirds of the studies were British, and in most European countries there was no sound evaluation of young offender treatment at all. This limits the generalization of results and underlines the policy need of systematic programme evaluation across the continent.

Introduction

Offender rehabilitation research and practice have made considerable progress since the claims of the 1970s that ‘nothing works’. There is now a large body of meta-analyses on ‘what works’ for general, violent, sexual, and other offender groups (Andrews et al., 1990; Andrews and Bonta, 2010; Aos et al., 2006; Hanson et al., 2002; Hollin and Palmer, 2008; Lipsey, 1992; Lipsey and Cullen, 2007; Lipsey and Wilson, 1998; Lösel, 1995, 2011; Lösel and Schmucker, 2005; McGuire, 2002; MacKenzie, 2006; Pearson et al., 2002; Tong and Farrington, 2006). They show relatively sound effects for cognitive behavioural treatment, structured therapeutic communities and multimodal systems-oriented programmes, whereas pure punishment, deterrence and supervision interventions revealed either negligible or even moderately negative outcomes.

In addition to specific programmes, more comprehensive characteristics of appropriate treatment have been recommended. Most popular became the model of Risk-Need-Responsivity (RNR; Andrews et al., 1990; Andrews and Bonta, 2010) that is supported by empirical research (e.g. Andrews and Bonta, 2010; Andrews et al., 2006; Hanson et al., 2009). According to RNR, treatment should correspond to the offenders’ risk level of reoffending, address their dynamic risk factors, and match their learning styles and capabilities (sometimes divided into ‘general’ and more individualised ‘specific responsivity’). Over time, the RNR concept has been expanded to include organisational, staff and other characteristics of evidence-supported offender treatment (e.g., Andrews, 1985; Andrews et al., 2011; Andrews and Bonta, 2010; Lösel, 1995).

These criteria are used for programme accreditation in European countries such as England and Wales, Netherlands, Norway, Scotland and Sweden (e.g., Maguire et al., 2010). Over the last decade the rollout of offender treatment programmes in Europe strongly

increased, but this was not based on local evidence. The vast majority of the above-mentioned ‘what works’ findings stem from North America and concentrate on English language material. Therefore, it is unclear how much these findings can be generalised to other cultures and countries. Very few systematic reviews addressed offender treatment in Europe. They focused on specific treatments in the German-speaking world (Lösel et al., 1987; Lösel and Köferl, 1989; Lösel, 2000) or rather heterogeneous samples of studies that aggregated analyses of the effects of diverse programmes for both juvenile and adult offenders in Europe (Redondo et al., 1999, 2001). To reach substantial study samples, these reviews had to include evaluations of moderate design quality. Because of their publication date they also could not cover more recent programmes. When the first findings of accredited programmes in England and Wales showed less positive results than North-American reviews, controversies arose surrounding the optimism of ‘what works’ and design of local evaluations (Harper and Chitty, 2005; Hollin, 2008; Maguire et al., 2010). These discussions highlight the issue that offender treatment is not like a natural science technology that may be easily transferred across cultural contexts, national legal regulations, criminal justice organizations, staff resources, offender populations, and so forth. This issue is already suggested by the meta-analytic findings that the programme contents only explain a limited amount of outcome variance and many other factors play a role (Lipsey and Wilson 1998; Lösel, 2011).

Against this background the present article reports a current systematic review and meta-analysis of correctional programmes for young offenders in Europe. We concentrate on this target group for four reasons: first, youth crime and violence has elicited heightened and immediate concern in many countries. Second, a potential positive intervention effect in young offenders is of particular value because it may cut off long-term criminal careers. Third, the focus on a specific offender population leads to a more homogeneous sample of

studies and programmes that would increase the potential for generalisation. And fourth, the field of young offender treatment is most promising with regard to the number of primary studies that are suitable for a meta-analysis.

Method

Study Inclusion

Eligibility Criteria

1. *Region.* Our literature search targeted evaluations of correctional interventions conducted in Europe.
2. *Target population.* The samples had to comprise young offenders, defined as previously adjudicated youths up to the age of 25. In instances where the evaluation was composed of subjects falling outside of these criteria (for example, where the sample contained at-risk youth with conduct disorders who had not offended), the study was included only if at least two-thirds of the participants met the eligibility criteria. We excluded studies if the sample of participants was deemed too idiosyncratic to allow the findings to be generalised. For example, we did not include treatments that were applied exclusively to sex offenders, psychopaths, or unique treatments that did not apply to a general offender population.
3. *Intervention.* The intervention had to contain a circumscribed correctional programme that aimed to reduce reoffending. Broader analyses of national-level policy changes were not included.
4. *Outcomes.* Data had to concern reoffending as an outcome, which was understood either as a formal institutional measure (e.g., re-arrest, re-conviction, re-incarceration, revocation of probation or parole), or as self-reported data pertaining to crime. Lower-level offences that did not constitute crime, such as anti-social behaviour, were not included.

5. *Evaluation design.* The evaluation had to compare the effect of the intervention in a treatment group with the level of reoffending in a control group. The latter could be defined as the application of no treatment, treatment as usual, or an alternative treatment. We excluded studies if the effectiveness of treatment was compared to national statistics of the general youth offender population. The control group had to show a clear indication of equivalence to the treatment group. This could have been achieved by randomisation, matching procedures or statistical comparisons of equivalence (Levels 3-5 on the Maryland Scale of Methodological Rigor; Sherman et al., 2002).
6. *Report of effects.* Any common statistics or raw data that allowed calculation of effect sizes were eligible. When primary studies reported data in a manner that was insufficient for the purposes of calculating an effect size, we made every effort to contact the author to acquire the necessary information.
7. *Publication.* We included published and unpublished evaluations appearing between 1980 and 2009.
8. *Language.* Reports in any commonly used European language were eligible. If no member of our research team was able to read a specific language, criminology students from the respective countries translated these documents.

Literature Search

In order to locate unpublished and published studies, we searched online computerised databases. The search terms and data bases are shown in Appendices Ai and Aii. In addition we searched meta-analytic and systematic review publications dealing with juvenile offending and reoffending treatment programmes as well as the references of primary studies. In an effort to acquire unpublished evaluations, we conducted a survey of European treatment programmes that aim to reduce juvenile reoffending, and requested the data when those

programmes had been evaluated. We distributed questionnaires to experts in all 27 EU countries, as well as to academics in European countries who were not EU members. We used snowball sampling methods to locate more inaccessible respondents.

Our bibliographic database search yielded a total of 26,989 titles¹, which, upon deletion of duplicates, yielded 21,223 discrete studies. Preliminary screening of titles and abstracts for *prima facie* relevance to our eligibility criteria yielded 14,001 studies. The abstracts were then screened in more detail according to method, location of study, sample population, and outcome of interest, in order to arrive at 88 studies (for details regarding the eligibility criteria, see above). We retrieved all of these documents in full and 22 studies met our inclusion criteria. A further ten studies had to be excluded because they had no control group or insufficient methodological rigour ($k = 5$), did not provide adequate data to compute an effect size ($k = 3$), or addressed exceptional offender populations ($k = 2$), thus resulting in 12 studies. A further nine studies that had not appeared in our electronic database search were added as a result of contact with other academics and experts and hand searching through relevant journals and bibliographies.

Our search uncovered no unpublished data. Although our European survey aimed to reduce the potential bias of simple document search (Wilson 2009), this strategy revealed clear limits. Of $n = 112$ received questionnaires 44% reported some kind of outcome evaluation (Hamilton et al., 2011; Appendix D). However, of those that had been conducted, none met our criteria of substantive or methodological rigour. As a result, this strategy yielded no further contributions to our study sample. We therefore proceeded with statistical analysis to examine a potential publication bias in our data set (see results section).

¹ The full number of studies yielded within each database is catalogued in Appendix Aii.

Coding of Variables

Outcome Measure

According to our eligibility criteria we used official indicators as well as self-reported reoffending. Although these measures suffer from their own unique threats to construct validity (Lloyd et al., 1994; Lösel, 1995), nonetheless, they are more relevant for policy than psychometric and other criteria that have also been used in a number of studies. We collected all the available outcome data in a given study, but used only the most relevant, general recidivism measure for computing an effect size (as opposed to multiple effects). Where various indicators of reoffending were available, we combined the data in a composite score. This also applied to different levels of seriousness of recidivism ($k = 1$; Lösel and Pomplun, 1998).

Effect Size

Where studies reported the prevalence of reoffending, we calculated an odds ratio (OR), whereas mean scores were used to calculate Cohen's d , which were then converted into OR. When necessary values were omitted in study reports, we calculated the effect size based on available p , t , or F values. Otherwise, every effort was taken to contact the author. In a few cases, we imputed effect sizes on the basis of other available data (see Ttofi et al., 2008).

Where possible, we sought to avoid using effect size data that were adjusted for pseudo-reconvictions (that is, when court data that refers to offences committed before the first reference offence indicate a 'reoffence'). Pseudo-reconvictions seem to decrease the observed reoffending rate by up to 7% (Lloyd et al., 1994). Therefore, this issue would introduce a conservative bias favouring the control condition.

Intent-to-treat data was reported in only five studies (20% of the sample). Therefore, we had to use treatment-as-treated outcome data in our computation of effect sizes. This could lead to biased results in favour of treatment (Lösel, 1995), although there was no

substantial difference in the few studies where intent-to-treat data have been available (see results section).

Where multiple comparison groups were presented ($k = 1$; Raynor and Vanstone, 1997), we used the data from the group that most closely approximated the normal criminal justice course without the respective intervention.

Follow-Up

We included data on all the follow-up periods provided within each study. Where data were incomplete, we contacted authors to acquire the necessary information. Only two studies (Curran et al., 1995; Lösel and Pomplun, 1998) reported follow-up data for more than two years after the end of the intervention. The mostly relatively short follow-up periods may lead to a positive bias; however, a recent study in England and Wales revealed that approximately two thirds of offenders discharged from custody reoffended within 24 months (Ministry of Justice 2010; see also Tournier and Barre (1990) for other European countries).

Intervention Type

Due to the considerable diversity of treatments, we had to compromise between creating treatment subcategories that were perhaps too broad on one hand, against the competing concern of low statistical power because of too few studies in each category on the other. We ultimately settled on three categories of treatment type: a) ‘Cognitive-Behavioural and Behavioural’ treatment (e.g., thinking skills programmes, social skills and problem solving approaches, reinforcement of behavioural change); b) ‘Intensive Supervision and Deterrence-Based’ interventions (e.g. amplified sanctions, boot camps without educational/therapeutic elements, and purely control-based supervision), and c) ‘Non-Behavioural’ treatment. This last category included a mix of programme types, ranging from educational and vocational skills training to programmes such as mentoring, restorative

justice, and intensive probation support. We coded these interventions separately to explore potential heterogeneity within that treatment category.

Studies were also coded based on whether the intervention was conducted in a community or a secure setting, whether the evaluation was conducted in the United Kingdom or elsewhere, and whether participation was voluntary or obligatory.

Risk, Needs, and Specific Responsivity

Because the studies did not contain detailed within-sample data on risk, we applied the aggregate sample approach to coding risk (e.g., Dowden and Andrews, 2003). The coding was based on data concerning prior contact with the criminal justice system and the primary study authors' own descriptions of sample risk. As no 'Low risk' study was observed, risk was coded on a three-level item, indicating 'Medium', 'Medium-High', and 'High' risk.

The studies were also coded according to two three-level items indicating: firstly, the degree to which treatments were designed to address specific criminogenic needs, as established in the rehabilitation programme literature (Hollin 2002); and secondly, to specific responsivity, or the degree to which programme implementers adapted the treatment delivery to the unique learning styles and capabilities of the offender. Both items were coded using the following categories: 'Low', 'Moderate', and 'High'.

Methodological Rigour

Common methodological features such as the type of research design, the total sample size, the length of follow-up measurement, and the level of attrition were coded for each study. To avoid too many subcategories with only few studies, we created dichotomous variables. For sample size, a total *N* of 100 provided the cut-off between 'Large' and 'Small'.

Follow-up measurement was dichotomised into 12 months and less as ‘Short’ and anything above as ‘Long’.

Research design was coded as a three-level variable, in which ‘Moderate Control’ denoted a quasi-experimental post-hoc formation of the comparison group that contained data suggesting overall similarity between treated and untreated groups. ‘Strong Statistical Control’ denoted that the two groups were matched according to key variables, and ‘Randomised Experiment’ denoted the randomisation of participant allocation to experimental groups (without obvious selective dropouts or other threats to validity). These three design categories correspond to levels 3, 4 and 5 of the Maryland Scientific Rigor Scale (Sherman et al. 2002), respectively.

Attrition was coded according to whether fewer than 15% of the total sample had dropped out, or more than 15% had dropped out. Where insufficient data was reported, we coded the study as ‘Not Reported’. We also collected data regarding whether the evaluation had been conducted as part of a demonstration or a routine practice project. Finally, we coded the level of fidelity to the programme objective on a two-level scale, using the categories ‘Low’ and ‘High’. This data was gleaned from primary study author’s reports of how closely the treatment delivery approximated the proposed treatment design or whether problems of integrity became apparent. Coding of all qualitative variables was handled through intensive discussion among all the authors.

Results

Description of the Studies

Our final sample comprised 21 evaluations with 25 discrete comparisons between treatment and control group (in the following named “studies”). Table 1 outlines key descriptive sample features. 87% of the studies have been published since 1990 and thus indicate relatively timely interventions. The majority of evaluations (67%) were conducted in the United Kingdom and the rest came from only four other European countries.

Table 1: *Study features*

Study Characteristic		<i>k</i>	Percentage
General			
Publication Year	1980-1989	4	16
	1990-1999	8	32
	2000-2009	13	52
Country	Germany	2	8
	Netherlands	4	16
	Norway	2	8
	Sweden	1	4
	United Kingdom	16	64
Sample			
Mean Age (years)	Жмр	8	32
	15 - 20	14	56
	21 - 25	3	12
Male (proportion)	Жфд	8	32
	Бфд	11	44
	Not Reported	6	24
Treatment			
Treatment Type	Behavioural/Cognitive-Behavioural	11	44
	Non-Behavioural	10	40
	Intensive Supervision/ Deterrence	4	16
Setting	Community	17	68
	Custody	8	32
Participation Type	Voluntary	14	56
	Obligatory	11	44
Risk	Medium	12	48
	Medium-High	3	12
	High	10	40
Needs	Low	9	36
	Moderate	6	24
	High	10	40
Specific Responsivity	Low	7	28
	Moderate	8	32
	High	10	40
Methodology			
Total Sample Size	Жмл	10	40
	>100	15	60
Group Allocation	Moderate Control	13	52
	Strong Statistical Control	8	32
	Randomised Experiment	4	16
Follow-Up²	Жмн а зүйКă	17	68
	> 12 Months	13	52
Dropout Rate	Жмр	8	32
	Бмр	7	28
	Not Reported	11	44
Research Type	Demonstration	7	28
	Practice	18	72
Programme Fidelity	Low	14	56
	High	11	44

² The percentages do not sum to 100, as studies may report multiple follow-up periods.

The sample consisted of a total N of 7,940 young offenders, with 3,883 in the TGs, and 4,057 subjects in the CGs. It was overwhelmingly male; in two thirds of the studies more than 90% of the participants were male. The average ages ranged from 14 to 23.7, with a mean of 17.9 years.

The intervention types were diverse: they included traditional 'boot camp'-style programmes (Farrington et al. 2002), self-help manuals (McMurran and Boyle, 1990), traditional cognitive-behavioural methods (e.g., Cann et al. 2005; Ogden and Hagen, 2006; Ogden 2007), mentoring programmes (Newburn and Shiner, 2005; St. James-Roberts et al. 2005), restorative justice (Shapland et al., 2008), and intensive supervision (Bottoms, 1995), among others.

Overall Effect Size

Figure 1 contains the effects and confidence intervals in the primary studies. The ORs ranged from 0.582 to 6.984, with fifteen studies showing an effect in favour of the treatment group, and nine pointed in the opposite direction. Six effects were statistically significant ($p < .05$) and all in favour of treatment.

The overall mean effect size for the sample was $OR = 1.34$ ($p < .05$; see Figure 1), which translates to a Pearson's r of .08 or Cohen's d of .16. Using the Binomial Effect Size Display (BESD; Rosenthal, 1991) and assuming a base rate of 50%, this effect value corresponds to 46% recidivism for the treatment group and 54% for the control group. The average effect displayed considerable heterogeneity ($Q_{\text{Total}}(24) = 64.48$; $p < .001$). Therefore, the meta-analytic results that follow are all derived from random-effects models. Although Borenstein et al. (2009:84-5) caution against choosing between fixed- and random-effects models purely on the basis of the Q statistic (largely due to the rarity with which it attains adequate power), the use of a random-effects model is further justified because our

distribution of effect sizes from a few European countries is not attributable to mere sampling error alone. Although small studies may be over-weighted in the random effects model, the mean effect was not much smaller and still highly significant when the fixed-effects model was applied for comparison ($OR = 1.19, p < .001$). The heterogeneity of outcomes justified further investigation in a moderator analysis.

Moderator Analysis

Treatment Characteristics

Table 2 contains the results on the differential effectiveness of the various treatment types and other intervention characteristics. Due to the small number of studies in the respective subcategories most of the differences between them were not statistically significant. However, some heterogeneity tests were only slightly above the significance threshold, in particular for the general treatment type ($Q_{\text{Between}}(2) = 5.56; p = .06$), adherence to the Responsivity principle ($Q_{\text{Between}}(2) = 4.63, p = .11$), and the overall RNR model ($Q_{\text{Between}}(2) = 4.20, p = .12$). As we used two-sided tests and most moderators contained theoretically meaningful and significant effects in specific subcategories, we presented the respective findings in Table 2.

Behavioural and cognitive-behavioural (CBT) treatments reported the largest OR of 1.73 ($p < .001$). This corresponded to a reduction in recidivism of 26% in the treatment group, compared to the control group. Non-behavioural treatments reported a smaller, statistically non-significant mean effect, although the direction was still in favour of treatment ($OR = 1.23, p = .21$). Intensive supervision and deterrence-based treatments showed a tendency towards *criminogenic* effects, favouring the control condition ($OR = 0.85$).

Figure 1: *Effectiveness of young offender treatment programmes*

With the exception of mentoring programmes, the effects of the various non-behavioural intervention types were roughly similar. Educational and vocational training programmes were most promising ($OR = 1.69$; $p = .16$); restorative justice programmes and guidance and counselling programmes also revealed positive, but nonsignificant effects. The two studies on mentoring programmes suggest nonsignificant undesirable effects.

Programmes that were conducted in community settings significantly reduced recidivism ($OR = 1.48$, $p < .05$), whereas the effect of custodial programmes was nonsignificant ($OR = 1.15$, $p = .46$). Treatments where participation was voluntary as well as mandatory programmes showed significant positive effects. The difference between both framing conditions was negligible.

Evaluations conducted within the United Kingdom reported smaller effects on average ($OR = 1.11$, $p > .05$) than evaluations conducted outside of the United Kingdom ($OR = 2.17$, $p < .001$). This difference was significant ($Q_{\text{Between}}(1) = 11.81$; $p < .001$). Seven of the nine evaluations that had been conducted outside of the United Kingdom were cognitive-behavioural programmes, and the majority of the eighteen practice programmes in our sample originated from the United Kingdom.

In comparison to medium risk, treatment of medium-high and high risk participants displayed more favourable effects and the OR of 1.63 for the latter group was significant. A similar systematic trend existed with regard to criminogenic needs. Programmes that were 'Moderate' in this category revealed a positive tendency ($OR = 1.41$, $p = .10$) and those that were 'High' reported a significantly positive effect ($OR = 1.59$, $p < .05$). The same pattern was observed for specific responsivity. Treatments displaying a 'Low' 'Moderate' level of responsivity were not effective, whereas programmes that were 'Moderate' or 'High' in this category showed nearly or in fact significantly positive outcomes (for the latter $OR = 1.64$; $p < .05$).

Table 2: Effectiveness of treatment according to treatment moderators

Moderator		<i>k</i>	<i>OR</i>	<i>CI</i> _{95%}		<i>Q</i> _{Between}	<i>BESD</i> (%)
				Lower	Upper		
Treatment							
Treatment Type						5.56	
	Behavioural/ Cognitive-Behavioural	11	1.73***	1.26	2.36		14.8
	Intensive Supervision/Deterrence	4	0.85	0.50	1.46		-3.7
	Non-Behavioural	10	1.23	0.89	1.69		5.7
						2.00 [†]	
	Vocational Training/Education	4	1.69	0.81	3.52		25
	Restorative Justice	1	1.56	0.54	4.50		22
	Guidance/Counseling	3	1.24	0.62	2.49		11
	Mentoring	2	0.79	0.38	1.65		-13
Setting						1.18	
	Community	17	1.48**	1.13	1.93		19
	Custody	8	1.15	0.79	1.66		8
Research Origin						11.81***	
	United Kingdom	16	1.11	0.91	1.36		6
	Non-United Kingdom	9	2.17***	1.57	3.00		35
Participation Type						0.00	
	Obligatory	14	1.359*	1.002	1.843		16
	Voluntary	11	1.347*	1.001	1.813		15
RNR Principle							
Risk						2.55	
	Medium	12	1.14	0.84	1.54		7
	Medium-High	3	1.58	0.88	2.82		22
	High	10	1.63**	1.14	2.34		23
Needs						2.00	
	Low	9	1.09	0.74	1.61		5
	Moderate	6	1.41	0.93	2.14		17
	High	10	1.59*	1.11	2.27		23
Specific Responsivity						4.63	
	Low	7	0.93	0.62	1.40		-3
	Moderate	8	1.37	0.99	1.88		16
	High	10	1.64**	1.20	2.24		24
Composite							
Adherence to RNR						4.20	
	Low	11	1.12	0.83	1.51		6
	Medium	7	1.34	0.90	1.99		15
	High	7	1.90**	1.27	2.85		30

k = Number of comparisons; *OR* = Odds Ratio; *CI*_{95%} = 95% Confidence Interval; *Q*_{Between} = Test of Between-Group Differences (G^2 with *df* = number of categories - 1); *BESD* = Binomial Effect Size Display (% reduction of reoffending against a 50% base rate)

[†]subgroup analysis of the Non-Behavioural programmes alone

**p* < 0.05.

***p* < 0.01.

****p* < 0.001.

The above findings already showed that programmes were most effective when they addressed high-risk offenders, targeted multiple criminogenic needs, and followed the

principle of specific responsivity. To assess the *combined* level of adherence to the RNR principle, we created a composite variable based on each study's score on the three components. According to Rosenthal (1991), this can be done by simply adding the constituent scores when they have similar standard deviations (maximal ratio < 1.5 when there are very few variables). In our dataset this ratio was 1.45; we therefore proceeded with this method. The resulting scores were divided into three levels of treatment adherence to RNR: 'Low' (3-5), 'Medium' (6-7), and 'High' (8-9).

Although the difference between these three categories was slightly above statistical significance there was a clear trend towards larger effects in programmes that adhered more to the RNR model. The effect for 'High' RNR was $OR = 1.90$ ($p < .05$) and indicated a 30% reduction in recidivism.

Methodological Characteristics

Table 3 contains the moderator analyses for various methodological characteristics. Again, the small number of studies in the respective subcategories provided insufficient power for clear significances. However, there were meaningful effects in some subcategories that could not simply be attributed to fishing for significances.

Higher effect sizes were observed among studies that were conducted as part of a demonstration project ($OR = 1.65$, $p < .05$) than studies that were evaluations of extant practice ($OR = 1.23$, $p = .07$).

The amount of control (internal validity) in the design revealed no significant difference ($Q_{\text{Between}}(2) = 1.67$, $p = .43$) and also a minor, nonlinear variation in the means (from $OR = 1.39$ over 1.28 to 1.43). Overall, the differences in quality of research designs and also the use of an RCT were not related to effect size.

Programmes that were delivered at a ‘High’ level of programme fidelity revealed a significant effect ($OR = 1.60, p < .05$), whereas ‘Low’ fidelity showed only a positive tendency. Studies reporting low levels of attrition observed a slightly better outcome (OR of $1.76; p < .05$) than studies with more dropouts, however both effects were significant.

There was a nearly significant trend ($Q_{\text{Between}}(1) = 2.98; p = .08$) of larger effect sizes in smaller samples ($OR = 1.86, p < .05$), but larger studies also showed a positive outcome trend. Effect sizes were a only slightly larger among studies measuring recidivism at short periods of follow-up ($OR = 1.38, p < .05$, versus OR of $1.28, p = .17$ for longer follow up).

Table 3: Effectiveness of treatment according to methodological moderators

Moderator		<i>k</i>	<i>OR</i>	<i>CI</i> _{95%}		<i>Q</i> _{Between}	<i>BESD</i> (%)
				Lower	Upper		
Research Type							
Total Sample Size						2.98	
	Жмлл	10	1.86**	1.22	2.84		29
	>100	15	1.23	0.99	1.52		11
Group Allocation						1.67	
	Moderate Control	13	1.39*	1.04	1.85		16
	Strong Statistical Control	8	1.24	0.84	1.84		13
	Randomized Experiment	4	1.37	0.70	2.68		18
Follow-Up						0.10	
	Жмн а 2yúKă	17	1.38*	1.08	1.77		16
	> 12 Months	13	1.28	0.90	1.84		13
Dropout Rate						2.81	
	Жмџ	8	1.74*	1.09	2.79		26
	Ѓмџ	7	1.51*	1.02	2.25		20
	Not Reported	11	1.10	0.77	1.56		5
Research Type						1.76	
	Demonstration	7	1.65**	1.15	2.37		24
	Practice	18	1.23	0.98	1.55		11
Programme Fidelity						1.87	
	Low	14	1.21	0.94	1.54		10
	High	11	1.60**	1.16	2.20		23
Composite							
Methodological Quality						0.48	
	Low	4	1.49	0.86	2.58		21
	Medium	14	1.35	0.98	1.85		15
	High	7	1.17	0.75	1.83		8

k = Number of comparisons; *OR* = Odds Ratio; *CI*_{95%} = 95% Confidence Interval; *Q*_{Between} = Test of Between-Group Differences (G^2 -distributed with *df* = number of categories - 1); *BESD* = Binomial Effect Size Display (% reduction of reoffending against a 50% base rate)

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

In an effort to synthesize the results from the moderator analyses of methodological variables, we computed a composite score by adding together each study's score on all the methodological moderators. Studies were given a score of 1 for each of the following values: large sample size, longer follow-up measurement, low attrition, high treatment fidelity, and strong statistical control between experimental groups. A randomised experiment was accorded a further point, to signify its particular internal validity. Furthermore, studies that did not report the level of attrition in their sample were given a score of 0.5 on this item. The ratio of the largest to the smallest standard deviation in the variables was 1.92, which is defensible because of the number of variables being combined. The composite score was divided into three categories of overall methodological quality: 'Low' (0 to 1), 'Medium' (1.5 to 3), and 'High' (3.5 and above).

Overall there was no significant difference between the three subcategories ($Q_{\text{Between}}(2) = 0.48, p = .79$) and also no significant effect for any of them. One can only mention a very slight trend of smaller effects in methodologically stronger studies.

We also conducted a hierarchical regression analysis to control for the effect of confounded moderators in our meta-analysis (Table 4). We controlled sequentially for proportions of outcome variance attributable to methodological, contextual, and delivery factors in order to isolate the influence of treatment type on reoffending outcome. We included a parsimonious collection of independent variables due to the small number of studies in our sample. In contrast to the random effects models employed in the preceding bivariate analyses, we used a fixed effects model on the assumption that the included variables would considerably reduce the observed heterogeneity in the overall model.

The final model accounted for 52% of the total outcome variance

Table 4: *Hierarchical regression*

Moderator		Standardized β			
Methodology	Composite Methods Score	.20	.14	.1	.06
	UK Evaluation vs. Non-UK Evaluation	-.50***	-.56***		-.35*
Delivery	Custody vs. Community Setting			-.33†	-.53*
	Voluntary vs. Obligatory Participation			.48**	.36†
Treatment Type	Cognitive Behavioral/Behavioral Intensive Supervision/Deterrence				.47†
					-.06
ΔR^2		.04***	.24***	.11***	.13***
Q_{Model} (df)		2.63 (1)	18.34 (2)***	25.33 (4)***	33.38 (6)***

† $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Sensitivity Analysis

Although we took every precaution to ensure that all eligible studies were represented in our meta-analysis we did not find any unpublished evaluations that met our criteria for inclusion (see above). To check the potential influence of a publication bias we conducted a Fail-Safe N test. This determines the number of studies with a null effect that would have needed to be included in our meta-analysis in order to render the effect statistically non-significant. The result was that 96 studies would have been required to render the mean effect statistically non-significant. It is highly unlikely that in our exhaustive literature search we may have overlooked so many studies.

We also carried out a funnel plot test for bias in the study sample. This displays the logged effect sizes on the x-axis against their respective standard errors on the y-axis (see Figure 2). Hence, studies with smaller sample sizes will concentrate at the bottom of the graph and large studies at the top. Asymmetry in the distribution of the effect sizes on either side of the vertical line would indicate a publication bias. The relative symmetry found in the funnel plot suggests that our meta-analysis was not heavily biased in any particular direction. The ‘Trim and Fill’ technique displays the difference in effect sizes that could be attributable to bias (Borenstein et al., 2009:286), by imputing effect sizes until the error distribution more closely approximates normality. The imputed effect sizes are visible on the funnel plot as

solid black dots, and the solid black diamond at the bottom renders the shifted summary effect size, had the distribution been normal. In Figure 2, the random-effects Trim and Fill analysis revealed that an imputed *OR* would have accounted for an increase of the overall effect size from 1.343 to 1.358. This difference is negligible and even in favour of a *lower* recidivism due to treatment. This suggests that our average is a conservative estimate.

As a third test of robustness we carried out an outlier analysis. Our meta-analysis contained not only a very broad range of effect sizes (*ORs* from 0.560 to 6.98), but also very different sample sizes (from 26 to 3,068). Although the exclusion of certain evaluations would affect the results of the moderator analysis (where statistical power is much lower), the exclusion of outliers had negligible effects on the total effect size. For example, the exclusion of the studies with the largest sample size slightly *increased* the overall effect size and resulted in $OR = 1.37$ ($p = 0.008$) when Cann et al. (2005) was excluded or $OR = 1.39$ ($p = 0.003$) when St. James-Roberts et al. (2005) was excluded. Furthermore, had we used only the most severe reoffending measure alone, rather than a mean of all reported outcome scores, the effect size for Lösel and Pomplun (1998) would have been positive ($OR = 1.77$, rather than 0.90), however, even with this analysis, our mean effect size would not have been substantially different ($OR = 1.37$, $p < .005$).

Using intent-to-treat outcome data in studies where they were reported changed the summary *OR* from 1.34 ($p < .05$) to 1.30 ($p < .05$). This difference is small, and suggests that our overall effect is robust.

Figure 2: *Funnel Plot of Standard Error by Log Odds Ratio*

N.B. Solid points represent imputed scores from the 'Trim and Fill' analysis.

Discussion

The mean overall effect size of $OR = 1.34$ ($r = .082$; $d = .16$) is noteworthy for five reasons: First, with the exception of a subset of data in Redondo et al. (1999, 2001) it is the only demonstration of an overall positive effect of European programmes for young offenders. Second, according to our sensitivity analyses it seems to be a relatively robust estimate of effect size. Third, it is in the same range as in other reviews of juvenile offender programmes (e.g., $r = .08$ in Cleland et al., 1997; $r = .09$ in Dowden and Andrews, 1999; $r = .09$ in Gensheimer et al., 1986; $r = .12$ in Gottschalk et al., 1987; $r = .09$ in Latimer et al., 2003; $r = .05$ in Lipsey, 1992; $r = .12$ in Redondo et al., 2001). Fourth, according to the BESD this small effect equates to approximately 8 percentage points or 16 per cent less recidivism in the treatment group, given a base rate of 50 percent in the control group. Fifth,

although we cannot provide sound European data on cost-benefits here, even such moderate reductions in recidivism seem to be practically significant in both a human and a monetary sense (e.g., see Aos et al., 2001; Welsh and Farrington, 2001).

As in other meta-analyses on offender treatment the mean overall effect tells only a small part of the story on the outcomes. Consistent with the literature, cognitive-behavioural and behavioural treatments showed larger effects than other types of programmes. Non-behaviourally-oriented programmes revealed no significant positive effect, whereas deterrence- and supervision-based interventions even resulted in slightly (although not significantly) *increased* recidivism. These findings are in accordance with North American research (Andrews and Bonta, 2010; Lipsey and Cullen, 2007; MacKenzie, 2006). With regard to non-behavioural programmes our review could not provide differentiated results because this category was very heterogeneous and contained too few studies on specific programme types. Therefore, one should only mention the nonsignificant positive outcome tendency of educational and vocational programmes (4 studies).

Applying the RNR model revealed the strongest mean effect ($OR = 1.90$) for those programmes that fulfilled all three principles. Here, the BESD indicated a substantial reduction of 30% in recidivism. In spite of repeated criticism of the RNR model (e.g. Ward and Brown, 2004; Ward and Maruna, 2007), our review showed that these principles are robust when it comes to empirical evidence. Our results support each of these constituent claims in isolation as well as their effect when in concert. Unfortunately, the sample of available studies did not contain sufficient information to analyze extensions of RNR that contain relationship, staff and organizational issues as applied in programme accreditation (Andrews et al., 2011; Maguire et al., 2010). The moderate number of studies also precluded estimating the relative importance of one or the other principle (e.g., Andrews and Bonta, 2010; Lowenkamp et al., 2006). However, one should also bear in mind that as in most RNR

research our studies addressed group-oriented risks, needs, and responsivity and not individualised approaches.

Although our findings support the particular strength of cognitive/behavioural and RNR-appropriate programmes we need to mention a few other aspects: Only 7 out of 25 studies addressed programmes that adhered closely to the RNR principle, i.e. a minority of evaluated treatments in Europe are following these replicated principles. The mean effect size of the most appropriate RNR interventions and of the cognitive/behavioural programmes was also roughly similar to what has been reported in other reviews. For example, in comparison to our circa 30% reduction in recidivism, North American meta-analyses of cognitive-behavioural juvenile offender treatment found mean effects of 16% (Wilson et al. 2005), 25% (Landenberger and Lipsey, 2005) and 33% (Lipsey et al. 2001). There is no simple regional explanation for this difference because Redondo et al. (1999, 2001) reported 23% reduction in reoffending for CBT programmes in Europe. The small number of primary studies limited our investigation of other moderators and multivariate analyses (see Lipsey, 2003). However, we found a number of relevant trends: The tendency of larger effects of programmes in the community as opposed to those conducted in secure institutions is in accordance with North-American research (e.g. Lipsey and Cullen, 2007; Lipsey and Wilson, 1998). The advantage of community programmes may partially result from criminogenic effects of imprisonment (Durlauf and Nagin, 2011). However, the latter research did mainly address incarceration without treatment. One must also take into account that our primary studies did not directly compare institutional and community treatment; rather, they examined differences to control groups within the same setting. However, from a clinical and educational perspective it is plausible that community programmes show smaller effects because they contain more opportunities for real life application and transfer (Lösel and Schmucker, 2005). Therefore, practice should aim for community instead of custodial programmes as far as criteria of

compensation of guilt and public security allow non-custodial sentences. It is also of practical relevance that we found no outcome difference between voluntary and mandatory programme participation. This speaks against older dichotomous assumptions about treatment motivation and supports more process-oriented concepts in which external triggers can lead to intrinsic change over time (McMurran, 2002).

Because of the small number of studies in the subcategories of our meta-analysis the effect sizes were not significantly related to individual components of methodological quality. However, the pattern of outcomes clearly followed similar trends as in the literature on correctional treatment. For example, we found larger effects among studies using small samples. This is not only in accordance with research on juvenile offender treatment (Lipsey and Wilson, 1998) but also on preventive skills training (Lösel and Beelmann, 2003). This finding may have been influenced by a publication bias (small samples require larger effects to become significant) or a better quality of implementation in smaller studies. As our review contained only published studies we could not test the first hypothesis. However, the second one is at least partially supported by our result of larger effects in studies with better fidelity, although one cannot exclude the possibility that integrity/fidelity problems were mainly reported when effects turned out lower than expected.

In accordance with the North-American research we also found a tendency of larger effects in studies that contained longer follow-up periods and evaluated demonstration projects (e.g. Lipsey and Cullen, 2007; Lipsey et al., 2006). The latter indicates the difference between ‘efficacy’ in model projects and ‘effectiveness’ in routine practice. Effectiveness evaluations typically reveal smaller effects, not only in correctional treatment. In demonstration projects the study authors are often involved in the development or delivery of programmes that also coincide with larger effect sizes (e.g. Lipsey and Wilson, 1998; Lösel and Schmucker, 2005). However, we could not investigate this aspect in the present review

because most primary studies did not contain such combinations. This may also have contributed to the somewhat smaller effects in European CBT programmes as mentioned above.

The type of evaluation design did not show a systematic relationship to effect size. This is in accordance with various reviews of correctional treatment (Lipsey and Cullen, 2007; Lösel, 1995), although the international findings on this issue are not fully consistent. Broader analyses of crime prevention programmes revealed smaller effects in studies of high internal validity such as RCTs (Weisburd et al., 2001) and our results on the integrative score of methodological quality points in the same direction. This supports the view that methodological quality should not only be reduced to one dimension but take various threats to validity into account (Farrington, 2003; Lösel and Köferl, 1989; Lösel, 2011).

Although most of our findings were consistent with the North-American ‘what works’ literature, we are not able to conclude that the latter can simply be generalised to Europe. This is due to the relatively small number of well-controlled European studies we could retrieve in our intensive search process. Two thirds of the studies in our sample were carried out in the United Kingdom where the culture and legal system is more similar to the United States than that of most continental nations. For the large majority of European countries we could not find any evaluation with sufficient methodological quality. And for those countries outside the UK that are represented in our review, there were only one or two studies on a specific type of programme.

For these reasons our review does not justify far-reaching conclusions on ‘what works’ in young offender treatment in the whole of Europe. One must bear in mind that juvenile justice systems vary even more than their adult counterparts (Doob and Tonry, 2004). Differences in the classification of offending, in responses to reoffending, in the age of

criminal responsibility, and in the handling of late adolescent or young adult offenders are only a few aspects of such variation across jurisdictions. For example, it is doubtful that the secure custody measure evaluated in Curran et al. (1995), which was applied to Northern Irish youths of roughly 14 years of age, would have been applied to their peers in the Netherlands, where the juvenile justice system is far more welfare-oriented (Weermann 2007:262), or Sweden, where the juvenile justice system is more welfare-oriented still (Janson, 2004). Furthermore, the very notion of ‘treatment as usual’ in the control groups may hold different meaning across different countries. For example, in our meta-analysis, the very same intervention was applied to youths of similar age and psychological profile in Sweden (Sundell et al., 2008) and in Norway (Ogden and Hagen, 2006; Ogden et al., 2007), yet only the latter two evaluations found a significantly positive effect. The authors attribute this discrepancy to the more punitive approaches of how youths are usually handled in Norway (Sundell et al., 2008; Andrée Löfholm et al., 2009). Similar aspects of punitivity need to be regarded in countries such as the United Kingdom (Bottoms and Dignan 2004).

Taking such issues into account, our review has three major conclusions: First, the overall effect of young offender treatment in Europe is encouraging, in particular for cognitive/behavioural programmes and the RNR model. Second, there are various moderators of effect that could partially be disentangled in our review. Third, our meta-analysis revealed a strong deficit of well-controlled programme evaluations in most of the European countries. As a consequence we need a European policy to increase well controlled programme implementation and evaluation of ‘what works’ in young offender treatment across the continent.

Appendix Ai: Keywords used to search electronic databases

<i>Target Population</i>	Youth Violen* OR Delinquen* OR Juvenile OR Hooligan* OR Hate AND
<i>Intervention</i>	Programme* OR Treatment* OR Interven* OR Correcti* OR Therap* OR Counsel* OR Mentor* OR Rehabilitati* OR Cogniti* OR Relapse OR Boot Camp* OR Wilderness Challenge* OR Intensive OR Incarcerat* OR Court* OR Probation OR Mandated OR Inmate* OR Institution* OR Non-Institution* OR Prison* AND
<i>Outcome</i>	Effect* OR Outcome* OR Eval* OR Experiment* OR RCT* OR Quasi* OR Trial* OR Empirical* OR ES* OR Recidiv* OR Reoffen*

Appendix Aii: Searched databases

Database	Number of Studies
<i>Bibliographic Databases</i>	
International Bibliography of the Social Sciences	795
PSYCINFO	4,018
PSYCARTICLES	186
PUBMED	336
COCHRANE Library	67
EMBASE	822
ISI Web of Knowledge	4,018
Cambridge Scientific Abstracts (CSA)	12,105
C2 SPECTR (ERIC)	1,672
Science Direct	2,282
<i>Governmental Publications</i>	
UK Home Office Research Database	5
Brå-Swedish National Council for Crime Prevention	0
<i>Internet Resources</i>	
Google Scholar	N/A
MetaCrawler	N/A
<i>Unpublished Resource Databases</i>	
Dissertation Abstracts International	683
Total	26,989

A Systematic Review and Meta-Analysis on the Effects of Drug Treatment Programmes to Reduce Reoffending in Europe*

Johann A. Koehler¹, David K. Humphreys¹, Thomas D. Akoensi¹,

Olga Sánchez de Ribera¹ and Friedrich Lösel^{1,2}

¹ Institute of Criminology

University of Cambridge, UK

² Institute of Psychology

University of Erlangen-Nuremberg, Germany

* This study was carried out within the European cooperation project “Strengthening Transnational Approaches to Reducing Reoffending”. The project was funded by the European Union. We thank our partners from the UK Ministry of Justice, London Probation Trust, European Organization for Probation, and the Ministries of Justice of Bulgaria, France and Hungary for their cooperation. We also thank primary study authors for providing us with the data to assist in the meta-analysis.

Abstract

North American evidence indicates that treatment programmes to reduce substance abuse can decrease reoffending, although this research often suffers from threats to internal validity. This article presents the results of a systematic review and meta-analysis on the effectiveness of treatment programmes to reduce recidivism among drug abusing offenders in Europe, using stringent eligibility criteria to control for such threats. A literature search of approximately 37,000 titles revealed 15 evaluations from six European countries that fulfilled the eligibility criteria (e.g., control group design, offending behaviour outcome). The study sample sizes ranged from 37 to 1,205, and in total the studies contained 1,698 treated offenders and 2,255 participants in the control groups. The average age was 33.17 ($SD = 6.9$). The length of follow-up varied between 3 and 36 months; treatment dropout ranged from 5% to 71%. The majority of studies used a randomised design and self-reports on reoffending. Effects ranged from *Hedges'* $g = 0.000$ to 1.912, and there was a significant positive mean effect ($g = 0.355$) that indicated a 32% reduction in reoffending against a baseline of 50%. This overall finding was robust in a sensitivity analysis. Most evaluations examined the effect of pharmacological treatment (maintenance, substitution) on opiate-dependent offenders who have failed in previous treatment. Such treatments were partially combined with counselling and other therapeutic measures, but evaluations of purely psychosocial interventions were a minority. Programmes that delivered primarily pharmacological substitution treatment were more effective than those that did not. Studies conducted in the United Kingdom reported smaller effects than those in other European countries, but this was mainly due to more non-pharmacological studies in the UK. Although our findings support the use of treatment programmes to reduce reoffending among substance abusing offenders, more research is needed on the effect of different programmes on different types of offender. Furthermore, in most European countries there was no sound evaluation at all. This limits the generalisability of the results across the continent.

Introduction

Despite continuing debate around the ‘drug-crime connection’ (e.g., see Hammersley, 2011; Bennett and Holloway, 2009), there remains some uncertainty about the strength and direction of this relationship. Many studies have reported convincing evidence identifying drug problems in offender populations. A systematic review of U.S., European, and Australasian prisoners found that drug abuse was prevalent in 10% to 48% of the male prison population, and in 30% to 60% of female prisoners (Fazel et al., 2006). A survey of arrestees in the UK found that up to 80% of the sample reported using illegal drugs (Holloway and Bennett, 2004). Results from a recent meta-analysis (Bennett et al., 2008) found that the odds of offending were between 2.8 to 3.8 times greater for drug using populations, and that this association was strongest among users of addictive, as opposed to recreational, drugs.

Such statistical associations can represent a number of possible relationships between drug use and crime. For example, the connection may be ‘direct’ (i.e. drug use could cause people to commit crimes); or ‘indirect’ (i.e. that drug use leads to crime indirectly as a moderating or mediating factor); the relationship could be ‘reciprocal’ (i.e. that drug use causes crime, and crime also causes drug use); or due to a ‘common cause’ (i.e. a factor causing both drug use and crime to occur). Uncertainty about the causal direction of this relationship remains an obstacle for developing effective interventions to reduce the harm (such as crime) associated with illegal drug use. Despite this obstacle, during the last fifty years there has been considerable innovation in the development (and implementation) of interventions to treat offending populations for symptoms of drug abuse.

The variety of theoretical explanations for the drugs-crime connection has prompted the development of many different approaches for treating drug abuse and dependency (see UK Drug Policy Commission (2008) for a thorough overview). Many of these treatments focus on helping individuals manage the symptoms of physical dependence, such as

pharmacological replacement treatments (i.e. methadone, heroin treatment, naltrexone). Other approaches focus on forming supportive environments that may help individuals reduce usage of illegal drugs, or cope with factors that result from, or lead to, their abuse of drugs. Examples include: therapeutic communities, psycho-social treatment, and group and individual counselling. Treatments for individuals convicted of crimes are also available; services such as those previously mentioned may be available for offenders to attend on a voluntary basis. Often, mandated treatments (i.e. referral to treatment, drug testing orders, drug courts, and boot camps) are used to force individuals into abstaining from drug abuse and criminal behaviour. Contrary to claims that 'nothing works' (Martinson, 1974), research synthesising the effects of drug interventions have generally suggested that doing something is better than doing nothing. Already in the 1990s meta-analytic studies have shown that pharmacological treatment (Marsch, 1998) as well as psychosocial treatment (Carroll, 1996; Irvin et al., 1999) can reduce drug use and prevent relapse.

More recent systematic reviews have focussed specifically on the effects of drug interventions in reducing criminal involvement and reoffending and have found general reductions. However, there are conflicting findings for the types of drug interventions that are considered most effective. For example, Prendergast et al. (2002) found that drug abuse treatment could help to reduce crime involvement regardless of the treatment modality (*Hedges' g* = .16, $p < .05$, fixed-effects model). Holloway et al. (2008) revealed an average 26% reduction in criminal behaviour in their meta-analysis, but instead found stronger effects for psychosocial and therapeutic communities. On the contrary, Egli et al. (2009) analysed the effectiveness of pharmacological treatments compared to other treatments. They found that maintenance programmes showed a significant reduction in offending (OR = 1.55, CI_{95%} = 1.18, 2.02) when compared to other modes. In other studies, statistically significant reductions were noted for the impact of therapeutic communities (Pearson & Lipton, 1999),

drug court interventions (Lowenkamp et al., 2005; Latimer et al., 2006), and prison-based drug treatments (Mitchell et al., 2006). Consistent with Bennett et al. (2008), Mitchell et al. found stronger effects for therapeutic communities and counselling interventions.

The findings of existing reviews present a promising outlook for the potential of drug interventions to reduce reoffending in different populations and settings. However, despite the growth in this research area over recent years, the current evidence base comes almost entirely from primary studies conducted in North America¹, thus limiting the generalisability of findings to European and other contexts. As Lösel (1995) or Pawson and Tilley (1997) advocate, it is necessary to go beyond ‘what works’ to investigate *why something is effective, for whom, and in what circumstances*. The latter issue is particularly relevant for the treatment of drug-addicted offenders because legal regulations, treatment practices, settings, and offender populations in many countries differ from North America. Therefore, this study aims to synthesise the European evidence of the effectiveness of drug interventions seeking to reduce reoffending, which is currently underrepresented by the international research literature.

Methods

Eligibility Criteria

- i. We limited our search to evaluations conducted in Europe.
- ii. The intervention had to be targeted at the drug use and/or related behaviours or attitudes of populations abusing illegal drugs.
- iii. The evaluation had to compare the effect of an intervention, as applied to a treatment group, to the level of reoffending in a control group. The latter could be defined as the

¹ In an initial scoping project we found 15 meta-analyses researching the effectiveness of various drug interventions. These 15 studies contained $k = 45$ treatment comparisons. Only $k = 29$ of these comparisons reported information of the country origin of primary studies, and of this subgroup, only 13% ($k = 4$) were primary studies conducted in European countries.

application of no treatment, treatment as usual, or placebo treatment. In an effort to achieve equivalence between experimental groups, we excluded studies if the effectiveness of treatment was compared to: national statistics of the general offender population; a sample of participants who were administered the same programme, but entered treatment through different routes (e.g., comparing outcomes of participants who entered treatment voluntarily with those who were court-ordered into the same programme); and treatment drop-outs and non-completers. We excluded studies that compared two types of experimental treatment (e.g., methadone maintenance vs. therapeutic community), and studies whose control group comprised subjects who received the same treatment but in a smaller dose. We did not exclude studies on the basis of how subjects were allocated to the experimental groups.

- iv. Data had to be reported in sufficient detail to allow the computation of effect sizes. This data had to concern reoffending as an outcome, which was understood either as a formal institutional measure (e.g., re-arrest, re-conviction, re-incarceration, etc.), or as self-reported data pertaining to crime. We excluded data that were based on indirect measures such as illegal income.
- v. We included studies written in any European language.
- vi. We included published and unpublished studies.
- vii. We included interventions only if they were formulated in conventional, programmatic terms, as opposed to broader analyses of national-level policy changes.

Literature Search

In order to locate unpublished and published studies, we searched online computerised databases² as well as meta-analytic and systematic review publications dealing with drug abusing offender treatment programmes. We also contacted academics and other experts in

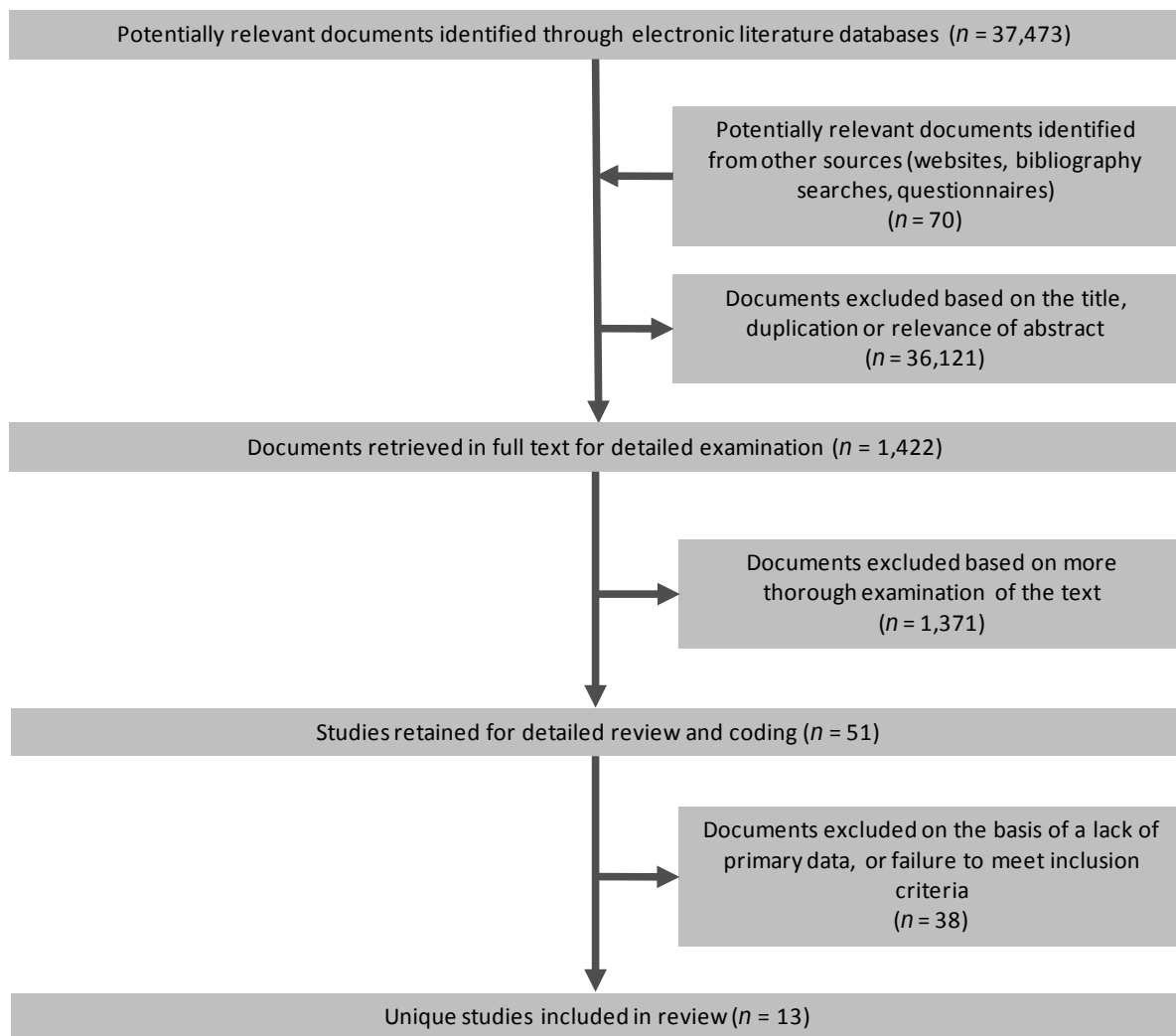
² A full list of the search terms employed in our database search can be found in Appendix Bi.

an effort to locate studies that might not have been accessed by the more conventional strategies. Snowball sampling methods were used to locate inaccessible respondents.

Our bibliographic database search yielded a total of 37,473 titles, which, upon deletion of duplicates, yielded 30,421 discrete studies. A further 70 potentially relevant studies were identified from bibliography searches and website searches. Preliminary screening of titles and abstracts for *prima facie* relevance to our eligibility criteria yielded 1,422 studies (for further detail regarding the eligibility criteria, see above). These studies were acquired in either electronic or hard formats and were screened according to method, location of study, sample population, and outcome of interest, in order to arrive at 51 studies. These studies were then retained for more detailed review and coding. Upon examination of the full text documents we excluded a further 38 studies that did not meet our inclusion criteria. This resulted in 13 discrete studies, comprising 15 controlled evaluations (see Figure 1).

Our search uncovered only one unpublished evaluation (McSweeney, 2009). Anticipating the potential bias this may have introduced into our results (Lipsey and Wilson, 2001; Wilson, 2009), we took the precaution of administering a survey of drug treatment reoffending programmes to experts and practitioners throughout the 27 countries of the European Union (Hamilton et al., 2011; Appendix D). However, the survey revealed that although some form of outcome evaluations have been conducted in 48% of our sample, none met our criteria of substantive or methodological rigour. As a result, this yielded no further contributions to our study sample. We therefore proceeded with statistical analysis to examine whether our data were significantly affected by publication bias (below).

Figure 1: *Literature search flowchart*



Description of the Studies

Our search uncovered 15 controlled evaluations, with a total N of 1,698 in the treatment groups and 2,255 in the control groups. The average age of the participants was 33.17 ($SD = 6.9$).

In the following we provide a description of the studies included in our sample. We defined all conditions approximating treatment-as-usual as the control group, and where newer treatments were compared to more commonly found treatment types, the new treatment was defined as the experimental condition. Given that methadone maintenance

treatment is the conventional programme delivered to opiate-dependent populations in many European countries, we coded this as a treatment-as-usual condition in all cases except Hartnoll et al. (1980; see below). A summary of the main study features can be found in Table 1.

Heroin Maintenance Programmes

Three studies compared the effectiveness of pharmaceutical heroin in the experimental condition to methadone in the control condition, in a sample of heroin addicts in Switzerland (Perneger et al., 1998) and in the UK (McCusker and Davies, 1996; Metrebian et al., 2001).

McCusker and Davies (1996) matched 39 control group participants attending a community drug treatment clinic to 27 roughly equivalent participants in the experimental group according to significant variables. Clients received pharmacologically equivalent doses of each prescribed drug, in addition to psychological counselling. Outcomes were collected by means of a structured interview and self-report questionnaire capturing data including the number of days, in the last 30, in which they had been involved in illegal activities. Although more than a third (36%, $n = 14$) of the experimental group participants had dropped out and were excluded from the follow-up analysis, only 4% ($n = 1$) dropped out of the control group; nonetheless, reported criminal activity was statistically significantly lower in the heroin group than in the methadone group.

Metrebian et al. (2001) offered the choice of receiving heroin or methadone to opiate-dependent participants who had experienced difficulties with oral methadone treatment. Of 58 participants recruited to the study, 37 chose heroin and 21 chose methadone. A ceiling dose of 200mg per day was set for both groups. Outcome data consisted of a self-report assessment instrument measuring criminal activity administered at three and twelve months

after admission into treatment. Results were based on analyses of the 15 (41%) participants who remained in treatment in the experimental group, and 11 (52%) who remained in treatment in the control group. Many participants elected to be in the control group on the basis that the ceiling limit for prescribed heroin was too low to support their level of drug use. Statistically non-significant reductions were observed in the heroin group, compared to the methadone group.

Perneger et al. (1998) evaluated an experimental heroin maintenance programme for patients who had experienced multiple unsuccessful attempts at drug treatment. 27 heroin addicts were randomly allocated to an experimental condition in which heroin was administered daily in an outpatient clinic; furthermore, patients occasionally received oral opiates, and clorazepate substitution. 24 roughly equivalent participants who were allocated to the control condition were encouraged to select any drug treatment programme; almost all control participants entered a methadone maintenance programme. Outcome was measured by a self-report questionnaire, capturing the number of drug and property crime charges in the six months previous to programme entry and at six months' follow-up, and another measure captured whether the participant had been charged with any offense in the last six months. Two experimental group participants discontinued treatment but were included in the final analysis. Statistically significant reductions in criminal activity were observed in the experimental group, compared to the control group.

Methadone Maintenance Programmes

Two studies examined the effectiveness of experimental methadone maintenance programmes among opiate addicts in the UK (Strang et al., 2000; Hartnoll et al., 1980).

Table 1: Summary description of primary studies in the systematic review

Author	Country	Treatment Type [†]		Group Allocation	Dropouts (%) [‡]		Follow-up (Months)	Outcome Measure
		Experimental	Control		Experimental	Control		
Hartnoll et al. 1980	UK	MA; PS (52)	HA; PS (44)	Random	71% (ITT)	36% (ITT)	12	OR
Lobmaier et al. 2010	Norway	NX (23)	MA (21)	Random	30% (ITT)	48% (ITT)	6	SR; OR
Löbmann & Verthein 2009	Germany	HA; MA* (515)	MA* (500)	Random	33% (ITT)	60% (ITT)	12	SR; OR
March et al. 2006	Spain	HA; MA* (27)	MA* (23)	Random	15% (ITT)	9% (ITT)	9	SR
Martin et al. 2003	UK	DT; PS (274)	NT (931)	Non-Random	(274)	(931)	12; 24	SR
McCusker & Davies 1996	UK	HA* (27)	MA* (39)	Non-Random	4% (26)	36% (25)	12	SR
McSweeney 2009	UK	DT (242)	NT (252)	Non-Random	(242)	(252)	12	OR
Metrebian et al. 2001	UK	HA (37)	MA (21)	Non-Random	41% (22)	52% (10)	3; 12	SR
Naeem et al. 2007	UK	DT; PS (35)	NT (38)	Non-Random	29% (25)	29% (27)	12	SR
Perneger et al. 1998	Switzerland	HA* (27)	MA* (21)	Random	7% (ITT)	(ITT)	6	SR
Robertson et al. 2006	UK	HA (108)	MA (110)	Random	18% (ITT)	3% (ITT)	36	SR
Strang et al. 2000	UK	MA* (19)	MA* (18)	Random	5% (18)	7% (15)	6	SR
van den Brink et al. 2003; Study 1	Netherlands	MA; HA* (76)	MA* (98)	Random	28% (ITT)	15% (ITT)	12	SR
van den Brink et al. 2003; Study 2	Netherlands	MA; HA* (117)	MA* (139)	Random	32% (ITT)	13% (ITT)	12	SR
van den Brink et al. 2003; Study 3	Netherlands	MA; HA* (119)	MA* (139)	Random	31% (ITT)	13% (ITT)	12	SR

[†]Numbers in brackets denote the original sample size in each experimental condition

[‡]Numbers in brackets denote the analysed sample size in each experimental condition; ITT denotes that the value is the same as in the 'Treatment Type' column

*Denotes that additional therapeutic treatment was offered

Legend: MA: Methadone-Assisted Treatment NX: Naltrexone Implants NT: No Treatment OR: Official Records ITT: Intent-to-Treat Analysis
HA: Heroin-Assisted Treatment PS: Psycho-Social Care DT: Drug Testing SR: Self-Report

Strang et al. (2000) compared the effectiveness of injectable methadone with oral methadone maintenance among opiate addicts who had experienced at least one previous episode of opiate substitution treatment. The authors randomly assigned 19 participants to the experimental condition, in which daily methadone was prescribed and injected in a dedicated clinic. In the control condition, oral methadone was dispensed daily to 18 participants and was consumed under supervision. Patients in both conditions received counselling and clinical support. Participants in each group were roughly equivalent on important variables. Outcome data were collected at six months after intake into the programme by means of a self-report instrument that captured the number of days on which acquisitive crime had been committed. 15 control group participants and 18 experimental group participants remained in treatment at follow-up. Treatment compliance was high in both groups. Acquisitive crime decreased more – although not statistically significantly so – in the experimental group, compared to the control group.

Hartnoll et al. (1980) randomly allocated 96 addicts who had both requested heroin maintenance and had rejected other alternative treatments to receive either their requested heroin course or methadone. The heroin condition most closely approximated ‘treatment-as-usual’, and was therefore defined as the control, whereas methadone treatment was defined as the experimental condition. The groups were roughly equivalent on all important variables. 74% of control group patients and 29% of experimental group patients were still receiving treatment at 12 months’ follow-up. Arrest data collected from official records revealed that more arrests were observed among patients in the experimental group at follow-up than in the control group. However, whereas the criminal activity in the control group was relatively homogeneous, the authors noted that experimental group participants could be disaggregated into two profiles of criminal activity: in the first, they observed almost total abstinence from both crime and drug consumption; in the second, where the level of drug consumption

remained roughly equivalent to the year before the trial, criminal activity increased significantly.

We did not include this study in our meta-analysis on the grounds that the treatments applied in the experimental conditions – heroin in the control, methadone in the experimental – were an inversion of the specified treatment conditions in the remainder of our study sample (see Table 1). This is in large part because the experiment was conducted during the early to mid-1970s, at a time when drug policy differed from modern applications. Furthermore, the experimental condition was applied coercively and contrary to the participants' stated desire for heroin treatment; this compromises the ability to draw conclusions regarding equal motivation between groups. Therefore, in order to compare 'like with like' in the meta-analytic model, the study was excluded.

Heroin and Methadone Treatment in Combination

Three studies examined the effectiveness of combined methadone and heroin treatment programmes among opiate addicts in Germany (Löbmann and Verthein, 2009), Spain (March et al., 2006), and the Netherlands (van den Brink et al., 2003).

Löbmann and Verthein (2009) compared the effectiveness of heroin-assisted treatment to methadone maintenance in outpatient facilities. Participants reported either no recent participation in therapy or a negative course of treatment. 515 participants were randomly assigned to receive diacetylmorphine and methadone over the course of three separate visits each day in the experimental group; 500 roughly equivalent control group participants received methadone on a single visit each day. Both groups received psychosocial support. An elevated attrition rate was observed in the control group (60%), compared to the experimental group (32.8%); missing data at 12 months' follow-up post admission into

treatment were replaced with baseline observations. Outcome data capturing the mean number of crimes per year in various categories (drug offences, violent crime, property crime, and fraud) were measured by means of a questionnaire, and also from police records. Statistically significant reductions in crime were observed in the experimental group, compared to the control group.

March et al. (2006) compared the effectiveness of diacetylmorphine to methadone among opioid-dependent individuals for whom standard treatments had failed. 31 participants were randomly assigned to receive diacetylmorphine twice daily in addition to methadone once daily under supervision in the experimental condition, and in the control condition methadone was dispensed to 31 roughly equivalent patients once daily. Both groups received additional psychological and social support. Outcome data measuring the number of days involved in illegal activities in the previous month were collected by means of a questionnaire administered at entry into the programme, and at nine months' follow-up. 74% of experimental group participants and 68% of control group participants completed treatment. Intent-to-treat analyses revealed that crime decreased statistically significantly more in the experimental group at nine months' follow-up, compared to the control group. The attrition rate was roughly equivalent between groups.

Van den Brink et al. (2003) conducted two separate randomised controlled trials to investigate the effect of prescribing heroin to treatment-resistant clients. Clients were allocated to one of the two trials based on how they habitually consumed heroin, whether through inhalation or injection. The first trial (involving the predominantly injecting subsample, hereafter referred to as Study 1) randomly allocated 98 participants to a control group that continued with methadone treatment at an existing clinic with existing staff and 76 participants to an experimental group in which a programme of co-prescribed methadone and heroin was administered in a new treatment facility with specially recruited staff. The other

participants, who predominantly inhaled their heroin, were allocated to the second trial. Here, an identical design was used to constitute Study 2 ($n = 139$ in the control group; $n = 117$ in the experimental group), and a final comparison between the control group and a treatment group that was administered methadone alone, and subsequently methadone plus heroin ($n = 119$) constituted Study 3. Psychosocial treatment was offered in all five treatment conditions. Outcome data collected from questionnaires at twelve months after admission into the programme captured the number of days of illegal activities in the 30 days preceding assessment, and these data were corroborated with official police records. Although the authors employed intent-to-treat analyses, treatment completion rates in the control and experimental groups were approximately 85% and 70%, respectively. No significant differences were observed between the groups at baseline, and although reductions in criminal activity were observed across the entire sample, the most marked differences were among the three experimental groups. Improvements were observed for treatment, compared to control, in all three studies, but these were only statistically significant in studies 1 and 2.

Other Pharmacological Programmes

Two studies examined the effectiveness of miscellaneous experimental pharmacological treatment programmes among opiate addicts in Norway (Lobmaier et al., 2010) and in the UK (Robertson et al., 2006).

Lobmaier et al. (2010) examined the effectiveness of experimental naltrexone implants compared to methadone maintenance as a control condition among in Norway prior to release from prison, and continued in the community upon release. The authors randomly allocated 23 inmates to the experimental group and 21 inmates to the control group. Participants were free to seek additional therapeutic and social service support, although they did so rarely (36.4% in the methadone group, 50% in the naltrexone group at six month follow-up).

Questionnaires administered before arrest and at follow-up six months after release from prison captured mean days per month engaged in criminal activity; this was supplemented with official incarceration data. Missing data at follow-up were replaced with baseline observations. Ten control group inmates and seven experimental group inmates did not initiate treatment. A further two control group inmates discontinued methadone use before release from prison. The experimental group manifested marginally greater reductions in self-reported criminal activity and reincarceration at follow-up than the control group, although this was not statistically significant.

Robertson et al. (2006) randomly assigned 119 participants to a control group to receive daily methadone; they then assigned 116 participants to an experimental group, to receive an equivalent daily dose of dihydrocodeine (DHC). A self-report questionnaire was administered every six months until the earliest recruit had been followed for over three years. Outcome data were analysed on an intent-to-treat basis, and comparisons were adjusted for significant baseline variables. Although some treatment crossover was observed from the experimental group to the control, both compliance and retention in treatment were high and approximately similar between groups at follow-up. There were no observed differences in outcomes between groups at 36 months' follow-up.

Non-Pharmacological Criminal Justice Programmes

Three studies examined the effectiveness of non-pharmacological criminal justice system-based treatment programmes for addicts (Martin et al., 2003; Naeem et al., 2007) and abusers (McSweeney, 2009) in the UK.

Martin et al. (2003) evaluated the abstinence- and psychology-based Rehabilitation of Addicted Prisoners Trust programme in male prisons. The programme comprises therapeutic

activity based on abstinence treatment and psychologically-informed case management. Official reconviction data were collected for 274 programme completers at one year post-discharge from the programme, and for 137 completers at two years post-discharge. These data were compared to a group of 931 offenders from other prisons who displayed a similar likelihood of reoffending, and who had not received treatment. The results showed a statistically significant reduction in criminal activity in the treatment group, compared to the control, at both follow-up periods.

McSweeney (2009) investigated the impact of the Drug Interventions Programme (DIP) in one English region. The DIP tests arrestees for drug use and requires subjects failing the test to enrol in a further assessment at a treatment centre, and this measure frequently entails further testing at regular intervals and required attendance at a drug treatment programme, often as a diversion from a more formal criminal justice measure. The author examined the offending data, compiled from official records, of an experimental group of 252 offenders who had been administered a DIP as compared with 263 offenders who had not. Although subjects in the experimental group reported more previous convictions than subjects in the control group, they nonetheless reported fewer convictions at 12 months' follow-up than subjects in the control group. The author further noted that half of the participants in the DIP group reported a large reduction in crime, and the remainder showed no change or a slight increase after follow-up.

Naeem et al. (2007) assessed the effectiveness of Drug Treatment and Testing Orders (DTTOs) compared to standard care. 35 subjects in the experimental DTTO group underwent frequent drug testing and 'treatment' – taking various forms – for a set period of between six months and three years in the community. 38 subjects in the control group received care as usual. Although the experimental group reported a more criminal history, the groups were approximately similar. The samples were drawn from two different sites (Portsmouth and

Southampton), and the authors note that the Southampton sample manifested a more problematic profile of drug use and criminal activity than their counterparts in Portsmouth. Questionnaire data measured at outcome captured crime at baseline and at 12 months' follow-up after admission into the programme. Nearly one-third of participants in each experimental group were lost at follow-up. After controlling for various significant baseline variable scores, the authors found no statistically significant differences in crime between groups at follow-up.

Meta-Analysis

Because of the deficits of the vote counting of significances (Wilson, 2001), we conducted a meta-analysis of our study sample in order to calculate a summary measure of the effectiveness of treatment programmes to reduce reoffending among substance abusing populations.

Coding of Effect Sizes

Primary study outcome results were coded using two effect size measures, depending on whether the reported data were provided in a continuous-level or a dichotomous-level format. For the former, we used the standardised mean difference, or Cohen's d , and calculated a Hedges' g adjustment to correct for the upward bias found in the computation of d when using small samples (Hedges, 1981). For the latter, we used the Odds Ratio and converted into a standardised mean difference so that all studies could be synthesised into the same meta-analysis. The resultant Hedges' g effect size obeys the same properties as Cohen's d , in that treatment outcomes favouring the experimental condition are represented by a score exceeding 0, and outcomes favouring the control fall below 0. Where necessary values were

omitted in study reports, we calculated the effect size based on calculations using available p , t , or F values; otherwise, every effort was taken to contact the author.

Conversion between continuous- and dichotomous-level constructs is endorsed by Lipsey and Wilson (2001:57) when the dichotomous variables reflect an inherently continuous variable, but have been artificially dichotomised by primary study authors. Although outcome constructs included in the meta-analysis vary considerably (ranging from the number of days engaged in illegal activity in the last month to whether the participants had been admitted into custody during the follow-up period), nonetheless the substantive construct that is measured is sufficiently homogeneous between studies to allay concerns of comparing ‘apples with oranges’ (Eysenck, 1995). We also coded whether the study was conducted in the UK, and whether the primary method of treatment was pharmacological substitution or otherwise.

Results

The summary effect size yielded by the meta-analysis was $g = 0.355$, $p < .001$ (see Figure 2). We used a random-effects analytic approach, as the Q test rejected the null hypothesis of homogeneity of variance $Q_{\text{Total}}(13) = 36.439$, $p < .001$. Although Borenstein et al. (2009:84-5) caution against choosing between fixed- and random-effects models purely on the basis of the Q statistic (largely due to the rarity with which the test attains adequate power), the use of a random-effects model is justified by the fact that the included effect sizes can most accurately be described as having been drawn from a *distribution* of effect sizes, with the dispersion in effects attributable to more than mere sampling error. The disadvantage of this approach is that it does not assign an inverse variance weight to those studies with low precision, as in a fixed-effects model. However, the substantive diversity among the studies in the sample suggests the need for a random-effects analysis. This debate notwithstanding,

the resultant effect size in the fixed-effects model yielded the only slightly more moderate summary effect size of $g = 0.277$, $p < .001$.

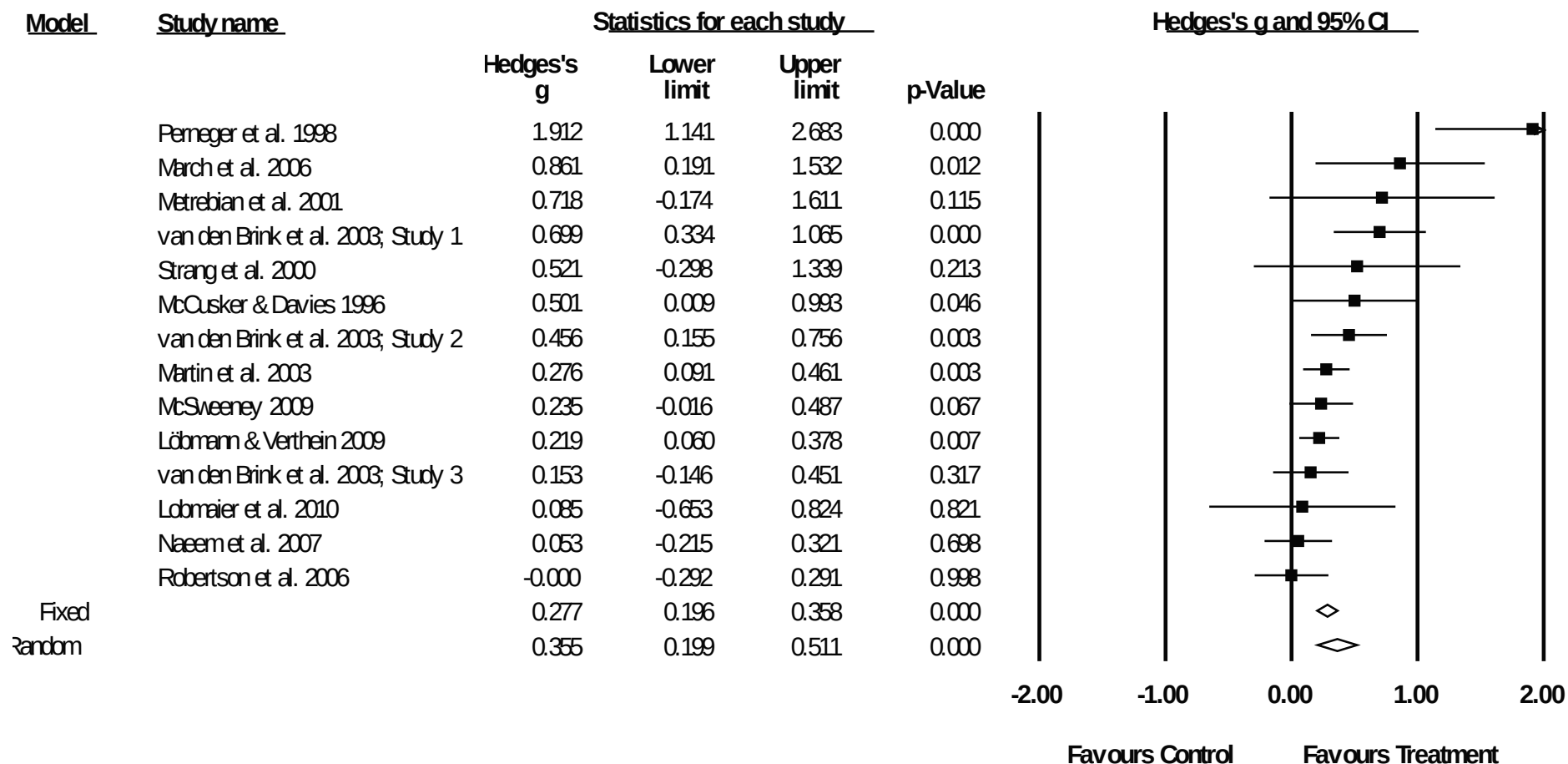
Large effects were observed in 6 studies: Perneger et al. (1998), March et al. (2006), Metrebian et al. (2001), van den Brink et al. (2003) Study 1, Strang et al. (2000), and McCusker and Davies (1996), with g values of 1.912, .861, .718, .699, .521, and .501, respectively. Effects reported in Perneger et al. (1998), March et al. (2006), van den Brink et al. (2003) Study 1, and McCusker and Davies (1996) were statistically significant ($p < .05$), whereas those found in Metrebian et al. (2001) and Strang et al. (2000) were not.

Van den Brink et al. (2003) Study 2, Martin et al. (2003), McSweeney (2009), and Löbmann and Verthein (2009) reported medium-sized effects ($g = .456$, .276, .235, and .219, respectively). Although the effects observed in van den Brink et al. (2003) Study 2, Martin et al. (2003), and Löbmann and Verthein (2009) were all statistically significant at $p < .05$, in McSweeney (2009) it was not. Van den Brink et al. (2003) Study 3, Lobmaier et al. (2010), Naeem et al. (2007), and Robertson et al. (2006) reported small effect sizes ($g = .153$, .085, .053, -.000³, respectively), and these were all statistically non-significant at $p > .05$.

The BESD (Rosenthal 1991) converts an effect size into a more practically meaningful measure of difference in recidivism rates between experimental groups (assuming an equal number of cases in each group, and a base rate of recidivism of 50%). It is calculated using the formula $BESD = 50\% \pm (\text{Pearson's } r / 2)$, for each experimental group. Hedges' $g = 0.355$ translates to a Pearson's r of 0.19 in the random-effects model, thus indicating a recidivism rate of roughly 40.5% in the experimental group and 58.5% in the control group at follow-up.

³ The Hedges' g observed in Robertson et al. (2006) was in fact -0.0004, and is therefore more sensibly interpreted as a non-effect, rather than a negative effect.

Figure 2: Meta-analytic results of main treatment effects of programmes to reduce reoffending



Sensitivity Analysis

Because random-effects models may overemphasise studies with small samples at the expense of studies with larger samples we conducted a sensitivity analysis to examine the variability of effects under different conditions. The removal of any single study would not have altered the random-effects summary effect size substantially, as the only substantially different g value if any single study had been removed would have been 0.280 in a random-effects model and 0.258 in a fixed-effects model, if Perneger et al. (1998) had been excluded. Even so, the summary effect would have remained statistically significant at $p < .001$. Therefore, although the summary effect is only approximate, we can be reasonably confident that the direction and relative magnitude remains fairly robust. The removal of that study from the model changed the results of the homogeneity test to $Q_{\text{Total}}(12) = 18.973, p = .09$. Finally, we conducted a Fail-Safe N test to investigate how many studies with a non-significant effect would have been required to render the significance level of the observed summary effect size greater than $p = .05$: the result was 191 studies, which underlines the robustness of our overall effect.

Moderator Analysis

Because of the small number of studies, the potential for moderator analysis is very limited. We examined the impact of the use of different measures of offending by comparing effect sizes calculated from self-report data ($k = 12$) to those calculated from official data ($k = 2$). The effect sizes were different between the two categories ($g = .257, p = .20$ for official data; $g = .423, p < .001$ for self-report data in the random-effects model). However, due to the presence of only two studies with official data, the difference in sizes of effect was not statistically significant ($Q_{\text{Between}}(1) = 0.561, p = .45$).

We also examined the impact of the length of follow-up. This revealed a tendency of larger effects in studies with shorter versus longer follow-up ($g = 0.733$, $p < .001$ and $g = 0.265$, $p < .001$, respectively; $Q_{\text{Between}}(1) = 5.978$, $p = .01$).

Primarily pharmacological substitution treatments ($k = 11$) reported a significantly large mean effect ($g = 0.437$, $p < .001$), and treatments that did not use primarily pharmacological substitution ($k = 3$) reported a lower, albeit non-significant, mean effect ($g = 0.193$, $p = 0.204$). This difference was not significant ($Q_{\text{Between}}(1) = 1.791$, $p = .181$). Studies that were conducted outside the United Kingdom ($k = 7$) reported a larger mean effect ($g = 0.488$, $p < .001$) than studies conducted within the UK ($k = 7$; $g = 0.240$, $p = .037$), although this difference was not statistically significant ($Q_{\text{Between}}(1) = 2.251$, $p = .134$). All three studies that tested primarily non-pharmacological programme types were conducted within the UK.

Discussion

A considerable literature has emerged within the field of drug treatment programme evaluation to demonstrate that such treatment ‘works’. The results of our analyses reiterate claims in support of the effectiveness of drug treatment programmes to reduce criminal behaviour among drug abusers. The added import of our review is the rehearsal of these claims within a specifically European context, with the use of a methodologically rigorous primary study sample.

Summary Effects and Previous Reviews

With the exception of the results reported in Hartnoll et al. (1980; see further below), our analyses revealed no negative effects of drug treatment. The generally positive results found in our meta-analytic review are consistent with some of the findings from previous

analyses, but are larger in magnitude and relatively robust. The roughly concordant effect size of $g = 0.13$ for crime-reductive outcomes reported by both Prendergast et al. (2002) and Holloway et al. (2008)⁴ did not reach statistical significance of $p = .05$ under random-effects models. Similarly, moderate and even some negative results were observed in reviews by Egli et al. (2009) and Mitchell et al. (2006), depending on which treatment modality was examined. The summary effect size of $g = 0.355$ in our review is not only more promising, but also statistically significant at $p < .001$. In practical terms, the significant monetary and human benefit of a 19 percentage point difference in recidivism between groups, assuming a 50% baseline according to the BESD, can be considerable. We are confident that these findings are fairly robust, considering the results of our sensitivity analysis. The size of the effect was not significantly related to type of outcome data. Furthermore, the results of the Fail-Safe N test indicate that 191 missing studies would have had to be included in order to render the summary effect size statistically non-significant. It is highly improbable that we missed so many studies in our exhaustive search of the literature.

We observed a positive and significant mean effect of interventions that were primarily pharmacological substitution based treatments. In comparison, although the mean effect for other forms of treatment was positive, this result was not significant. However, we attribute this in part to the small number of studies in this latter category. We also believe that this may, in part, explain the more moderate mean effect observed among studies conducted within the United Kingdom compared to those conducted elsewhere.

The amplified overall effectiveness of treatment reported in our analysis invites consideration of the differences between our data and those used in previous reviews. One

⁴ Holloway et al. (2008) provide a summary fixed-effects Odds Ratio in their original report; using their data, we re-computed a summary effect size and found that for the random-effects model, the resultant summary effect ($d = 0.133$) failed to reject the null hypothesis at $p = .05$.

plausible explanation is the higher threshold of methodological rigour applied in our eligibility criteria. The threats to internal validity in single group, pre-post designs have long been noted (e.g., Cook and Campbell, 1975), and were eliminated in our sample by the inclusion of exclusively two-group experimental and sound quasi-experimental designs. Although conspicuously little is known about the ‘life course’ of criminal activity among drug addicts (Hammersley, 2011), we do know that drug use tends to peak in the year before offenders enter treatment (Gossop et al., 2006). As a result, any evaluation design that measures outcomes post-treatment must take into account the peculiarity of ‘regression to the mean’, lest false confidence be attributed to observed crime reductions. The fact that *almost all* the experimental groups in our study sample observed a crime reduction over time should stand as a cautionary tale in this regard, and should act as an admonition to reserve claims of crime-reductive treatment benefit to comparison group designs.

A further consideration that is difficult to manage in such research is the confounding factor of *motivation* to undergo and complete treatment. This is problematic for a number of reasons, not least of which is noted by Holloway et al. (2008:33), that “it is possible that quasi-experimental designs are prone to selection bias, whereby the most promising clients are allocated to the experimental treatment.” The relationship between an offender’s willingness to complete treatment and the likelihood of positive results upon discharge is well established (e.g., see Lösel (1995) for a discussion of this topic concerning rehabilitation research in general, and Zarkin et al. (2002) for drug treatment research in particular). Unfortunately, controlling for such a factor is extremely difficult in drug abuse research, largely as a result of the evaluation designs commonly employed in that field (Stevens, 2011). Typically, drug treatment evaluations consist of either a comparison of treatment completers against drop-outs (e.g., see Egg, 1992; Fernández-Montalvo et al., 2008; Pettersson et al., 1986), or treatment participants sampled from different populations. An example for the latter

is the multi-national QCT Europe study, in which participants who were administered treatment as part of a ‘quasi-compulsory’ court order were compared to treatment participants who chose to undergo treatment voluntarily (Uchtenhagen et al., 2006). If the elimination of motivational bias is the goal, then the use of “control groups especially from voluntary treatment are mostly not equivalent.” (Uchtenhagen, 2002:80)

Thus, the moderately different effects observed in previous reviews may in part be attributable to the inclusion of exactly such a set of evaluation designs, in which results are artificially biased *in favour* of treatment in the former case, and *against* it in the latter. It is for this reason that we applied a high threshold of eligibility, in order to ensure that the control group constituted a set of participants with a reasonably equivalent level of motivation to pursue treatment and, concomitantly, an approximately equivalent propensity to reoffend.

Primary Study Biases

Although we included only evaluations with a sound design, our study samples may have contained participants who were selectively biased in certain respects that introduced non-equivalence between treatment and control groups. For example, the participants in all of the primary study samples were adults, and with the exception of McSweeney (2009) and Naeem et al. (2007), they were opiate addicts who had either failed, or had demonstrated a propensity to fail, in conventional treatments. The experimental treatment applied in each of the primary studies was intended as in some way innovative and exceptional from convention. If the sample of participants in each of the studies had already repudiated such courses of treatment, as was indeed the case, then it is plausible that this could conceivably introduce an *a priori* motivational bias against treatment altogether. We saw this as an acceptable bias because there was no strong evidence that it would operate in favour of treatment.

Perhaps the potential violation of the stable unit treatment value assumption, or SUTVA, is a more salient concern in the present analysis. This concept reflects the statistical prerequisite that participant outcomes are not affected by their assignment to any particular experimental group (Berk and Freedman 2003). Despite our efforts to include studies that approximate differing levels of motivation, we cannot control for a given addict's predispositions towards a particular drug of choice. In both Metrebian et al. (2001) and McCusker and Davies (1996), subjects were assigned to groups based on their habitual drug of choice. The authors noted that as a result, it is difficult to disentangle the differences in treatment outcome from being attributable either to the treatment itself, or to the differences in characteristics between groups. Moreover, the differences in outcome may simply be a function of participants being given the option of receiving their drug of choice (Metrebian et al., 2001:274).

Randomisation is also not always the best solution. In both of the randomised controlled trials where participants were invited to express a desire for a particular drug, the authors recognised a patient-treatment interaction: subjects who were randomly allocated to their preferred option displayed an improved outcome, and they displayed a poorer outcome if assigned to the opposite (Strang et al., 2000; Hartnoll et al., 1980). As mentioned, we therefore excluded Hartnoll et al., (1980) from the meta-analysis. Doing so changed the summary effect size in the random-effects model only slightly to Hedges' $g = 0.313$, $p < .001$.

As well as a potential bias in primary study samples, there may have also been a bias embedded within each of the treatment comparisons. For example, the pharmacological properties of the different treatments may have dictated different behavioural or physiological responses between experimental groups. Despite attempts to equate the treatment dosage between methadone and prescribed heroin in March et al. (2006), Metrebian et al. (2001), and

Löbmann and Verthein (2009), the effects of prescribed heroin decay much more rapidly than methadone. This necessitated more frequent visits to the experimental facilities by participants in both studies. The authors noted that this may have strengthened a personal relationship with their caseworker and resulted in more care. Increased attendance could also have influenced motivation more than in the control groups.

Lobmaier et al. (2010) also remarked on the potential bias in favour of the treatment group in their study, but they employed the contrary case of *less* frequent contact with experimental group participants. Naltrexone implants operate by releasing their dose over a protracted time period, requiring very little interaction with clinical facilities. In this case, the authors note that it is the greater convenience and independence afforded to participants that may have contributed to the positive effects observed in this study.

The consequence is that there may be confounding factors within the treatment group comparisons that do not lend themselves to being partialled out in the final analysis. Our results do not accommodate parsing the differential effects of one treatment modality over another, nor do they allow for an identification of the particular component of success within a treatment. Multiple studies within our sample applied a combination of treatments in each experimental group: for example, Perneger et al. (1998) and McCusker and Davies (1996) offered both a pharmacological as well as a social service treatment within the experimental condition. As a result, it is difficult to determine which component was the more operative in reducing criminal activity. Similarly, the participants in Martin et al. (2003) received abstinence treatment and psychological counselling; again, the principally operative component is largely unidentifiable.

Factors Influencing Effectiveness

It was repeatedly found that new treatments that are applied in demonstration studies with dedicated resources and facilities report significant crime reductions, but that these effects disappear or become weaker in routine practice (Lipsey & Wilson, 1998; Lösel & Schmucker, 2005). Accordingly, substantial crime reductions were reported in both Strang et al. (2000) and in van den Brink et al. (2003), where treatments in the experimental condition were delivered in newly constructed, dedicated clinics, and outcomes were measured at six and twelve months' follow-up respectively. It would be prudent to conduct follow-up evaluations of treatments applied at the same facilities later to investigate whether such positive effects were maintained. Our tendency of a lower effect in studies with longer follow-up periods suggest that this is a relevant issue.

Attrition rates may also play an important role in influencing the observed effectiveness of a drug treatment programme for two reasons: Firstly, an elevated drop-out rate in one group suggests an unequal proportion of motivated participants between groups, and thus a higher effect size might be expected in the group with a smaller drop-out rate. The second reason is that some studies (e.g., Lobmaier et al., 2010) applied the 'last observation carried forward' technique to impute follow-up data from baseline values for subjects who have dropped out of treatment and are no longer reachable. As offending tends to peak in the year before entry into treatment, differential success may be attributable in part to the artificial over-representation of high conviction rates in whichever group reported the higher drop-out rate. Differential attrition rates are reported in many studies in our sample (see Table 1): Löbmann and Verthein (2009), Lobmaier et al. (2010), Metrebian et al. (2001), and van den Brink et al. (2003).

Broader Implications

Meta-analytic results are ultimately “based on ‘found’ data and are thus limited to the characteristics of studies that [are sampled]. What can be known through meta-analysis largely depends upon the state of the literature regarding the research question.” (Prendergast et al., 2002:66). Although it is improbable that our exhaustive literature search omitted a significant number of European evaluations, the sample consists of mainly evaluations of primarily pharmacological substitution programmes for opiate-dependent offenders. In addition, studies conducted in the United Kingdom are over-represented. This sets a clear limit to generalisation. Substitution programmes make up only one part – albeit a significant one – of the drug treatment repertoire. Estimates from the UK report that methadone treatment accounted for roughly 50% of treatments delivered to substance mis-users in 2010 (Roxburgh et al., 2010). Our survey in 27 EU countries suggests that other programmes are even more common in other parts of Europe (Hamilton et al., 2011). Had we relaxed our eligibility constraints to include a greater diversity of evidence, it is likely that we would have discovered a broader range of treatment programme evaluations. However, then we could not control for the various threats to validity that commonly plague research in this field (Stevens 2011).

In conclusion, we need more, and better, evaluations of other drug treatment programmes, such as therapeutic communities and psychological programmes. More good quality evaluations are also needed on addictions to other substances than opiates (e.g., crack or marijuana⁵), for young offenders, for female substance addicts, and for offenders who have

⁵ For example, Bennett et al. (2008) found the strongest relationship between drug use and crime among crack users; moreover, a considerable proportion of drug abusers entering treatment are found to be addicted not just to opium, but to many other drugs as well (58% in the UK, for example; Roxburgh et al., 2010:19). Evidence for effective solutions to deal with this problem is clearly an exigent need.

not demonstrated any lack of receptivity to drug treatment. And finally it is urgent to reduce the heavy imbalance of evaluation activity across Europe. Whereas most research comes from the UK and Spain, Norway, the Netherlands, Switzerland, and Germany are at least represented, we could not find one sound study in the majority of European countries. For these reasons there is clearly an imminent need to commence a programme of evaluation research of treatment programmes for a much broader population of drug abusing offenders in whole Europe.

APPENDIX Bi: Search terms

<i>Target Population</i>	Offender* OR Inmate* OR Prison* OR Probation OR Correction*
	AND
<i>Intervention</i>	Drug* OR Substance OR Addict* AND Therapeutic communit* OR Methadone* OR Cognitive behav* OR Drug couns*
	AND
<i>Study</i>	Evaluat* OR Outcome* OR Effect* OR Assess* OR Treatment* OR Interven* OR Rehab*
	AND
<i>Outcome</i>	Prevent* OR Reduc* OR Arrest OR Recidiv* OR Reoffen*

A Systematic Review of the Effectiveness of Domestic Violence Perpetrator Programmes to Reduce Repeat Abuse in Europe*

Thomas D. , Johann A. '

David K. and Friedrich .²

¹ Institute of Criminology

University of Cambridge, UK

² Institute of Psychology

University of Erlangen-Nuremberg, Germany

* This study was carried out within the European cooperation project “Strengthening Transnational Approaches to Reducing Reoffending”. The project was funded by the European Union. We thank our partners from the UK Ministry of Justice, London Probation Trust, European Organization for Probation, and the Ministries of Justice of Bulgaria, France and Hungary for their cooperation. We graciously acknowledge the assistance provided by the many experts and practitioners who responded to our requests for information and data, and by the Daphne II Work with Perpetrators Survey who furnished us with data that assisted in our literature search.

Abstract

Systematic reviews of North American evaluations of the effectiveness of domestic violence perpetrator programmes in reducing episodes of further abuse are partially positive, but overall inconclusive and not simply transferrable to other cultures. Therefore, we present the results of a systematic review of European evidence. After searching through 10,446 titles, we discovered only eleven studies that evaluated the effectiveness of a perpetrator programme in some systematic manner. The review systematically documents and analyzes the characteristics and findings of the various studies. The studies applied treatment to a total of 1,413 domestic violence perpetrators, and the sample sizes ranged from 9 to 322. The mean participant age was 37.74. Follow-up periods ranged from immediately upon treatment completion to 12 months after the programme concluded. Typically these studies used single group pre-post measurements of a range of outcomes. Only one study contained a comparison group, but this was not equivalent to the treatment group. Although the evaluations showed various positive effects after treatment, with violent behaviour reductions ranging from 5 to 30 percentage points, methodological problems do not allow us to attribute these findings to the programmes. Overall, the methodological quality of evaluations of domestic violence perpetrator programmes in Europe is insufficient to derive firm conclusions and estimate an effect size. Accordingly, one cannot claim that one programmatic approach is superior to another. Evaluation of domestic violence perpetrator treatment must be improved and programmes should also become more tailored to the characteristics of the participants.

Introduction

Recent estimates of the lifetime prevalence of domestic violence among European women report that roughly 12% to 16% of women throughout Europe have, at some point in their lives, experienced an episode of physical abuse at the hands of their partner since reaching the age of 16 (Council of Europe, 2008). These findings approximate the estimates of lifetime prevalence found in other countries, including North America, Canada, and New Zealand, among others (Archer, 2006). Given that domestic violence victimisation has been associated with an increased likelihood of substance abuse, depression, post-traumatic stress, suicidal ideation, injury, and death (Alhabib et al., 2009; Campbell, 2002; Golding, 1999), a panoply of measures have emerged that attempt to address the risk of repeat victimisation in abusive relationships.

The first legislative provisions to address domestic violence were passed in the United Kingdom and the United States during the 1970s, largely in response to successful lobbying efforts on the part of victims' rights and advocacy groups (United Nations, 2008). As longitudinal estimates of the rate of repeat violence in domestically violent relationships were found to be in the region of around 40% to 80% (Garner et al., 1995; Shepard, 1992), dedicated programmes dealing with the perpetrators' violent behaviours were seen as a promising approach to reducing the incidence of victimisation. Therefore, in addition to victim-oriented protection and punishment programmes that seek to re-orient a perpetrator's attitudes, beliefs, and behaviours have become popular. Gondolf (2002) enumerates the prevailing paradigmatic approaches to domestic violence perpetrator programmes as cognitive-behavioural, psycho-dynamic, and pro-feminist. Cognitive-behavioural programmes attribute violence to learned behaviours that perform an expressive, instrumental function; as a consequence, programmes adhering to this model emphasise that desistance must be *learned* through a process of cognitive restructuring. Psycho-dynamic approaches

emphasise the personality and emotional disposition of the perpetrator as being central to desistance, by allowing him to recognise and reconcile latent feelings of emasculation that precipitate abusive impulses. Pro-feminist approaches view violence as originating from patriarchal values about the role of women, and typically aim to re-orient the way men seek to exert power and control over their partner. More specific treatment approaches to dealing with domestic violence perpetrators also exist, such as anger management and integrated substance abuse/domestic violence treatment programmes. However, evaluations of these latter types of treatment are rare, and those that do exist contain many methodological flaws (Barnish, 2004).

The distinction between the above-mentioned treatment categories is often more clear in theory than in practice. In actuality, many programmes overlap and coalesce around unitary principles of how to reduce repeat abusive behaviour (Scourfield and Dobash, 1999; see also Hamilton et al., 2011). For example, the most prevalent programme in the world is the Duluth model (Rothman et al., 2003), which is an integrated, multi-agency approach that derives from the psycho-dynamic paradigm and incorporates elements from the cognitive-behavioural and pro-feminist models.

Evaluations of the effectiveness of Duluth and other programmes have been inconclusive; in part this is attributable to a continuing contention surrounding the correct definition of domestic violence. Where some advocate using only violence as an indicator of abuse, others maintain that reductions in various emotionally and verbally assaultative behaviours are more suitable. Some promote the use of the victim's psychological well-being or feelings of safety as being a more valid indicator of domestic violence. In addition to outcome criteria, the evaluation design is a problematic issue. Davis and Taylor (1999) refer to three generations of research in the evaluation of domestic violence perpetrator programme effectiveness: the first generation comprised evaluations of low methodological quality,

typically employing research designs that eschewed comparison groups, and outcomes were often measured either after treatment had ended or at intake and upon completion of the programme. The second generation employed non-equivalent comparison group designs, and the third generation of evaluation designs employed random assignment of participants into various treatment/non-treatment groups. Davis and Taylor claim that the advent of the third generation is still yet to take its full form. Accordingly, very few evaluations exist that qualify in the later generations of domestic violence research; of those that do, all originate from North America.

These methodological issues notwithstanding, various systematic reviews and meta-analyses have been carried out in North America (e.g., Feder et al., 2008; MacKenzie, 2006; Babcock et al., 2004; Davis and Taylor, 1999; Hamberger and Hastings, 1993). For example, the most recent review of Feder et al. (2008) contained 10 randomised controlled trials on court mandated offender programmes. It found that on average official reoffending (e.g. arrest or reconviction) declined by 13%; however, there was no significant effect in victim reported outcome. Overall, the North-American reviews agree insofar that they could not yet ascertain the precise effectiveness of domestic violence perpetrator programmes.

Much less systematic information in this field has been gathered in Europe, although preliminary efforts have been conducted by the Daphne II's Work with Perpetrators Survey. However, one cannot simply generalise to North-American contexts because of differences in definitions, conceptualisations, measurements, legal frameworks, and responses to domestic violence that even vary across the continent (Gracia and Herrero, 2006). The likelihood that male partners will refer themselves, or be reported to, perpetrator programmes in response to an abusive incident is commensurate with internationally differing levels of permissiveness surrounding partner abuse. Consequently, the population of partners enrolled in treatment in different countries may vary along important dimensions. All this may have important

implications for the success of a programme. Therefore, we endeavoured to conduct a systematic review of the effectiveness of European domestic violence perpetrator programmes to reduce repeat victimisation.

Method

Inclusion criteria

We anticipated that evaluations of perpetrator programmes would be less accessible than evaluations of other types of rehabilitation programme (Dalton, 2007); therefore, we sought to maximise the studies captured in our literature search by using fairly relaxed eligibility criteria, and by employing a comprehensive search protocol.

1. We limited our search to evaluations conducted in Europe.
2. Both published and unpublished formats were acceptable for inclusion in our sample.
3. The target sample had to comprise domestic violence perpetrators, defined as either offenders who had been convicted of a domestic violence offence, or partners who had commenced a course of treatment to deal with their self-reported partner-violent behaviours.
4. The evaluation had to examine the effectiveness of a treatment programme that was designed to alter the attitudes and/or behaviours of domestically violent partners.
5. We deemed either attitudinal or behavioural outcome measures as acceptable.
6. At minimum, the study had to measure outcomes before the commencement of treatment, and at the conclusion of treatment.

Search results

In order to locate unpublished and published studies, we searched online computerised databases and specialist journal archives¹, as well as meta-analytic and systematic review

¹ A full list of databases used in our search, as well as the search terms employed, can be found in Appendices Ci and Cii.

publications dealing with domestic violence perpetrator treatment programmes (e.g., Babcock et al., 2004; Davis and Taylor, 1999; Feder et al., 2008; Rothman et al., 2003;). We also contacted academics and experts in an effort to locate studies that might not have been accessed by the more conventional strategies. We also conducted a survey of all domestic violence perpetrator programmes throughout the European Union during 2010 to 2011, in which we asked respondents to furnish us with any evaluations of their practice. Moreover, we consulted the database of the Daphne II Work with Perpetrators Survey, which had compiled a network of domestic violence perpetrator programmes in each of the 27 EU countries during 2007 to 2008. Although the database did not include information concerning evaluations of practice, we individually contacted each respondent to that survey and asked them to provide us with any available evaluations of their practice.

Our bibliographic database search yielded a total of 10,446 titles, which, upon deletion of duplicates, yielded 8,325 discrete documents. The titles and abstracts were then screened in detail according to method, location of study, sample population, and outcome of interest, in order to arrive at seven studies (for further detail regarding the eligibility criteria, see above). These were supplemented by seven further studies which were retrieved as a result of our questionnaire survey, and a further five studies which were added through consultation with respondents to the Daphne II Work with Perpetrators Survey. This resulted in nineteen studies, which we retrieved in full. Eight studies were excluded on the basis of a lack of outcome measurements ($k = 4$), or because measurements were taken at only one point in time ($k = 4$). Our final study sample consisted of eleven evaluations of domestic violence perpetrator programmes (see Figure 1).

Figure 1: *Flowchart of literature search*

Results

The eleven studies that constituted our final study sample originated from six European countries: Cyprus ($k = 1$), Finland ($k = 1$), Germany ($k = 1$), Spain ($k = 4$), Sweden ($k = 1$), and the United Kingdom ($k = 3$). Six studies were published, and five were unpublished². Nine of the studies were written between 2000 and 2010; only two studies were written before this time, of which the oldest was published in 1997.

We have provided an in-depth narrative review below of the six studies that report outcomes related to offending or violent behaviours. Five studies in our sample used data that pertained only to attitudes and beliefs surrounding women and psychological variables related to impulsivity, self-esteem, anger, etc. Because these variables are at best only a proxy

² One study (Bowen 2004) was an unpublished doctoral dissertation. Although this work has appeared in published form in peer-reviewed journals, we referred to the original source document on the basis of its comprehensive explanation of the primary study project and its outcomes.

indicator of reoffending activity, we eschew in-depth description of these studies, and instead refer readers seeking elaboration to Table 1 for further details.

Dobash et al. (1999)

Dobash et al. (1999) compared the effectiveness of two court mandated domestic violence perpetrator programmes with traditional criminal justice based sanctions (e.g., fines, probation, and prison) in Scotland. Both programmes were court-mandated, as participants had all been found guilty of domestic violence and were fulfilling a condition of their probation by attending the programmes. Both programmes were predicated on the belief that violence in the home is a learned behaviour, resulting from issues of power and control between partners. The programmes are cognitive-behavioural, and emphasise educational rather than psychodynamic methods.

The experimental condition was composed of 51 men who participated in one of the two programmes and 47 of their women partners. The control condition was composed of 71 men who were adjudicated to other types of criminal justice sanction and 87 of their women partners. The authors administered in-depth interviews to the men and the women partners of the participants at the beginning of the programme (Time 1), and sent postal questionnaires at three and twelve months following the first measurement (Time 2 and 3). The authors received responses at Time 2 from 80% of men and 83% of women in the experimental condition, and from 72% of men and 77% of women in the control condition. At Time 3, the authors received responses from 57% of men and 60% of women in the experimental condition, and from 49% of men and 57% of women in the control condition.

There were very few differences between the two groups on key demographic, criminal, and attitudinal variables at Time 1 measurement. The only significant baseline differences between groups pertained to employment and marital status. During the follow-up

period, 7% of men in the experimental condition and 10% of men in the control condition appeared in arrest and prosecution records. Women's reports of subsequent violence based on the questionnaire data revealed that 30% and 33% of the men in the experimental condition used violence at Times 2 and 3, respectively. The corresponding figures provided for the men in the control condition are 61% and 69%, respectively, which is a statistically significant difference. This difference obtained when the authors compared the use of frequent violence between the two groups, as well.

The authors also observed reductions in the experimental condition of controlling and intimidating behaviours, both over time and compared to the control condition. Moreover, women partners of men in the experimental condition reported more positive and statistically significant improvements in quality of life measures such as feelings of happiness, contentment, and safety than women partners of men in the control condition.

Table 1: Summary of primary studies

Author(s)	Country	Participation	Theoretical Paradigm	Programme Structure	Sample Size and Drop-out Rate	Maryland Scale	Outcome Measure and Follow-up Period	Results
Dobash et al. (1999)	UK	Court mandated	Duluth Model; CBT	Two group programmes: 24 weeks and 27 weeks	51 men and 47 women in treatment group. 71 men and 87 women in comparison group. 47% male and 40% female dropout in treatment group, 51% male and 42% female dropout in comparison group.	3	Various psychometric and psychological assessment instruments, official crime reports, and women partner's self-reports. Outcomes measured at intake and three and twelve months after programme completion.	7% recidivism in treatment group, 10% in comparison group; reductions in violence corroborated by women partner's reports. Improvements in treatment group's men's controlling behaviours and women's well-being, compared to control group.
Bowen (2004)*	UK	Court mandated	Pro-feminist psycho-educational (Duluth model)	24 150-minute group sessions, and 5 150-minute follow-up sessions.	120 men, 32% dropped out.	2	Various psychometric and psychological assessment instruments, and official crime reports. Outcomes measured at intake and at 11 months' follow-up.	15% recidivism for treatment group and 33% for drop outs, equating to a small, marginally significant effect. Offenders achieved modest psychological improvements, although these were not related to offending behaviours.
Adva (2008)*	UK	Voluntary	Duluth Model; CBT	10 individual CBT sessions; 30 loosely defined group sessions.	115 men, 63% dropped out. Data collected from 12 women partners and 20 children of perpetrators.	2	Self-reported risk of re-abuse, psychological variables (e.g., self-esteem, locus of control), abusive incidents, measured throughout treatment over rolling period of 30 months.	Initial increase in self-reported abusive behaviour, then gradual decline. Women partner report decrease in abuse. Strong decrease in risk of re-abuse among treatment completers; corroborated by women's report. Significant psychological improvement among perpetrators and among women and children.
Socialstyrelsen (2010)	Sweden	Voluntary	Duluth Model; CBT	Varies from 3 individual to more than 20 individual/ group sessions.	188 men, 43% dropped out. 16 women partners, 25% dropped out.	2	Self-assessment instrument capturing various psychologically and physically violent behaviours, mental health, and substance abuse, captured at Time 1 (entry into programme) and at Time 2 (12 months thereafter). Men's and partner's views of treatment gathered at Time 2.	Continuation of various forms of violent behaviours, although some reductions observed in the sample. Partners reported reductions in violent behaviours. Improvement in perpetrators' mental health and substance abuse. Perpetrators and partners reported satisfaction with the programme.

* Denotes the study was unpublished

Table 1: Summary of primary studies (continued)

Author(s)	Country	Participation	Theoretical Paradigm	Programme Structure	Sample Size and Drop-out Rate	Maryland Scale	Outcome Measure and Follow-up Period	Results
Törmä and Tuokkola (2009)*	Finland	Voluntary and court mandated	Psycho-dynamic	Individual or group sessions. Treatment lasts at minimum two months.	80 men responded to questionnaire. Six women spouses of perpetrators.	2	Self-report questionnaire capturing recalled behavioural and psychological information from before and after treatment. Interviews with women partners.	Reported reductions of violent incidents, intimidating behaviours, and an increase in well-being. Spouses reported feeling safer as a result of men's participation in treatment programme.
Association for Family Violence Prevention (2009)*	Cyprus	Not specified	Psycho-educational	20 two-hour seminars over 12 weeks.	9 men	2	Questionnaire measuring psychological variables at intake and upon programme completion.	Modest psychological improvements in anger, well-being, anxiety, and self-esteem, and modest reductions in physical assault towards partners.
Hagemann-White et al. (2004)	Germany	Both voluntary and court mandated	CBT	Roughly 20 group sessions	322 men, 57% dropped out.	2	Self-report questionnaire capturing recalled behavioural and psychological information from before and after treatment.	Improvements in gender-stereotypical attitudes, accepting responsibility, and admitting guilt.
Echeburúa et al. (2009)	Spain	Voluntary	CBT	20 weekly two-hour group sessions.	148 men, 32% dropped out.	2	A battery of psychological assessment instruments, administered at intake and immediately after completion of the programme.	Promising change on various psychological variables related to impulsivity, anger, and self-esteem.
Echeburúa et al. (2006)	Spain	Voluntary	CBT	20 weekly two-hour group sessions.	52 men, 8% dropped out.	2	A battery of psychological assessment instruments, administered at intake and immediately after completion of the programme.	Promising change in cognitive distortions, hostile attitudes, and uncontrolled anger.
Echeburúa and Fernández-Montalvo (1997)	Spain	Voluntary	CBT	15 weekly one-hour individual sessions	31 men, 48% dropped out.	2	A battery of psychological and psychometric assessment instruments, administered at intake, end of treatment, and at 1 and 3 months' follow-up.	Promising change in various psychological variables, e.g., anxiety, self-esteem, depression, anger.
Echauri Tijeras (2010)*	Spain	Voluntary and remand	CBT	20 weekly two-hour group sessions.	170 outpatient men, 53% dropped out; 80 men in prison, 73% dropped out.	2	A battery of psychological assessment instruments, administered at intake, end of treatment, and at 1, 3, 6, and 12 months' follow-up.	Promising change in various psychological variables, e.g., anxiety, self-esteem, depression, anger.

* Denotes the study was unpublished

Bowen (2004)

Bowen (2004) is the only study to date to have measured both psychological change and reoffending (i.e. police records) as outcomes in the assessment of effectiveness for domestic violence offender programmes in Europe. The author investigated the effectiveness of a programme delivered by the West Midlands Probation Area in the UK, administered to 120 participants who had been mandated to attend the programme as a result of a court order.

The programme consisted of 24 two-and-a-half-hour group sessions held once or twice a week, and five follow-up sessions lasting two-and-a-half hours each were held once a month. Attendance at fewer than 21 of the 24 core sessions constituted a drop-out. The modules adopted a psycho-educational, pro-feminist approach and concerned men's issues of power and control over women.

Outcomes were measured by official police records as well as by means of a number of psychological batteries administered at 11 months' follow-up, which captured data about violent behaviours and attitudes surrounding anger, violence, and dependency. These psychological variables were statistically adjusted to compensate for social desirability bias.

68% ($n = 82$) of offenders completed the programme. The author compared results for the completers to the results for the dropouts. In the eleven month follow-up period, 15% of completers were alleged to have committed another domestic violence offence, compared to 33% of dropouts; this was a marginally statistically significant difference. Recidivism was not associated with actuarial risk, offender type, the therapeutic environment, programme attendance, and variations in programme implementation. Reoffending was significantly associated with pre-treatment criminal history. Although a small positive psychological change was observed in the offender sample, this change was not related to reoffending.

Adva (2008)

The authors investigated the effectiveness of a Duluth-based programme located in Devon, UK. The community-based treatment incorporated a multi-agency approach to working with domestic violence, and targeted perpetrators who had been referred by other domestic violence agencies or who had approached the service directly. The intervention comprised meetings with police domestic abuse units and social workers, and also with women's support workers as well as children's and young people's workers to address the full panoply of problems in the family. The treatment consisted of a minimum 42-week course, involving ten individual cognitive-behavioural therapy sessions, and 30 loosely-defined group sessions with the perpetrator. The modules adopted cognitive-behavioural and Duluth-based psycho-dynamic methods.

Of the 115 participants who began assessment, 23% ($n = 26$) of participants completed the course, and 14% ($n = 16$) were still on the programme at the conclusion of assessment. Continuous assessment was undertaken throughout the duration of participation in the treatment. The percentage of men self-reporting abusive behaviour increased sharply after the first month, and then decreased over the next eleven months until the percentage of self-reported abuse is somewhat lower than at the commencement of assessment. The risk level of abuse decreased for the majority of participants who completed treatment, and this change was less pronounced among participants who did not complete the programme. Furthermore, this was corroborated by the women partners' perceptions of the risk of abuse. Statistically significant improvements on a number of psychological variables were observed among 19 participants who completed treatment. Women partners of programme participants reported a decline in the number of abusive incidents over the course of the intervention, and an improvement in well-being and safety regardless of the man's progress through treatment.

The majority of the 20 children of perpetrators who were assessed towards the beginning and end of their support programme reported psychological, behavioural, and academic improvements.

Socialstyrelsen (2010)

The author investigated the effectiveness of eight voluntary programmes located throughout Sweden. The programmes were broadly similar in approach, and comprised individual and group sessions incorporating psycho-dynamic and cognitive-behavioural approaches. The length of treatment varied for each participant from three individual sessions to more than 20 individual and group sessions each week.

A questionnaire was administered at entry into the programme (Time 1) and at twelve months thereafter (Time 2) to 188 male participants and 16 female partners. Questionnaire items concerned physical and psychological violence, mental health, and substance abuse. At Time 2, data was gathered from 12 women and from 140 programme participants; however, the author used a 'last observation carried forward' analysis for the male participants who were not available at post-test measurement. At Time 2, 38% of participants had completed treatment, 43% had dropped out, and 19% were still in treatment. The assessment comprised dichotomous-level data capturing whether or not a violent behaviour had been used in the year preceding measurement.

Although there were statistically significant reductions in the use of various forms of violence at Time 2, the majority of men continued to use minor psychological violence, and a substantial proportion reported continued physically violent behaviours. Significant improvements were also observed in programme participants' mental health and substance abusive behaviours at Time 2, compared to Time 1. The majority of participants reported

satisfaction with the programme at Time 2. 92% ($n = 11$) of the women reached at Time 2 who were still in contact with their partners reported that violent behaviours had decreased.

Törmä and Tuokkola (2009)

Törmä and Tuokkola investigated the effectiveness of a treatment programme called Jussi-työ in Finland. The programme can either be voluntary or as a result of a court mandate, and involves one-on-one or group discussion therapy depending on individual consultation with the client at intake into the programme. The psycho-dynamic treatment is intended to last at minimum two months.

The authors distributed questionnaires to clients who had participated in the programme, although the selection process was not described in detail. It is not clear how much time had elapsed since participation in the programme, before questionnaires were distributed; probably this time varied among respondents. 61% ($n = 80$) of questionnaire recipients responded, providing data about violent behaviours, psychological change, and satisfaction with the programme, from their recollections of before and after their participation in the programme. The authors report reductions in self-reported 'violent incidents', 'intimidating behaviours', and an increase in 'well-being' across all respondents. Furthermore, many respondents reported that the programme had 'been useful' to them. The authors also interviewed six spouses of programme participants, all of whom reported 'feeling safer' as a result of the man's participation in the programme. The precise behaviours, attitudes, and beliefs captured within each of the constructs have not been specified.

Association for Prevention and Handling of Violence in the Family (APHVF; 2009)

The authors investigated the effectiveness of a community-based perpetrator programme delivered to domestically violent men in Cyprus. The programme comprised

twenty two-hour psycho-educational group sessions delivered over the course of twelve weeks. Nine participants completed a questionnaire at intake into the programme and immediately upon programme completion. The questionnaire captured the respondents' sense of anxiety, well-being, anger, and self-esteem, as well as instances of psychological and physical aggression towards the partner.

Modest improvements were observed in all of the psychological variables, and five of the six participants who had reported physical assault at intake, reported a reduction upon programme completion. Of the three cases in which respondents reported injuring their partner at intake, reductions were reported in two cases. The authors conducted no statistical analysis of the data.

Discussion

We had hoped to establish the effectiveness of domestic violence perpetrator programmes at reducing reoffending and further abuse. Given our prior lack of knowledge about the state of evaluation research in this field in Europe, we relaxed the constraints of our eligibility criteria in order to accommodate a variety of research designs. Unfortunately, the promising results that the authors of each of the primary studies attribute to the programmes are specious, as the methodological quality of the studies is generally weak.

The methodological problems that characterise the studies within the sample comprise the nature of the outcome variable, the drop-out rates of the samples involved, the follow-up periods used before outcome measurements are taken, and the evaluation research design. Moreover, these methodological problems are accompanied by programme issues, such as the documentation of the treatment components, the similar modalities of treatment programme, and the absence of programme tailoring to offender typologies.

Methodological Issues

Outcome Measure

As in other fields of programme evaluation, we have to ask *who* is the most appropriate provider of information about programme effectiveness, and *what* characterises a suitable measure of that effectiveness (Westmarland et al., 2010; although some progress has been made recently, see Stover 2005). There is still no consensus in the scientific literature about whether official police data, perpetrator self-reports, victim interviews, or programme deliverer testimonies provide the most reliable indicator of repeat violence, as each form entails unique disadvantages (Straus 1991). For this reason, it is generally advised that evaluations incorporate data from a variety of sources, in an attempt to ‘triangulate’ often inconsistent information (Gondolf, 2002; Rosenbaum, 1988).

Only one study in our sample (Dobash et al., 1999) gathered outcome data from official police report, offender questionnaire, and women partner’s self-reports. One study (Bowen 2004) corroborated offender self-report data with police records, and two studies (Socialstyrelsen, 2010; Törmä and Tuokkola, 2009) complemented offender self-report data with information gathered from questionnaires administered to the women partners. The remaining seven studies collected outcome data from the offenders’ self-reports alone, thus raising suspicion surrounding the available data.

With regard to the *content* of the outcome measure, data can vary from criminal justice measures such as arrest, over rates of physical assault or episodes of verbal abuse, to perceptions of chronic intimidation and a general sense of lack of safety and well-being. Only two of the studies in our sample collected official crime data from police records (Dobash et al., 1999; Bowen, 2004), and four studies collected data concerning self-reported violent behaviours such as slapping and beating. All eleven studies gathered data concerning psychological change in some form or another; however, the utility of such data in

determining future violence remains unclear. For example, Bowen's (2004) analysis of the relationship between offenders' responses to items capturing psychological change reveals that there is little concordance with reoffending behaviours.

Sample Generalisability

A further concern in the determination of programme effectiveness relates to the generalisability of the sample of treatment participants. Feder et al. (2008) observed in their North American meta-analysis of domestic violence perpetrator programme evaluations that studies using a general population that was representative of "typical" perpetrators observed a lower overall mean effect size. The authors attributed this finding to the observation that such programmes may "work for a selected (presumably more motivated) subset of offenders." (Feder et al., 2008:15) It is thus necessary to distinguish when the evaluations have isolated those participants who are most likely to manifest reductions in violent behaviours, whether through sample size, selection, or attrition.

The sample sizes in the primary studies of our European review ranged considerably, from 9 to 322 (APHVF, 2009; Hagemann-White et al., 2004, respectively). The participants were selected from a range of sources, such as voluntary referrals and court-mandated diversion orders. In any cases there are relatively high rates of attrition. In the prison subsample of Echaury Tijeras' (2010) evaluation, for example, the number of participants who completed the programme approximated a mere quarter of the original sample, and in the 10 remaining primary studies in our review, attrition rates were rarely below 30%. Problems of high drop-out rates have been prominent since the early stages of domestic violence research. The predictors of drop-out seem to be related to risk factors for the resumption of violent behaviours among perpetrators (Daly and Pelowski, 2000). Moreover, one must assume that the women victims lost at follow-up such as those in Dobash et al. (1999) and in Socialstyrelsen (2010) are more likely to be abused with greater frequency and

severity (Sullivan et al., 1996). Therefore, on both methodological and ethical grounds, low attrition rates are a key issue.

It is highly plausible that the remaining samples in each of the evaluations where attrition was high were composed of the participants with the highest level of motivation to change. Therefore, a selection bias would have exaggerated the results in favour of finding a reduction in abuse at follow-up. This phenomenon has been labelled “creaming”, and has been observed in systematic reviews from North America (Feder et al., 2008; Babcock et al., 2004; Davis and Taylor, 1999). Furthermore, the conflation of outcome data for both voluntary *and* court-mandated samples together (e.g., Törmä and Tuokkola, 2009; Hagemann-White et al., 2004; Echaury Tijeras, 2010) dilutes the effectiveness of the programme, as voluntary referrals are also likely to manifest a greater motivation to change.

Evaluation Design

The selection of an appropriate follow-up period at which to gather outcome data is an additional problem in our study sample. The available literature has alluded to a “honeymoon period” of either a cessation or subsidence of abusive behaviours in the duration and immediate aftermath of enrolment in a perpetrator programme (e.g., Rosenfeld, 1992; Rosenbaum, 1988). As a consequence, the collection of outcome data immediately upon programme completion risks inflating the probability of asserting the false-positive of an encouraging treatment effect. The majority ($n = 6$) of studies in our sample collected data immediately upon programme completion; of the five remaining studies, none collected data more than twelve months after the programme had concluded. Some researchers have advocated dismissing results that have been gathered any sooner than six months after the treatment programme has ended (e.g., Feder et al., 2008). Had we done so, we would have excluded seven of the eleven studies in our sample.

Perhaps the most significant methodological shortcoming in our sample of studies was the near-total absence of comparison group evaluation designs. Of the eleven studies in our sample, only one used such a design (Dobash et al., 1999). However, the equivalence between the two groups in that study was suspect because the participants were allocated to groups based on their court sentence: the authors note that participants may have been mandated to attend the programmes only because the local law enforcement officials deemed them to be less severe offenders (Dobash et al., 1999:213). It is therefore plausible that this exaggerated the effectiveness of the treatment. The remaining studies only gathered outcome data from the participants who had undergone treatment. Although all eleven studies reported reductions in one or the other outcome measures, one cannot draw a causal conclusion. These modest to significant reductions may simply be an artefact of the “honeymoon period” phenomenon mentioned above, or they may be a true indicator of perpetrator programme effectiveness. However, in the absence of a comparison group, a number of threats to the internal validity hamper any claims regarding programme effectiveness (Shadish et al., 2002).

Programmatic Issues

Treatment Modalities

Even if the methodological problems of the primary studies in our sample would have been less serious, it would be difficult to identify the precise components of effective treatments. All eleven studies adopted an approach that mixed both cognitive-behavioural, educational, and ‘pro-feminist’ techniques. The specific method of treatment delivery in most cases was somewhat opaque and in other cases it was not described in sufficient detail. With one exception (Echeburúa and Fernández-Montalvo, 1997), all the studies used group therapy sessions which usually took place each week and lasted between one and two hours. These sessions were typically spread over twenty weeks, ranging from three to roughly thirty sessions (Socialstyrelsen, 2010; Dobash et al., 1999, respectively).

Although complex and theoretically heterogeneous programme packages seem to be most common and perhaps appropriate in practice, it is difficult to evaluate what components of the treatment may have led to more or less positive results. Therefore, it is not surprising that recent reviews, such as those by Babcock et al. (2004) and Feder et al. (2008), could not discern which treatment approaches (e.g., psycho-dynamic, cognitive-behavioural, etc.) work better than others. Day et al. (2009) explicated some of the theoretical inconsistencies in popular perpetrator programmes, and made a case for the need to base programmes on a more empirically verifiable and theoretically sound footing.

The documentation of programme delivery in the primary studies is also a problem. This seems to be a common feature of domestic violence perpetrator programme research (e.g. Mears, 2003). Deficits in delivery information may partially be due to limited space in journal articles, but also other forms of reports contain only descriptions such as “psycho-educational seminars” (APHVF, 2009). In the absence of further specification of the treatment through detailed documentation of process *and* outcome, we will remain unable to determine how best to improve practice in perpetrator programmes.

Offender Typology

There have been some advances in the literature supporting the notion that domestic violence offenders cannot be distilled into one unitary profile, and that perpetrator programmes would benefit from tailoring their treatment to their unique patterns and learning styles (Cavanaugh and Gelles, 2005). Consequently, large drop-out rates may partially be an indicator that programmes have successfully targeted a particular type of offender, whereas criminogenic needs of many other men have not been adequately addressed (Graham-Kevan, 2007:221). This hypothesis has ample support from elsewhere in the offender rehabilitation literature, where meta-analyses reveal the particular efficacy of programmes that adhere to

the risk, need and responsivity principles (RNR; Andrews and Bonta, 2010; Koehler et al., 2011; Lösel, 2011).

Of the eleven studies in our sample, none explicitly targeted the delivery of the treatment programme to specific characteristics of the intake sample. Although analysis was conducted on the effectiveness of the treatment on different subtypes of offender in Bowen (2004) and in Dobash et al. (1999), this was *post hoc* and therefore not used to match the treatment to the individual needs of the offender. Despite both repeated calls for perpetrator programmes to eschew the ‘one-size-fits-all’ philosophy of treatment delivery and Bowen’s (2004) finding of considerable variation in reoffending patterns across offender types within her study sample, our review could not detect any systematic evidence on individualised perpetrator treatment in Europe. This is clearly an area that requires more attention in practice and research.

Explanation of Low Methodological Quality

Although it is beyond the scope of this systematic review to explain why the methodological quality of evaluations in this field is so low, there are some reasonable avenues for speculation. In a recent survey of domestic violence perpetrator programmes throughout Europe (Hamilton et al., 2011), programme deliverers reported two misgivings about conducting a rigorous evaluation: firstly, they referred to the exigent need to ensure victim safety, regardless of whether the efficacy of the measure taken to do so had been proven; secondly, they expressed a fear that a negative outcome from an evaluation could lead towards a termination of financial support.

Five of the eleven evaluations in our sample were conducted by the developers or administrators of the programme. It is plausible that in these instances, the evaluation was

conducted not only as a test of the programme's effectiveness but rather as an administrative tally of the programme's intake and general performance. The role of programme deliverers as evaluators has been demonstrated to inflate scores in favour of treatment effectiveness (Petrosino and Soydan 2005), and this may perhaps also affect the choice of evaluation design.

Another possible explanation for the low methodological quality of the studies relates to the fact that most of the evaluations ($n = 6$) were conducted in the community, as opposed to custody. The formation of a control group is encumbered by the difficulty in acquiring perpetrators in the community who are interested in participating in evaluation studies. And last but not least one must assume cultural influences that go along with generally less openness to sound evaluations of offender treatment in Europe than in North America (e.g. Hamilton et al., 2011).

Although evaluative practice in European perpetrator programme research is clearly in need of improvement, it is nonetheless encouraging to note that evaluations are being conducted. Because five of the eleven studies in our sample were unpublished, practice and policy making would probably welcome that these documents become more easily accessible.

Conclusion

The studies in our sample do not, in aggregation, provide much guidance in determining what steps to take to improve domestic violence perpetrator programmes in Europe. The key message of this review remains therefore similar to the conclusions reached in previous reviews from North America, namely that we do not yet know what works best, for whom, and under what circumstances (e.g., Feder et al. 2008; Babcock et al. 2004; Davis and Taylor 1999; Hamberger and Hastings 1993). We cannot even establish validly that

domestic violence perpetrator programmes are successful in reducing episodes of future re-abuse at all.

Criminologists have become increasingly aware both of the “moral imperative” (Weisburd 2003) to deliver policy recommendations grounded in the highest internal validity and to the need to safeguard against recommending “cures that harm” (McCord 2003). The clarion call for more rigorous evaluation in domestic violence perpetrator programmes has been sounded for decades; it is high time that the call be heeded, and that we progress more beyond the above-mentioned ‘first generation’ of research in this field.

Appendix Ci: Search strategy terms

Subject Words	Programme Words	Outcome Words Methodology
Domestic violence	Programme(s)	Effect*
Domestic assault	Treat*	Outcome*
Batterer	Intervention(s)	Eval*
Family violence	Therapy	Experiment*
Physical Abuse	Counsel*	Randomised Controlled Trials (RCT)
Spousal abuse	Rehab*	Quasi (experiment*)
Inter-family violence	Court Decisions	Trial
Intimate partner violence	Mandated Court Decisions	Empirical
Duluth	Within Prison	Recidiv*

*search term wildcard

Appendix Cii: Databases searched

Electronic Databases	Hand Search of Gender Violence Journals	Government Publications
IBSS	Journal of Interpersonal Violence	UK Home Office Research Database
PsycInfo	Advances in Psychiatric Treatment	Brå-Swedish National Council for Crime Prevention
PsycArticles	Feminist Criminology	
PubMed	Feminist Theory	
Cochrane Library	Violence against Women	
C2-SPECTR	Journal of Family Violence	
EmBase	The Family Journal	
ISI Web of Knowledge	Feminism and Psychology	
CSA Illumina:		
<i>Criminal Justice Abstracts</i>		
<i>Applied Social Sciences Index and Abstracts</i>		
<i>Conference Papers Index</i>		
<i>ERIC</i>		
<i>Medline</i>		
<i>CSA Social Services Abstracts</i>		
<i>CSA Sociological Abstracts</i>		

Programmes to Reduce Reoffending Throughout Europe: Three Surveys on Current Practice*

Leah Hamilton¹, Johann A. Koehler¹,

and Friedrich Lösel^{1,2}

¹ Institute of Criminology

University of Cambridge, UK

² Institute of Psychology

University of Erlangen-Nuremberg, Germany

* This study was carried out within the European cooperation project “Strengthening Transnational Approaches to Reducing Reoffending”. The project was funded by the European Union. We thank our partners from the UK Ministry of Justice, London Probation Trust, European Organization for Probation, and the Ministries of Justice of Bulgaria, France and Hungary for their cooperation. We graciously acknowledge the assistance provided by the many experts and practitioners who responded to our questionnaire, and by the Daphne II Work with Perpetrators Survey who furnished us with data that assisted in search for programmes.

Introduction

In addition to the STARR Cambridge Research Team's three systematic reviews, we have conducted three questionnaire surveys investigating the state of current practice in programmes to reduce reoffending among young offenders, alcohol and substance abusing offenders, and domestic violence perpetrators.

Previous efforts to identify the current range of measures to reduce reoffending throughout the European Union were limited in scope. Our initial search of the literature revealed that no comprehensive list of reoffending programmes in each of the 27 countries existed. Although available documents enable some overview, they typically contain a summary of evidence on the effectiveness of national policies (e.g., Wartna and Nijssen, 2006), lack of a detailed description of the programme components (e.g., EMCDDA, 2011), or provide rudimentary information on programme theories without data on empirical evaluation (e.g., Rothman et al., 2003; Work with Perpetrators Survey, 2008).

The surveys administered by the STARR Cambridge Team aimed to reduce these knowledge gaps and to advance our understanding of current measures to reduce reoffending in the three above-mentioned domains. This could be achieved by compiling detailed descriptions on a larger sample of programme practices across all 27 EU.

Method

In our search for programmes to reduce reoffending, we looked for any service that saw reducing reoffending as a goal or by-product of treatment. We were specifically interested in *re-offending* programmes, which necessitate that the participant has had previous interactions with the criminal justice system in the wider sense (e.g., official charges, arrest,

or warnings from police or other justice representatives. We excluded primary prevention measures such as awareness campaigns or early work with children and families at risk.

The questionnaires were composed of multiple choice and semi-structured questions, and gathered information from the programmes on five themes in each offense category: (1) the theoretical model of change, (2) administrative data, (3) programme content, (4) process evaluation, and (5) outcome evaluation. The questions addressed: the type and theory of the programme; participant and staff selection and characteristics; programme organisation; programme implementation and effectiveness; and general features of the infrastructure of rehabilitation programming.

We recruited respondents from four information sources. These streams of recruitment fed into each other and informed us of other programmes elsewhere:

Ministries of Justice

Ministry of justice officials provided us with the details of government- and externally-run programmes to which they refer offenders. In the case that such programmes were the responsibility of a different ministry, we were referred onwards. For example, social-welfare ministries are often responsible for addressing all matters related to domestic violence, and ministries of health are often responsible for drug and alcohol abuse treatment.

Expert Contacts

We contacted well-known European experts in the fields of research within the STARR Project's purview. This was supplemented by contacts made both during the STARR conferences and seminars, and those resulting from networks arising out of our own survey searches.

Pan-European Associations

We also contacted international monitoring agencies (e.g., EMCDDA) and practitioner networks (e.g., CEP). These organisations either provided contact details of programme administrators, or they were linked to local experts who had knowledge of national or regional programmes.

Community Sector Search

Lastly, we contacted third sector and/or private organisations to locate community-based programmes.

The questionnaires were distributed sequentially. The young offender questionnaire was distributed first, in July 2010. The domestic violence questionnaire was distributed in October 2010. Lastly, the substance abuse questionnaire was distributed in February 2011. All responses were completed and returned by the end of June 2011.

Contact was made with participants via telephone or email. We allowed people to respond to the questionnaire either in their mother tongue or we translated the questionnaire completely. The questionnaires were translated into French, German, Spanish and Hungarian to accommodate participants. After initial contact we followed-up with the respondents on their progress and collected completed questionnaires. Most programmes returned their questionnaires in one to two months, and were in contact with the research staff multiple times. Occasionally this process took many months with frequent contacts by phone and email to gather back completed responses to the questionnaire. A total of 250 programmes were located with representation from all 27 countries.

While we strived for representation of all EU countries, some countries have greater representation in the sample than others. This can be explained by various reasons. For example, it is plausible that larger countries have more programmes in place than smaller

ones. This is supported by high programme numbers in Germany or the UK. However, other large countries such as France and Italy reported only a few programmes and some smaller countries such as the Netherlands or Czech Republic reported a higher number. With regard to Germany one must also take into account that criminal justice is organised federally; therefore we received questionnaires from different states (“Länder”). A relatively small number of programmes should not be interpreted simply as less activity in offender rehabilitation, but may indicate a more uniform approach instead of a range of programmes. In addition, we must assume differences in the response behaviour of the local practitioners. In some countries they may have been more ready or better informed to fill in questionnaires on different programmes than in other countries. As there is no representative documentation of young offender, drug abusing offender or domestic violence perpetrator treatment programmes in Europe, we cannot validly estimate what factors determined the different prevalence rates in Figures 1, 3 and 5.

A second difficulty of the survey is related to the content of the responses. As these are self-reported questionnaires, the data gathered can be influenced by various response biases that are well known in research (Lösel and Schmucker, 2002). For example, respondents may have given socially desirable answers and presented their programmes in a more favourable manner. There may also have been information gaps with regard to specific programmes. In addition, different informants may have invested more or less time in answering the questions (which is particularly relevant for the open questions). Last but not least language problems may have played a role, although we often helped by providing translations or intensive guidance via telephone. Although we cannot exclude the above-mentioned and other influences, it needs to be emphasised that they are never fully avoidable in surveys of our type.

The answers to the semi-structured questions were coded by two members of the research team. Average Kappa scores for each offence category all displayed strong inter-rater reliability: Young Offenders: $\kappa = 0.98, p = 0.02$; Substance Abuse: $\kappa = 0.92, p = 0.04$; Domestic Violence: $\kappa = 0.95, p = 0.03$.

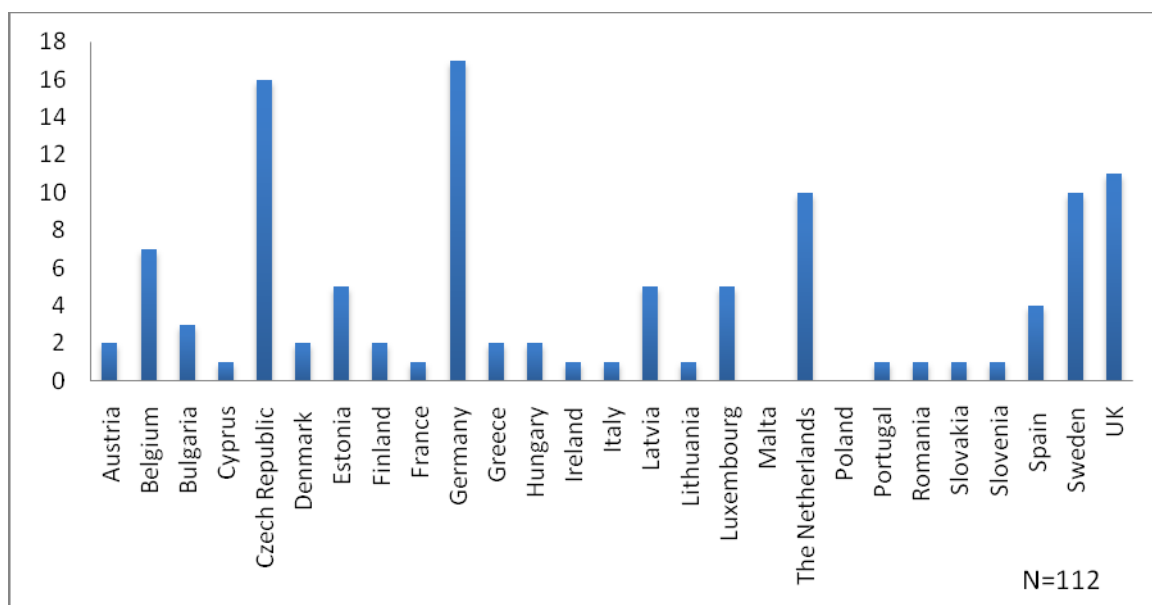
Young Offender Programmes

Results

The young offender survey located a total of 112 programmes from 25 of the 26 European Union countries with young offender programmes. Figure 1 shows the distribution of reported programmes across the various countries.

Some countries had greater representation in the survey, particularly those from a few countries in North-western Europe. However, even small countries provided information on some programmes, which enables a reasonable reflection of European young offender rehabilitation programmes. Table 1 contains a selection of results from our survey.

Figure 1: *Number of programmes for young offenders in different countries*



Programme Characteristics

One third ($N = 37$) of the programmes were adapted from another country. The most frequently cited country of origin was the United States ($N = 14$), though there were a few inter-European transfers in the survey. Cognitive-behavioural treatments tended to be most often transferred ($\chi^2_{(2)} = 9.64$, $p = 0.08$, $\phi = 0.293$). There was no relationship between programme transfer and programme evaluation or programme accreditation.

Almost three-quarters of the programmes were described as being structured to a great extent (71%). Barring 2 programmes that were indicated as having very little structure, the rest of the programmes (25%) were reported to have somewhat structured treatment services. Programmes that were transferred were more likely to be highly structured ($N = 30$ out of 37). This is in accordance with the above result on cognitive-behavioural treatment modalities.

Programmes can be classified into five main types: cognitive/behavioural programmes, education programmes, non-behavioural therapy programmes, therapeutic communities and deterrence-based measures. As Table 1 shows, cognitive/behavioural programmes were the most common form of treatment by far (78%). The second most frequent form of treatment was education-based programmes (44%), followed by non-behavioural therapy (35%) and therapeutic communities (10%). Purely punitive and deterrence-based measures were the least often used intervention type (5%). The sum of proportions exceeded 100% because many programmes contained a combination of the various approaches.

Table 1: *Characteristics of the programmes for young offenders*

Programme Component	Variables	Frequency	Percent	N
<i>Transfer</i>	Transferred	37	33%	112
	Not Transferred	75	67%	
<i>Structure</i>	Little to None	2	2%	109
	Somewhat	28	26%	
	To a Great Extent	79	72%	
<i>Treatment Modality†</i>	Cognitive/Behavioural	87	78%	112
	Education-based	49	44%	112
	Non-behavioural	39	35%	112
	Therapeutic Communities	11	10%	112
	Deterrence/Intensive Supervision	5	5%	112
<i>Assessment†</i>	General Assessment	78	70%	112
	Systematic Risk Assessment	50	45%	112
	Tailored to Risk	87	78%	112
<i>Programme Location</i>	Community	43	38%	112
	Custody	54	48%	
	Community & Custody	14	14%	
<i>Funding</i>	Government	100	90%	111
	Non-governmental Organisation	5	4%	
	European Union	3	3%	
	Client Fees	2	2%	
	Multiple Sources	1	1%	
<i>Accreditation</i>	Accredited	75	67%	112
	Not Accredited	37	33%	

† percentages do not sum to 100, as respondents could indicate multiple responses

The average duration of the programmes is 11 weeks, but there is a large variation of programme length ($SD = 88$ weeks). The mean duration of sessions is 1.5 hours ($SD = 77$ minutes).

Offender Characteristics

When dealing with young offenders, it is essential to identify and address key risk/need factors (Andrews and Bonta, 2010:72-74). In order to determine the risk level of reoffending a systematic risk assessment should be carried out. However, this form of assessment accompanied less than half of the programmes (45%) in our survey. Despite this fact, it was reported that 78% of the programmes had been tailored to the participants' risk

levels to some extent. A part of this difference may be accounted for by the 70% of respondents who indicated that general assessments were carried out. Systematic risk assessment did not go along with a risk-tailored programme. Of the 50 programmes that carried out *systematic* assessment, two-thirds reported that the typical clientele of the programme was medium to high risk.

Although European countries vary in their age of criminal responsibility, most programmes address offenders above 14 years. The most frequent response to the item concerning participant age ('Other'; 66%) suggests that the programmes are not targeting a narrowly defined age group, but a broader age range. Where specific age groups were mentioned, the focus is on 15 to 18 years (24%). Only a small proportion (9%) address young adult offenders of age 19 to 25. Many respondents indicated that specific ethnicity data is not collected in their country due to legal restrictions; therefore, we could only analyse whether programme participants were local nationals of the country in which they received the programme. On average 73% ($SD = 20$) of the participants in young offender programmes were local national residents.

Programme Organisation

Correctional programmes for young offenders in Europe are more frequently located in a custodial setting (48%) than in a community setting (38%). An additional 12.5% of programmes operate in both locations. The specific setting of these programmes varies greatly from youth and adult custodial institutions to schools, residential care facilities and home visits in the community.

In most cases (82%) the developers of a programme are also involved in its implementation. This indicates that many programmes are not national or international standard routines but are rather more locally constructed or adapted interventions.

Government funding accounts for the vast majority of the programmes' financial resources. 89% of the programmes received more than half of their funds from a governmental branch. Though the majority of programmes were state-funded, there was a more diverse pool of alternate sources of support such as non-governmental organisations ($N = 5$), European Union ($N = 3$) and private contributions ($N = 2$) than in the substance abusing offender survey or domestic violence perpetrator survey. For 67% of the programmes it was reported that they had been accredited. Over three-quarters of those programmes received their certified status from a government body.

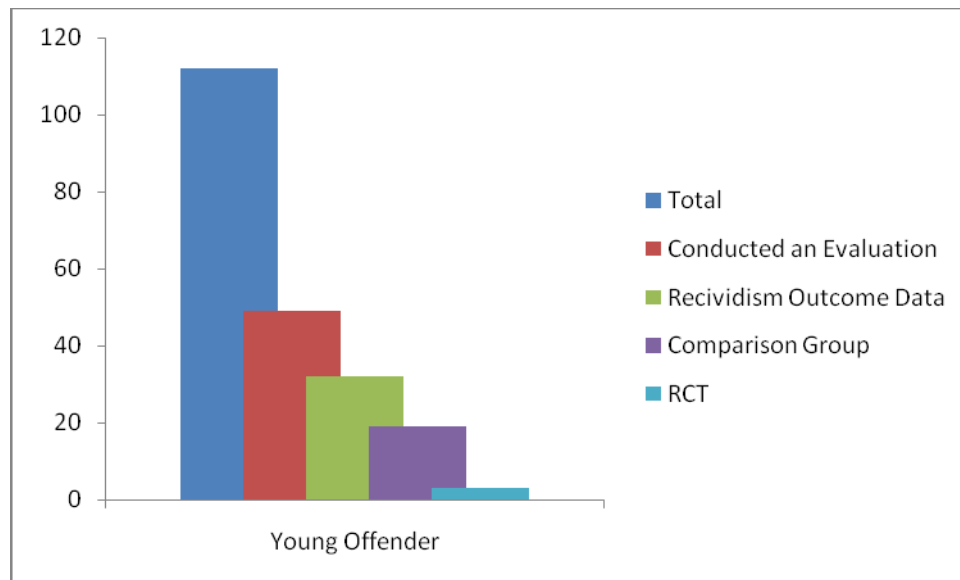
Process and Outcome Evaluation

The survey gathered information on what kind of process and outcome data were collected, and what the results were of these evaluations. In the young offender programmes surveyed, 71% of respondents stated that they took steps to ensure the quality of the respective programme. However, this positive result is not supported when it comes to concrete measurements of implementation quality. For example, only 21% measured whether participants attended the programme on a regular basis and less than half (45%) could report whether participants had completed the treatment.

At first glance, a relatively large proportion of programmes underwent an outcome evaluation (44%). Over two-thirds of these ($N = 32$) used indicators of recidivism as their outcome variable. However, only 59% ($N = 19$) of the latter used a comparison group design.

This corresponds to 17% of the original 112 programmes. Figure 2 shows the decrease in outcome evaluations when minimal criteria were applied.

Figure 2: *Findings on outcome evaluation of young offender programmes*



It must be noted that the comparison groups are not necessarily control groups that were equivalent to the treatment group. Only three of the 19 programmes that used comparison group designs were randomised control trials (RCTs). 13 of the comparison groups had some matching and the remaining 3 used quasi-experimental designs. A further analysis revealed no significant relationship between programme evaluation and the statement of some kind of accreditation.

Respondents were also asked in an open format to provide further thoughts and experiences on why programmes matter, what the obstacles were to running treatment programmes and what they believed was necessary to improve practice. 71% of respondents indicated an interest in maintaining programme quality; however, respondents frequently cited lack of funding, of professional support, of staff training, of participant suitability and organisational difficulties as challenges to successful programme implementation.

Discussion

The STARR survey of young offender programmes is the first of its kind to gather detailed information on existing practices in reducing reoffending programmes. The survey contains information on a broad range of programmes and represents 25 of the 26 EU countries that have some form of young offender rehabilitation regime. In spite of these positive facts there are limitations regarding the representativeness of the survey data (see above).

Taking such cautions into account, our survey's results on the type of treatment draw a rather positive picture. The most common type of intervention (78%) is a cognitive-behavioural approach. This treatment mode has not only shown the strongest effects in numerous North American evaluations (Aos et al., 2006; Lipsey and Cullen, 2007; Lösel, 2011; MacKenzie, 2006), but also in our STARR systematic review of European studies (Koehler et al., 2011; see Appendix A). In this respect European practice seems to adhere to the evidence-based 'what works' literature. However, as mentioned, we cannot rule out that some respondents may have labelled their programmes as 'cognitive-behavioural' although they in fact take a more unspecified approach.

The second and third most common forms of treatment in our survey, education-based treatment and non-behavioural therapy, are less strongly supported by the above-mentioned research. Meta-analyses show lower effects and also more mixed results (Aos et al., 2006; Lipsey and Wilson; MacKenzie, 2006). More evidence is needed to determine which non-behavioural and education programmes may be more effective. For example, mentoring programmes appear to be promising, but the evidence is still sparse, particularly in Europe (Joliffe and Farrington, 2008; Koehler et al., 2011). Therapeutic communities, which

represent a rather small portion of our sample, are also a promising form of treatment (Lösel and Egg, 1997; Pearson et al., 2002), but most of this research addresses adult offenders.

The least common form of intervention within our sample was deterrence or intensive supervision. According to the international ‘what works’ literature, such correctional programmes show no or sometimes a slightly criminogenic effect (Andrews and Bonta, 2010; Lipsey and Cullen, 2007; MacKenzie, 2006). This tendency is also confirmed in our STARR review of European studies (see Appendix A, Koehler et al. 2011). Although calls for more punitive measures for young offenders are rather popular in many European countries, practice seems to adhere to a more evidence-oriented strategy.

In contrast to the type of intervention, our findings on the programme setting are somewhat less positive. More programmes are carried out in custody than in the community, although the Anglo-American ‘what works’ research and our European review showed larger effects in the latter context (see Andrews and Bonta, 2010; Lipsey and Cullen, 2007; Lösel, 2011; Appendix A). Of course, one cannot avoid placing high-risk violent offenders in custody. However, our survey and the systematic review suggest that Europe should invest more in the development of community treatment programmes for those young offenders for whom community treatment is legally and empirically appropriate. This could be offenders at moderate to high risk (Andrews and Bonta, 2010), but in relatively low categories of potential harm doing.

Approximately a third of the respondents indicated that their programme had been transferred and adapted from another country. Most of these programmes were cognitive-behavioural. However, there was no correlation between programme transfer and evaluation. This lack of relationship may suggest that programme administrators are relying on evidence of their programmes’ effectiveness in a different context. As Sundell et al. (2008) reported on

the transport of Multi-Systemic Therapy from the US to Sweden, differences in the social infrastructure and the cultural context may have led to a lack of effect of MST on recidivism in Sweden. In addition, most transferred programmes seem to undergo some modification. This is indicated by the fact that in 90% of all programmes the developers have been involved in the delivery in some capacity. Although some modification may be necessary to adapt to cultural and organisational differences, this has the consequence that one cannot simply rely on evaluations from other countries. The latter are also often carried out as demonstration projects that typically show larger effects than programmes in routine practice (Landenberger and Lipsey, 2005; Lösel, 2011).

Nearly three quarters of the respondents noted that their programmes were structured to a great extent. This may be because programme administrators sometimes believe that young offenders require a more structured approach in order to learn. However, it is possible that at least some of the responses about structure may have been influenced by a social desirability bias because the ‘what works’ findings also suggest structured approaches have a positive impact on programme efficacy (Andrews and Bonta, 2010). The relationship between programme transfer and the level of programmes’ structure is a logical one: programmes with more structure are normally manualised, and thus, may require less developer involvement in the implementation. Another possible explanation is that this result is spurious because cognitive-behavioural programmes are both more often transferred and highly structured (Wilson et al., 2005).

Research has shown that thorough offender assessment and programme adaptation to the participants’ risk and need factors leads to larger effects (Andrews and Bonta, 2010; Andrews et al., 2011). With regard to this issue our survey found mixed results. Although about three quarters of programmes included some kind of general assessment (e.g., intake interview, clinical judgment or standardised structured procedures), only 45% of respondents

said that they carried out *systematic* measures using risk assessment tools. Therefore, less than half of the programmes include a diagnostic approach that is testable with regard to its predictive power. This also suggests that the reported frequent tailoring of programmes to the offender's risk level cannot be very detailed and precise. We do not know how much this may undermine programme effectiveness.

Our survey shows that the vast majority (90%) of the reported programmes are organised and funded by government bodies. This suggests that NGOs and private institutions still play a much less important role in the organisation of correctional treatment for young offenders in Europe. However, we believe that the real involvement of the 'third sector' and private organisations is stronger than the figures indicate. A first reason for this assumption is that government-controlled programmes may sometimes be delegated to NGOs or private organisations. Secondly, small non-governmental organisations may be somewhat underrepresented in our survey. Although we targeted a broad range of experts in the various countries, smaller non-governmental organisations may have partially fallen through the net of contacts.

Thorough programme implementation will impact how effective a programme is at reducing reoffending (Lipsey, 1999:159). Therefore, process evaluation is an important part of sound correctional treatment. Our survey data suggest, however, that the collection of process data needs to be improved. Nearly one third of respondents did not indicate that they were taking any steps to maintain the quality of the programme. Less than half did not keep track of the completion rates of their programmes and only about a fifth stated that they regularly record the participants' attendance. This is problematic insofar as our systematic review showed that programmes with unrecorded dropout rates had no significant effect (Appendix A).

At first glance the proportion of programmes that had gone through, or were in the process of undergoing, an outcome evaluation was relatively high (44%). This is remarkable as our survey addressed programmes in practice and not academic demonstration studies. However, as Figure 2 shows, this proportion shrinks to 17% when we ask for minimal criteria such as an untreated comparison group and recidivism outcome data. Unfortunately, this final number is still too large, as our systematic review revealed that only three of the 112 programmes reported in the survey used a randomised design. When comparison groups are not equivalent, one cannot draw valid conclusions about programme effectiveness. Therefore, the findings of our survey clearly suggest that we need a better practice of sound outcome evaluations of young offender treatment in Europe.

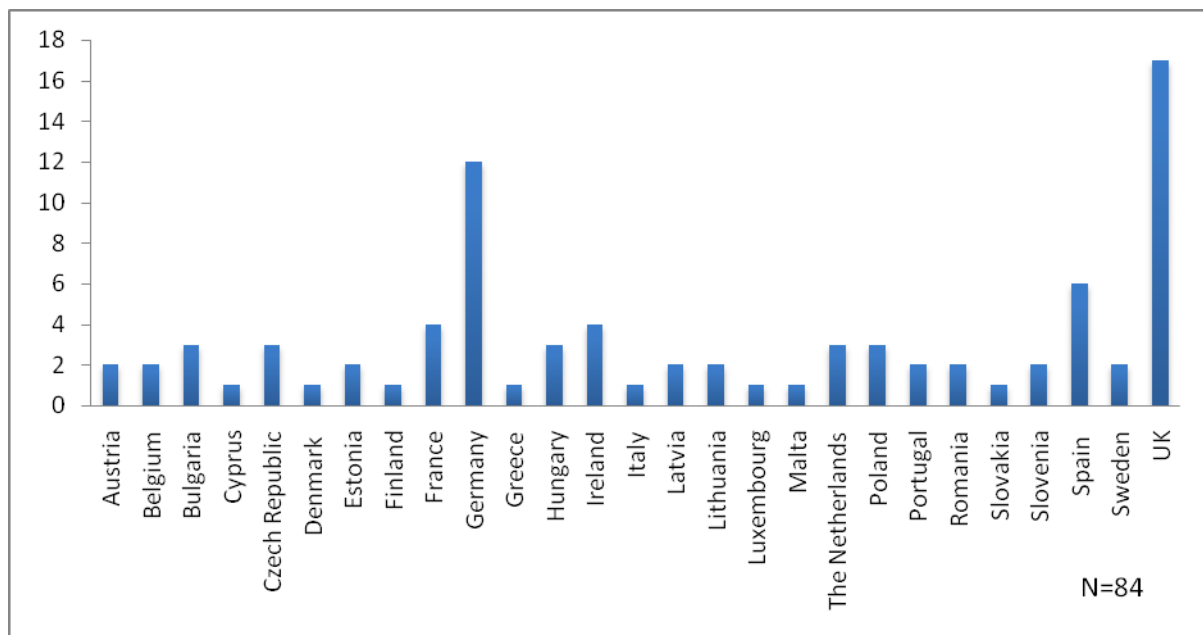
Programmes for Substance Abusing Offenders

Results

Responses

The survey on programmes for substance abusing offenders located a total of 84 programmes from all 27 European Union countries. Figure 3 shows the distribution across the different countries. Some countries had greater representation in the survey, particularly North-western European countries; however, sufficient information was provided from the questionnaires to give a reasonable reflection of European substance abusing offender programmes. Table 2 contains a selection of findings on programme and related characteristics.

Figure 3: *Number of programmes for drug abusing offenders in different countries*



Programme Design

Respondents indicated that they apply up to four different treatment modalities. The most commonly cited type of treatment was cognitive/behavioural-based therapy (70%) such as, for example, the UK's Prison Addressing Substance Related Offending (ASRO). Non-behavioural therapy methods like the 12-Step Programme were also cited as a common form of treatment (49%). Pharmacological programmes were reported less frequently (14%). Approximately one fifth (21%) of the programmes were therapeutic communities. The sum of proportions exceeded 100% because programmes often contained a combination of treatment modes.

Table 2: *Characteristics of the programmes for drug abusing offenders*

Programme Component	Variables	Frequency	Percent	N
<i>Treatment Modality†</i>	Cognitive/Behavioural	59	70%	84
	Non-behavioural/Counselling	41	49%	84
	Therapeutic Community	18	21%	84
	Pharmacological	12	14%	84
<i>Transfer</i>	Transferred	25	30%	83
	Not Transferred	58	70%	
<i>Structure</i>	Little to None	4	5%	84
	Somewhat	29	35%	
	To a Great Extent	51	61%	
<i>Assessment†</i>	General Assessment	60	71%	84
	Systematic Risk Assessment	42	50%	84
	Tailored to Risk	57	68%	84
<i>Programme Location</i>	Community	34	41%	84
	Custody	44	52%	
	Community & Custody	6	7%	
<i>Recruitment</i>	Voluntary	19	23%	84
	Coerced	38	45%	
	Mixed	27	32%	
<i>Funding</i>	Government	82	93%	84
	Client Fees	1	3%	
	Other	1	3%	
<i>Accreditation</i>	Accredited	38	45%	84
	Not Accredited	46	55%	

† percentages do not sum to 100, as respondents could indicate multiple responses

Nearly a third (30%) of respondents indicated that their programme had been adapted from another country. While there was no dominant transferred programme, three respondents explicitly stated that their work was an adaptation of the American 12-Step Programme. All of the transferred programmes in the sample were funded from government sources. There was a statistically significant relationship between non-transferred programmes and a lack of accreditation ($\chi^2_{(2)} = 8.08, p = 0.02, \phi = 0.32$) indicating that grass-root programmes are more often not accredited.

Almost all the respondents stated that their programmes were somewhat structured (35%) or structured to a great extent (61%). Only 4 programmes had little to no structure in their programme design.

Offender Characteristics

The most common response with regard to age of participants was 'Other' (52%), indicating that most respondents did not specify the age of the participants. 16% explicitly stated an age range of 20-29 and 12% stated a range of 30-39. Again, as in the case of young offender programmes, in some countries data could not be collected on ethnicity; instead data on nationality was gathered. The mean percentage of programme participants that were local national residents to the country of treatment was 78% ($SD = 23$).

General assessments of participants were conducted in 71% programmes before intake. Only 50% of programmes were accompanied by a systematic risk assessment for reoffending. Over two-thirds of the respondents indicated that their programmes were somewhat (44%) or to a great extent (24%) tailored to the participant's risk level of reoffending. Tailoring to risk level is significantly correlated with conducting a systematic assessment of risk level for reoffending ($\chi^2_{(1)} = 5.90, p = 0.02, \phi = 0.27$), though the effect is small.

Programme Organisation

The survey results show that over half (52%) of the drugs and alcohol programmes took place in a purely custodial setting. Two fifths (41%) of the programmes were located in the community and 7% of programmes were carried out in both settings. The modal

recruitment type was mandatory (45%). Less than one quarter (23%) of the respondents stated that participation in the programme was purely voluntary.

The vast majority (93%) of the programmes received government funds as their primary source of support. One programme had the European Union as its main source of financial support. Only one programme used participant fees to support the service. Less than half (45%) of the programmes received some kind of accreditation. Government boards were responsible for three-quarters of programme accreditation.

Process and Outcome Evaluation

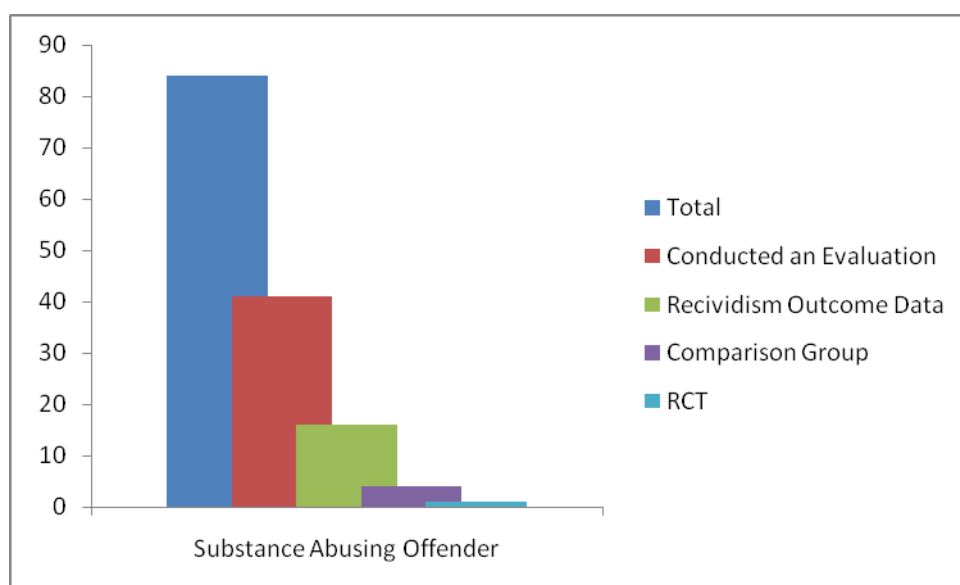
Many respondents said that process evaluation data was gathered throughout implementation. 87% stated that they took measures to ensure the quality of the programme. The most frequently cited forms of quality assurance were supervision (52%) and training (27%). It appears that few records were kept of basic measurements on the execution of programmes. For example, only 45% of programmes kept track of participant attendance, and 35% of programmes gathered participant feedback data.

The results on outcome evaluation are shown in Figure 4. Just under half of respondents indicated that they had conducted an evaluation, but on further inspection only 39% of the evaluated programmes used either official or self reported recidivism as the outcome measure. Outcome data was mostly collected both before and after an intervention. Only 4 (5%) of the evaluated programmes used both recidivism as the outcome measure and had a comparison group design. Only one programme was evaluated in an RCT.

Respondents were also asked in an open format to provide further thoughts and experiences on why programmes matter, what the obstacles were to running treatment

programmes and what they believed was necessary to improve practice. While nearly all the respondents were committed to carrying out high quality programmes that would reduce reoffending, there were a number of challenges in executing their programmes accordingly. Respondents frequently cited lack of funding, of professional support, of staff training and of participant suitability as sources of difficulty in programme implementation. Only one informant mentioned lack of evaluation as a critical problem.

Figure 4: *Findings on outcome evaluation of drug abusing offender programmes*



Discussion

This survey is the first of its kind to gather data from all 27 European Union countries. A previous survey by the European Monitoring Centre for Drugs and Drug Addiction's programme 'Exchange on Drug Demand Reduction Action' located only 33 such programmes and 14 countries were not represented (EMCDDA, 2011). Therefore, our survey advances the knowledge-base considerably and also provides a specific focus on reoffending programmes. Despite these strengths one must be aware of the limits of representativeness and potential response biases (see above). In addition, it needs to be mentioned that the

questionnaire gathered data on substance abuse programmes that *seek to reduce reoffending*. Many drug and alcohol treatment service in Europe operate separate of the criminal justice system. For example, only 29% of referrals to community treatment services in the UK in 2008 came from the criminal justice system (National Treatment Agency for Substance Misuse, 2010). Therefore, one should not generalise our findings to drug treatment as a whole.

Drugs and crime can be related because of pharmacological, economic, and systemic reasons (Kleiman, et al., 2011). The national conceptualisations of drug and alcohol abuse and their relationship to crime plays a critical role in determining the framework, whether it be public health or criminal justice, in which substance abuse is addressed. For countries such as Portugal where possession, acquisition and use of drugs is decriminalised the approach is public health-oriented (Hughes and Stevens, 2007), whereas in countries such as Sweden and the UK the criminal justice system is intimately involved in dealing with substance abuse.

The results of our survey indicate that programmes were 10 percentage points more frequently implemented in custodial than community setting. This is probably a result of specifying reoffending as one of the treatment goals.

Research on general drug treatment suggests that treatment of ‘coerced’ participants is as effective as treatment for voluntary participants (Stevens et al., 2005). As this is the case, it seems to be adequate that nearly three quarters of the programmes in our survey addressed offenders that were referred in part from a criminal justice agency. Purely voluntary participation was much less frequent (23%). Given the high representation of compulsory, or at least “quasi-compulsory” programmes (Stevens et al., 2005), practitioners seem to agree with the research that supports the use of mandatory programmes for substance abusing offenders.

Unlike the American system where private treatment clinics are more common (Prendergast et al., 2002), in our sample governmental bodies, whether they be justice- or health-oriented, support nearly all the interventions. Though less than half of the programmes have undergone some form of accreditation, three-quarters of those programmes were accredited by a governmental or government affiliated body. These data indicate at least moderate state control of the programmes for drug abusing offenders.

The results of our survey show that cognitive-behavioural therapy is most frequent in practice (70% of all reported programmes). Cognitive-behavioural programmes have been proven effective at reducing reoffending with the general offender population and there are also various Anglo-American reviews that demonstrate mean positive outcomes for drug- or alcohol-addicted offenders (Lipsey and Wilson, 1998; Pearson and Lipton, 1999; Mitchell et al., 2007). However, our STARR systematic review of drug treatment programmes did not reveal substantial research on this type of programme in Europe (Koehler et al., 2011 see Appendix B). The same applies to therapeutic communities. International reviews on therapeutic communities demonstrated desirable effects for drug addicted offenders (Holloway et al., 2008; Mitchell et al., 2007; Pearson and Lipton, 1999). However, according to our survey this promising type of programme is not particularly frequent in European practice (21%). The least common form of intervention within our sample was pharmacological treatment, i.e. substitution programmes (14%). In contrast, our STARR systematic review revealed that pharmacological programmes were based on a number of sound evaluations and significantly reduced reoffending (see Appendix B). In our view this is an important discrepancy that may be due to political controversies about legal heroin provision or substitution in a number of European countries. Of course, empirical evidence cannot solve such conflicts and one must also be aware that competing definitions and

conceptualisations of drug abusing offender programmes hold considerable influence in the choice of programme adopted.

Since many European countries offer some form of substitution programme, the dearth of these treatments in our sample can result from organisational structures. Substitution or legal heroin provision are often public health oriented and either do not receive clients from the criminal justice system or do not directly address reducing reoffending. One must also take into account that substitution is still a politically controversial topic in various countries. Another explanation given by a substitution treatment agency in our survey was that pharmacological treatment services do not conceive of themselves as 'programmes' because their treatment varies greatly depending on the client. Given the demonstrated effectiveness of pharmacological treatments to reduce reoffending, this approach should be considered more frequently in Europe. However, it should not be seen as the only solution, but also systematically investigated in combination with the most prevalent, but not sufficiently evaluated cognitive-behavioural programmes.

A third of programmes in the sample were transferred from other countries. As discussed in our survey of young offender programmes (see above), recipient states must consider that programmes may be more or less effective depending on where the programme takes place. Factors such as developer involvement, cultural difference and systemic differences may affect how well a programme works in a given context (Sundell et al., 2008). In our survey there is neither a significant relationship between programme transfer and evaluation nor between programme evaluation and accreditation. This suggests that little is being done to determine whether programmes maintain effectiveness after transfer. It also indicates that accreditation requires a detailed programme assessment and locally sound evidence. At the moment neither home-grown programmes nor transferred programmes seem to be the subject of sufficiently valid quality standards.

Compared to the other offender categories evaluated in the STARR project, substance abusing offender programmes conduct far more *systematic* assessments of risk (instead of general assessments). This is plausible because drug treatment services have ties to the medical system as they take a physical toll on the user. This may influence programme practitioners to use more systematic assessment methods, similar to medical records. It is noteworthy that systematic assessments are more prevalent than in the youth offender field and that also in both areas cognitive-behavioural programmes are the dominating mode of intervention. Regular use of systematic risk assessment tools would address some of the concern expressed by respondents regarding participant suitability to their programmes. Over two-thirds of the programmes were tailored to risk level. Though some of the programmes were tailoring according to a systematic risk assessment, others seem to rely more on clinical judgments. There was a relationship between systematic risk assessment and risk-oriented programme tailoring, but this was only weak.

Although our respondents stated that 87% of the programmes gathered process data during the implementation, the picture is less positive with regard to some elementary records such as participants' attendance and feedback (both below 50%).

However, more alerting is the situation with regard to outcome evaluation. Less than half of the respondents mentioned conducting some kind of evaluation, but these seem to be mainly reports on drug-consumption, and psychological measures of before-after data on deviant behaviour. Only four programmes (5%) conducted or are conducting evaluations based on recidivism data and containing a comparison group. And even these few outcome evaluations did not include truly equivalent control groups, thus excluding these evaluations from our systematic STARR review (see Appendix B). The latter located 13 studies conducted in Europe, whereas Perry et al. (2008) found 23 American randomized control trials (RCT) alone. As most of the sound studies in Europe focussed on pharmacological

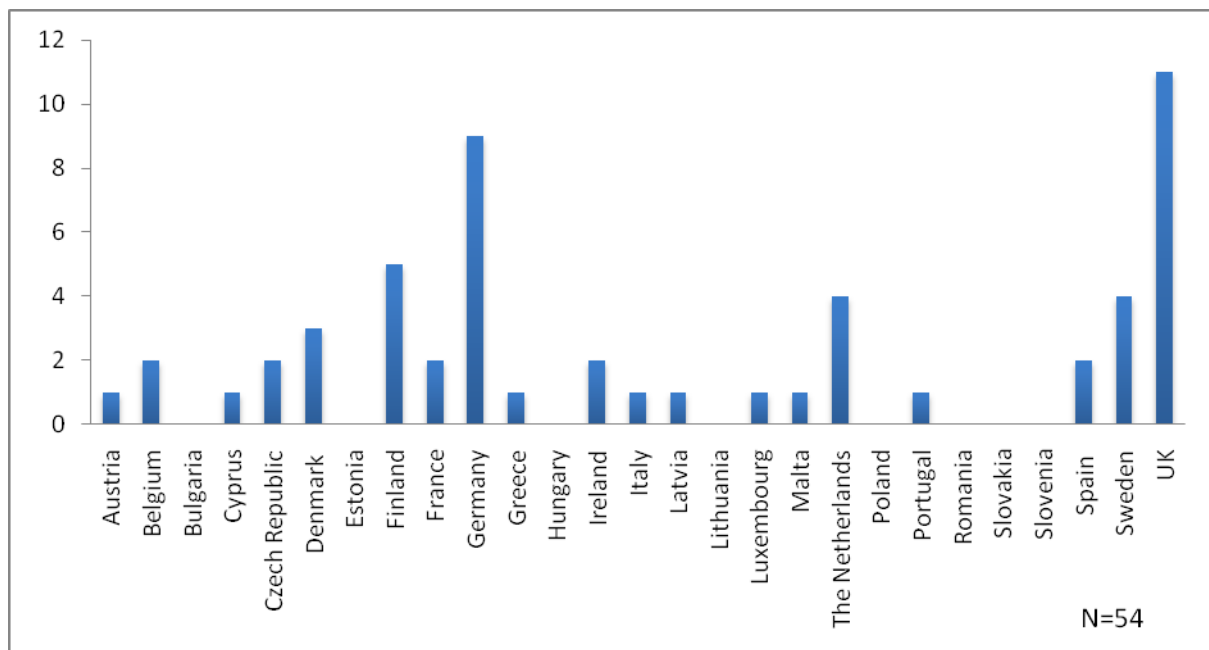
treatment, there is a clear need for further systematic evaluation on psychosocial programmes for drug abusing offenders. However, in our open questions on challenges and difficulties, very few respondents indicated the lack of evaluation as a problem. It appears that there is insufficient information on why evaluation should be done and how to go about doing one. Because programme effectiveness is highly relevant for the participating offenders, potential victims, service deliverers, criminal justice and the society as a whole (also for financial reasons), systematic evaluation of programmes for drug abusing offenders should receive more political attention.

Domestic Violence Perpetrator Programmes

Results

The domestic violence survey located a total of 54 programmes from 19 of the 23 European Union countries with existing domestic violence programmes. The distribution of programmes across the various countries is shown in Figure 5. Some countries had greater representation in the survey, particularly those from a few countries in North-western Europe. However, sufficient information was provided from the questionnaires to give a reasonable reflection of the current European treatment programmes in this field. We located four countries that had not yet developed any domestic violence perpetrator interventions (Bulgaria, Estonia, Hungary and Romania). While we were aware of at least one programme in 3 of the 4 missing countries, we were unable to obtain a response within the time constraints of the project. Although our survey provides new information in an under-investigated field, one must be aware of the same limits of representativeness and potential response biases as in the survey on programmes for young offenders (see above).

Figure 5: *Number of programmes for domestic violence perpetrators in different countries*



Programme Design

The various interventions can be grouped into four main categories: cognitive-behavioural therapy, non-behavioural psychological therapy, systemic therapy, and feminist approaches. Almost all programmes reported multiple treatment modalities. 91% of programmes indicated they used cognitive-behavioural therapy. Despite the dominance of the latter, it is more accurate to refer to popular cognitive-behavioural *techniques*, because fewer programmes are routed in cognitive-behavioural *theory*. Only 47% of respondents indicated that cognitive-behavioural theory was part of the main theoretical background of their programme. 47% of respondents stated that they used non-behavioural psychological therapy and 42% said that one of their treatment modalities was systemic therapy. Less than a quarter (22%) of respondents explicitly stated that feminism was a part of their main theoretical background; all of these used cognitive-behavioural techniques. When asked whether gender specific power and control was included in their programme, 68% of respondents said ‘yes’.

34% of domestic violence programmes were transferred from another country. The most frequently transferred programme was the Duluth model from the United States. Five programmes in the survey were transfers of this concept; however, many more programmes reported that aspects of the Duluth model had informed their programmes' designs.

Most respondents reported their organisation's services as being somewhat structured (43%) or structured to a great extent (48%). Only four programmes reported having little to no structure.

Offender Characteristics

Only 6 respondents indicated a specific age range of participants (30-39) for their programme. The vast majority selected 'Other', which indicates that programmes are not specified to certain age groupings. Data gathered on participants' nationality indicated that on average, 76% ($SD = 18$) of programme participants were local national residents.

Approximately two thirds (67%) of the programmes carried out general assessments of participants prior to assigning them to a programme. Only 33% of the programmes contained a systematic form of risk assessment despite the fact that 72% of the respondents reported tailoring their programmes to the risk level of reoffending. 63% of programmes that did not conduct systematic risk assessments were still seen as tailored to the participants' risk level of reoffending.

Programme Organisation

Most domestic violence perpetrator programmes took place in the community (85%); only 12% were delivered purely in custodial settings. While most programmes allowed participants to enrol in their service voluntarily (63%), the majority of recruitment came from mandatory referrals from various parts of the criminal justice system. Only 17% of the programmes relied on voluntary participation solely.

Table 3: *Characteristics of the programmes for domestic violence perpetrators*

Programme Component	Variables	Frequency	Percent	N
Treatment Modality†	Cognitive/Behavioural	48	91%	54
	Non-Behavioural	25	47%	54
	Systemic Therapy	22	42%	54
	Feminist	12	22%	53
Transfer	Not Transferred	35	66%	53
	Transferred	18	34%	
	Duluth Programme Transfer	5	10%	
Structure	Little to None	4	8%	53
	Somewhat	23	43%	
	To a Great Extent	26	49%	
Assessment†	General Assessment	36	67%	54
	Systematic Risk Assessment	18	33%	54
	Tailored to Risk	39	72%	54
Programme Location	Community	44	85%	52
	Custody	6	12%	
	Community and Custody	2	4%	
Recruitment	Voluntary	9	17%	52
	Coerced	10	19%	
	Mixed	33	63%	
Funding	Government	39	75%	53
	Client Fees	8	15%	
	Other	6	11%	
Accreditation	Accredited	20	39%	51
	Not Accredited	31	61%	

† percentages do not sum to 100, as respondents could indicate multiple responses

While only 17% of the programmes were directly run by governmental authorities, three-quarters received more than half of their funding from local, regional or national governments. 15% relied on participant contributions for the majority of their funding. 38% of programmes had been accredited by an accreditation board that was almost always a government-run or government-sponsored body.

Process and Outcome Evaluation

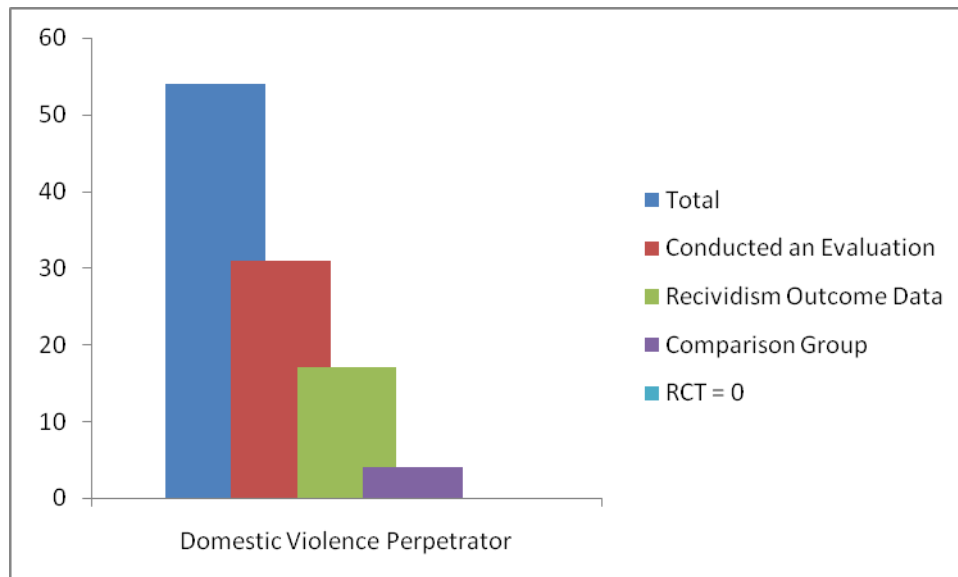
Measures were taken to ensure and maintain programme quality in 80% of the sample. The most common actions taken to address treatment integrity were staff supervision (48%), training (28%) and feedback (24%). While most programme administrators seemed to care for maintaining programme quality, fewer kept records of basic measurable data. For example, only 31% gathered data on participant attendance and 44% recorded programme completion.

58% of respondents stated that there was some form of outcome evaluation. Of these evaluated programmes, 55% had used recidivism as its outcome measure and 24% of these programmes applied a comparison group design. Therefore, of the 54 programmes surveyed only 7% adhered to a reasonable methodological standard (see Figure 6). None of the comparison group evaluations used a RCT design.

The responses to open questions showed that nearly all our informants were committed to carrying out high quality programmes that would reduce reoffending. However, they mentioned a number of challenges in executing their programmes. Respondents frequently cited problems of funding, organisational difficulties, lack of participants and

offender difficulties or disorders as critical issues in programme implementation. Nobody mentioned the lack of evaluation as an obstacle to success.

Figure 6: *Findings on outcome evaluation on domestic violence perpetrator programmes*



Discussion

Though the sample from the domestic violence survey was smaller than in the other two surveys of the STARR project ($N = 54$ versus 112 and 84), it gathered new and rich data. Thus far, there have been only two surveys which addressed domestic violence perpetrator programmes: An international survey conducted by the World Health Organisation (WHO) in 2003 (Rothman, Butchart and Cerdá) located 20 programmes within Europe. The Daphne II Work with Perpetrator Survey (2008) gathered data on more programmes, but did not collect information on implementation details, process and outcome evaluations. As in our other two surveys one must be aware of the limits of representativeness and potential response biases (see above).

There is some American evidence that perpetrator programmes have modest effects on reducing reoffending (Babcock et al. 2004; Davis and Taylor, 1999; Feder et al. 2008). For example, Feder et al., reviewed 10 randomised controlled trials on court mandated offender programmes, and found that for official outcome data (e.g., arrest or reconviction) reoffending declined by 13%. However, there was no significant effect in victim reported outcome. The Anglo-American research contains some promising findings for cognitive-behavioural programmes and the educational feminist psycho-educational Duluth model, but there are little specifics on treatment structure, participant assessment and other details. The evidence is much narrower in Europe where our systematic STARR review could not find any methodologically robust studies to demonstrate desirable effects on reoffending (Akoensi et al., 2011; see Appendix C). Therefore, we cannot put our survey findings in the context of systematic evidence regarding what works best for whom.

Despite the lack of clear evidence on what makes an effective treatment for domestic batterers, European practice seems to adhere to the most promising approaches. While fewer respondents indicated that cognitive-behavioural *theory* was their main foundation (47%), cognitive-behavioural techniques have become preferred forms of treatment (70%). Many programmes also seem to have integrated both psycho-educational and cognitive-behavioural methods into their practice. Because programmes include both treatment modalities, it becomes more difficult to determine what type of treatment works best. Nearly half the programmes stated that they were structured to a great extent. This may be due to the prevalence of cognitive-behavioural methods in the sample, but a social desirability bias could also have an influence.

Only 22% of the programmes surveyed were explicitly feminist, though many more cited that the gender-based power and control was a core aspect of their practice. Interestingly, all respondents who reported feminist programmes said that these

interventions have cognitive-behavioural theoretical foundations. While feminist cognitive-behavioural programmes have become more prevalent, feminist theory has typically been associated with psycho-educational types of treatment like the Duluth model (Day, Chung and O'Leary, 2009). Possible implications of this finding are that these informants either over-stated their preference for cognitive-behavioural theory, or have begun to perceive domestic violence as a learned behaviour that can be changed rather than just recognised and controlled, which tends to contradict feminist theory.

Programme transfer is a particularly salient issue among programmes for domestic violence perpetrators because cultural and legal frameworks differ across Europe. In our survey we found approximately one third of European programmes to be adaptations from another country. The American Duluth model is the most transferred programme in the survey. Only one of the transferred programmes was evaluated at its new site, but again the level of rigour of this evaluation was low. There are potential implications to programme effectiveness when a transferred programme is not evaluated in its location, even if it was successful in its original context. For example, the cultural context may require that significant adaptations need to be made which would result in a rather different programme. There may also be problems in taking the programme to scale, especially without developer involvement (Welsh et al. 2009). In addition, the risk factors targeted in the new location may be different from those in the other culture (e.g., when wife battering is more culturally accepted). Programme transfer has the potential to disseminate effective practices; however, if these transfers are not evaluated in their new locations we cannot know whether we are spreading good ideas or potentially doing harm (McCord, 2003).

There is some evidence that certain types of offenders tend to do better in treatment programmes than others (Feder, Wilson and Austin, 2008). Cavanaugh and Gelles (2005) suggest that there are different types of domestic violence perpetrators that need to be

identified and addressed in the formulation of treatment. Therefore, a certain amount of participant assessment and programme tailoring would probably improve the efficacy of perpetrator programmes. With regard to this issue our survey shows that most programme administrators seem to acknowledge the need to avoid ‘one size fits all’ programmes and aim to tailor their programmes to the participants’ risk of reoffending. However, far fewer programmes contained systematic assessment procedures. There was also no significant correlation between carrying out systematic assessments of risk and programme tailoring to risk. It is likely that those tailoring their programmes often base this on professional judgment, which is not problematic in itself, but without recording this data one cannot determine whether the treatment was appropriately differentiated.

Though most participation in the surveyed perpetrator programmes was compulsory (ordered by a branch of the criminal justice system), very few programmes are explicitly run by criminal justice authorities. Most programmes in the survey received funding and referrals from governmental bodies but operated independently of the criminal justice system. This relationship likely is a result of the third sector’s established role in perpetrator programmes. Private or NGO perpetrator programmes are also the norm in America; however, the US programmes receive significantly less government funding, and instead rely more on participant fees (Dalton, 2008). The dominance of government-funded programmes in Europe provides a direct and tractable route for EU states to set good practice standards and monitor programming. However, given that only 38% of programmes in the survey were somehow accredited and that only four programmes were evaluated using a (non-equivalent) comparison group, few governments seem to capitalise on this fiscal relationship to promote best practice.

Our survey, as well as our systematic review, shows that both process and outcome evaluation of domestic violence perpetrator programmes in Europe is underdeveloped.

Despite three decades of advocacy that evaluation is critical, there is not much improvement in European evaluation practice. Without more systematic evaluation, programmes run the risk of not only being ineffective, but could be harmful to participants or their family victims.

While 80% of respondents indicated their commitment to ensuring treatment integrity, far fewer programmes gathered basic systematic data on this issue. Even the most commonly cited methods of maintaining programme integrity, namely supervision and staff training, are only applied in 45% and 30% of programmes, respectively. The percentages of collected data on attendance, completion rates, and satisfaction data are equally low. Regular information on such issues is necessary to determine what is working in a programme and what needs to be improved.

As mentioned above, the current state of outcome evaluation of domestic perpetrator programmes in Europe is in poor condition. We found less than a handful of evaluations with untreated comparison groups and recidivism outcome data and no randomised experiments or studies with equivalent control groups. Simple before-after or other weak quasi-experimental designs cannot be properly interpreted because they may suffer from many threats to internal validity (Shadish et al., 2002). For example, domestic violence perpetrators tend to show positive behaviour during and shortly after treatment, but often fall back into violence when they are less under official supervision.

Our survey suggests that large parts of European practice in our field are not yet fully aware of the need for establishing a more solid evidence-base. Nobody mentioned evaluation as a problem in the qualitative open questions. Over half of the respondents stated that there was an outcome evaluation of their programme; however, on closer inspection nearly all those evaluations used only outcome variables such as participant satisfaction or

psychological measures. Although one must acknowledge serious practical obstacles to well-controlled evaluation designs (Lösel, 2007), it is important that more programmes gather outcome indicators such as official, self- and victim-reported recidivism and compare treated perpetrators with appropriate untreated groups (and not dropouts or other non-equivalent groups). Government bodies, particularly those who fund domestic violence perpetrator programmes, should provide an organisational framework, resources and assistance to improve evaluation in practice. Such a development would not only serve the victims and offenders, but also lead to longer-term financial benefits for the society.

Policy Implications of the Findings in the Three Surveys

Because of the differences between European countries in legal and cultural frameworks, criminal justice organisation, resources and many other issues, any political and practical consequences from our findings must be locally developed. However, our surveys enable some more general recommendations to be made that can be adapted across Europe.

Programmes that employ cognitive-behavioural interventions and adhere to the principles of Risk, Need, and Responsivity have produced greater reductions in reoffending in young offenders. Although the evidence base for substance abusing offenders is less developed, we found that primarily pharmacological programmes appear to be consistently effective. In contrast there is no clear evidence of what works best for domestic violence perpetrators. Although many of our informants seem to be aware of basic evidence on ‘what works’, this is not yet sufficiently realised in practice. A central feature of the disparity between existing practice and best practice is the rarity and low-quality of offender rehabilitation programme evaluation. This issue is particularly salient in transferred programmes, as these are very rarely evaluated in the new context.

Therefore, the most important conclusions and recommendations from our surveys address systematic programme evaluation. Improving rehabilitation practices requires the inculcation of ‘evaluative sensibility’ by both criminal justice authorities and practitioners. Evaluation assists not only in determining whether current practice is in fact best practice, but also helps to develop an understanding of *what else* contributes to effective practice beyond treatment modality.

In our qualitative questions respondents reported that obstacles to developing better evaluation practices stem from three key problems: (1) a lack of funding, (2) a lack of resources and (3) a lack of incentive.

1. Lack of Funding

Most programmes rely on government funds to support their services. Funding bodies — particularly governmental institutions — should provide more dedicated funding to accommodate evidence-based evaluations. Since evidence shows that effective interventions are highly cost-efficient (Aos et al., 2001), this strategy should be seen as a means of most efficient resource allocation and decreasing the costs in the longer term.

2. Lack of Resources

Practitioner evaluation should be facilitated through knowledge sharing and resource assistance. Following the model of medicine, centres of excellence could provide the knowledge and tools needed to evaluate programmes in routine practice. Given the prevalence of transferred programmes in the surveys, a relatively small European centre of excellence could be highly cost-efficient in coordinating national and international efforts.

3. Lack of Incentive

Practitioners often have some reservation toward methodologically sound evaluation strategies and partially fear negative consequences in case of undesirable outcomes. They do not yet view objective evaluation as a necessary feature of effective practice. Because evaluation cannot simply be delegated into practice, governments, other funding bodies and accreditation or similar expert panels should encourage 'evaluative sensibility' by providing incentives for those programmes that develop a sound evidence-base.

Although addressing these basic issues will require time, some resources and, in particular, the commitment of governments, practitioners and researchers, it is a milestone for a further improvement of rehabilitation practices to decrease recidivism in Europe.

Bibliography

§ denotes the study was included in the systematic review and meta-analysis of young offender treatment programs

- § Andrée Löfholm C, Olsson T, Sundell K, et al. (2009) Multisystemic therapy with conduct-disordered young people: Stability of treatment outcomes two years after intake. *Evidence and Policy*: 5(4): 373-397.
- § Curran D, Kilpatrick R, Young V, et al. (1995) Longitudinal aspects of reconviction: Secure and open intervention with juvenile offenders in Northern Ireland. *The Howard Journal of Criminal Justice* 24(2): 97-123.
- § Farrington D, Ditchfield J, Hancock G, et al. (2002) *Evaluation of Two Intensive Regimes for Young Offenders*. Home Office Research Study 239. London, UK: Home Office.
- § Kruissink M (1990) The Halt program: Diversion of juvenile vandals. *Dutch Penal Law and Policy: Notes on Criminological Research from the Research and Documentation Centre* The Hague, Netherlands: Ministry of Justice.
- § Kury H (1989) Treatment of young remand prisoners: Problems and results of a research project. In: Wegener H, Lösel F, and Haisch J (eds) *Criminal Behavior and the Justice System: Psychological Perspectives*. New York: Springer, 356-381.
- § Little M, Kogan J, Bullock R, et al. (2004) ISSP: An experiment in multi-systemic responses to persistent young offenders known to Children's Services. *British Journal of Criminology* 44(2): 225-240.
- § Lobley D, Smith D (2007) *Persistent Young Offenders: An Evaluation of Two Projects*. Aldershot, UK: Ashgate.
- § Lösel F, Pomplun O (1998) *Jugendhilfe statt Untersuchungshaft: Eine Evaluationsstudie zur Heimunterbringung*. Studien und Materialien zum Straf- und Massregelvollzug, vol. 7. Pfaffenweiler, Germany: Centaurus.
- § McMurran M, Boyle M, (1990) Evaluation of a self-help manual for young offenders who drink: A pilot study. *British Journal of Clinical Psychology* 29(1): 117-119.
- § Mitchell J, Palmer E, (2004) Evaluating the 'Reasoning and Rehabilitation' program for young offenders. *Journal of Offender Rehabilitation* 39(4): 31-45.
- § Newburn T, Shiner M (2005) *Dealing with Disaffection: Young People, Mentoring and Social Inclusion*. Cullompton, UK: Willan.

- § Ogden T, Hagen K, (2006) Multisystemic treatment of serious behaviour problems in youth: sustainability of effectiveness two years after intake. *Child and Adolescent Mental Health* 11(3): 142-149.
- § Ogden T, Hagen K, and Andersen O (2007) Sustainability of the effectiveness of a programme of Multisystemic Treatment (MST) across participant groups in the second year of operation. *Journal of Children's Services* 2(3): 4-14.
- § Raynor P, Vanstone M (1997) *Straight Thinking on Probation (STOP), The Mid Glamorgan experiment*. Probation Studies Unit Report, 4. Oxford, UK: Centre for Criminological Research.
- § Scholte E, Smit M, (1988) Early social assistance for juveniles at risk. *International Journal of Offender Therapy and Comparative Criminology* 32(3): 209-218.
- § Shapland J, Atkinson A, Atkinson H, et al. (2008) *Does Restorative Justice Affect Reconviction? The Fourth Report from the Evaluation of Three Schemes*. Ministry of Justice Research Series, 10/08. London, UK: Home Office.
- § Slot N (1983) The implementation and evaluation of a residential social skills training program for youth in trouble. In: Everaerd W, Hindley C, Bot A, et al. (eds) *Development in Adolescence: Psychological, Social, and Biological Aspects*. Amsterdam, Netherlands: Marinus Nijhoff, 192-205.
- § Slot N, Bartels A (1983) Outpatient social skills training for youth in trouble: Theoretical background, practice and outcome. In: Everaerd W, Hindley C, Bot A, et al (eds) *Development in Adolescence: Psychological, Social, and Biological Aspects*. Amsterdam, Netherlands: Marinus Nijhoff, 176-191.
- § St. James-Roberts I, Greenlaw G, Simon A, et al. (2005) *National Evaluation of Youth Justice Board Mentoring Schemes 2001 to 2004*. London, UK: Youth Justice Board.
- § Sundell K, Hansson K, Löfholm CA et al. (2008) The transportability of multisystemic therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology* 22(4): 550-560.
- § Bottoms AE (1995) *Intensive Community Supervision for Young Offenders: Outcomes, Process and Cost*. Cambridge, UK: Institute of Criminology.
- § Cann J, Falshaw L, and Friendship C (2005) Understanding 'What Works': accredited cognitive skills programmes for young offenders. *Youth Justice* 5(3): 165-179.

† denotes the study was included in the systematic review and meta-analysis of drug abusing offender treatment programs

- † Hartnoll RL, Mitcheson MC, Battersby A, et al. (1980) Evaluation of heroin maintenance in controlled trial. *Archive of General Psychiatry* 37: 877-884.

- † Lobmaier PP, Kunøe N, Gossop M, et al. (2010) Naltrexone implants compared to methadone: Outcomes six months after prison release. *European Addiction Research* 16(3): 139-145.
- † Löbmann R, Verthein U, (2009) Explaining the effectiveness of heroin-assisted treatment on crime reductions. *Law and Human Behaviour* 33(1): 83-95.
- † March JC, Oviedo-Joekes E, Perea-Milla E, et al. (2006) Controlled trial of prescribed heroin in the treatment of opioid addiction. *Journal of Substance Abuse Treatment* 31(2): 203-211.
- † Martin C, Player E, and Liriano S (2003) Results of evaluations of the RAPt drug treatment evaluation programme. In: Ramsay M (ed.) *Prisoners' Drug Use and Treatment: Seven Research Studies*. Home Office Research Study 267. London: Home Office, 97-112.
- † McCusker C, Davies M, (1996) Prescribing drug of choice to illicit heroin users: The experience of a UK community drug team. *Journal of Substance Abuse Treatment* 13(6): 521-531.
- † McSweeney T (2009) *The impact of 'quasi-compulsory' drug treatment in one English region*. Presented at the 10th Meeting of the Expert Forum on Criminal Justice, 10th-11th September. Strasbourg, France. Accessed July 11th, 2011:
<http://www.kcl.ac.uk/content/1/c6/01/22/42/DIPandRe-Offending.pdf>
- † Metrebian N, Shanahan W, Stimson GV, et al. (2001) Prescribing drug of choice to opiate dependent drug users: A comparison of clients receiving heroin with those receiving injectable methadone at a West London drug clinic. *Drug and Alcohol Review* 20(3): 267-276.
- † Naeem F, Bhatti F, Pickering R, et al. (2007) A controlled trial of the effectiveness of drug treatment and testing orders (DTTO) with standard care. *Journal of Substance Use* 12(4): 253-265.
- † Perneger TV, Giner F, del Rio M, et al. (1998) Randomised trial of heroin maintenance programme for addicts who fail in conventional drug treatments. *British Medical Journal* 317: 13-18.
- † Robertson JR, Raab GM, Bruce M, et al. (2006) Addressing the efficacy of dihydrocodeine versus methadone as an alternative maintenance treatment for opiate dependence: A randomized controlled trial. *Addiction* 101(12): 1752-1759.
- † Strang J, Marsden J, Cummins M, et al. (2000) Randomized trial of supervised injectable versus oral methadone maintenance: Report of feasibility and 6-month outcome. *Addiction* 95(11): 1631-1645.
- † Van den Brink W, Hendriks VM, Blanken P, et al. (2003) Medical prescription of heroin to treatment resistant heroin addicts: Two randomised controlled trials. *British Medical Journal* 327: 310-315.

*** denotes the study was included in the systematic review of domestic violence perpetrator programs**

- * Adva (2008) *Adva Against Domestic Violence and Abuse in Devon: The Repair Programme*. Unpublished report.
- * Association for the Prevention and Handling of Violence in the Family (APHVF) (2009) *Love without Hurt*. Unpublished report.
- * Bowen E (2004). *Evaluation of a Community Based Domestic Violence Offender Rehabilitation Programme*. Unpublished doctoral dissertation, Centre for Forensic and Family Psychology. University of Birmingham, UK.
- * Dobash RP, Dobash RE, Cavanagh K. et al. (1999) A research evaluation of British programmes for violent men. *Journal of Social Policy* 28(2): 205-233.
- * Echaury Tijeras J (2010) *Eficacia de un Programa de Tratamiento Psicológico Para Hombres Maltratadores en la Relación de Pareja: Características Psicopatológicas y Resultados Terapéuticos*. Unpublished doctoral dissertation, Departamento de Psicología y Pedagogía, Universidad Pública de Navarra. Pamplona-Iruña, Spain.
- * Echeburúa E, Fernández-Moltavo J, (1997) Tratamiento cognitivo-conductual de hombres violentos en el hogar: Un estudio piloto. *Análisis y Modificación de Conducta* 23(89): 355-384.
- * Echeburúa E, Fernández-Moltavo J, (2009) Evaluación de un programa de tratamiento en prisión de hombres condenados por violencia grave contra la pareja. *International Journal of Clinical and Health Psychology* 9(1): 5-20.
- * Echeburúa E, Fernández-Moltavo J, and Amor P (2006) Psychological treatment of men convicted of gender violence: A pilot study in Spanish prisons. *International Journal of Offender Therapy and Comparative Criminology* 50(1): 57-70.
- * Hagemann-White C, Kavemann B, and Beckmann S (2004) Täterarbeit im Kontext von Interventionsprojekten gegen häusliche Gewalt: Wissenschaftliche Begleitung der Interventionsprojekte gegen häusliche Gewalt (WiBiG), Abschlussbericht 2000 bis 2004. In: Beckmann S, and Hagemann-White C (eds) *Gemeinsam gegen häusliche Gewalt: Kooperation, Intervention, Begleitforschung*. Berlin: Bundesministerium für Familie, Senioren, Frauen, und Jugend, 691-834.
- * Socialstyrelsen (2010) *Behandling av Män som Utövar Våld i Nära Relationer - en Utvärdering*. Stockholm: Socialstyrelsen.
- * Törmä S, Tuokkola, K (2009) *Jussi-työ: Miesten Perhe-ja Lähisuhdeväkivaltatyön Ulkoinen Arviointi*. Unpublished report.

Other works cited

- Akoensi T, Koehler JA, Sánchez de Ribera O & Lösel FA (2011) A systematic review of the effectiveness of domestic violence perpetrator programmes to reduce repeat abuse in Europe. *STARR Probation Final Report: Appendix C*.
- Alhabib S, Nur U, and Jones R (2009) Domestic violence against women: Systematic review of prevalence studies. *Journal of Family Violence* 25(4): 369-382.
- Andrews D, Zinger I, Hoge R, et al. (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28(3): 369-404.
- Andrews DA, Bonta J (2010) *The Psychology of Criminal Conduct*. (5th ed.) Newark, NJ: LexisNexis.
- Andrews DA, Bonta J, and Wormith S (2011) The Risk-Need-Responsivity (RNR) model: Does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior* 38: 735-755.
- Aos S, Miller M, and Drake E (2006) *Evidence-based Adult Corrections Programs: What Works and What Does Not*. Olympia, WA: Washington State Institute of Public Policy.
- Aos S, Phipps P, Barnoski R, et al. (2001) *The Comparative Costs and Benefits of Programs to Reduce Crime*, 4th ed. Washington State Institute for Public Policy. Accessed 27 June, 2011. <http://www.wsipp.wa.gov/rptfiles/costbenefit.pdf>
- Appelt B, Kaselitz V (2000) *Prevention of Domestic Violence Against Women: European Survey of Good Practice Models, WAVE Training Programme*. Austrian Women's Shelter Network. Vienna, Austria: WAVE Office.
- Archer J, (2006) Cross-cultural differences in physical aggression between partners: A social-role analysis. *Personality and Social Psychology Review* 10(2): 133-153.
- Armeliu B-Å, Andreassen TH (2007) *Cognitive-behavioural treatment for antisocial Behavior in youth in residential treatment*. The Cochrane Database of Systematic Reviews 4.
- Babcock JC, Green CE, and Robie C (2004) Does batterers' treatment work? A meta-analytic review of domestic violence treatment. *Clinical Psychology Review* 23(8): 1023-1053.
- Barnish M (2004) *Domestic Violence: A Literature Review*. London: HM Inspectorate of Probation.
- Bennett TH, Holloway K, (2009) The causal connection between drug misuse and crime. *British Journal of Criminology* 49: 513-531.

- Bennett TH, Holloway K, and Farrington DP (2008) The statistical association between drug misuse and crime: A meta-analysis. *Aggression and Violent Behavior* 13(2): 107-118.
- Berk RA, Freedman DA (2003) Statistical assumptions as empirical commitments. In: Blomberg TG, Cohen S (eds) *Punishment and Social Control: Essays in Honor of Sheldon Messinger*, 2nd ed. New York: de Gruyter, 235-254.
- Borenstein M, Hedges LV, Higgins JPT et al. (2009) *Introduction to Meta-Analysis*. Chichester, UK: Wiley.
- Bottoms A, Dignan J, (2004) Youth justice in Great Britain. In: Tonry M, Doob A (eds.) *Youth Crime and Youth Justice: Comparative and Cross-National Perspectives. Crime and Justice: A Review of Research* vol. 31. Chicago: University of Chicago, 21-183.
- Bowen E, Brown L, and Gilchrist E (2002) Evaluating probation based offender programmes for domestic violence perpetrators: A pro-feminist approach. *The Howard Journal* 41(3): 221-236.
- Campbell JC, (2002) Health consequences of intimate partner violence. *Lancet* 359(9314): 1331-1336.
- Carroll KM, (1996) Relapse prevention as a psychosocial treatment: A review of controlled clinical trials. *Experimental and Clinical Psychopharmacology* 4(1): 46-54.
- Cavanaugh, MM, Gelles RJ, (2005) The utility of male domestic violence offender typologies: New directions for research, policy, and practice. *Journal of Interpersonal Violence* 20(2): 155-166.
- Cleland C, Pearson F, Lipton D, et al. (1997) Does age make a difference? A meta-analytic approach to reductions in criminal offending for juveniles and adults. Presented at the Annual Meeting of the American Society of Criminology, San Diego.
- Cook TD, Campbell DT (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Council of Europe (2008) *Combating Violence Against Women: Towards a Council of Europe Convention*. Report of the Committee on Equal Opportunities for Women and Men, Doc. 11702. Accessed August 14th, 2011:
<http://www.assembly.coe.int/Documents/WorkingDocs/Doc08/EDOC11702.pdf>
- Dalton B, (2007) What's going on out there? A survey of batterer intervention programs. *Journal of Aggression, Maltreatment & Trauma* 15(1): 59-74.
- Daly JE, Pelowski S, (2000) Predictors of dropout among men who batter: A review of studies with implications for research and practice. *Violence and Victims* 15(2): 137-160.

- Davis RC, Taylor BG, (1999) Does batterer treatment reduce violence? A synthesis of the literature. *Women & Criminal Justice* 10(2): 69-93.
- Day A, Chung D, and O'Leary P (2009) Programmes for men who perpetrate domestic violence: An examination of the issues underlying the effectiveness of intervention programmes. *Journal of Family Violence* 24(3): 203-212.
- Doob AN, Tonry M (2004) Varieties of youth justice. In: Tonry M, Doob A, (eds.) *Youth Crime and Youth Justice: Comparative and Cross-National Perspectives. Crime and Justice: A Review of Research*. vol. 31. Chicago: University of Chicago, 1-20.
- Dowden C, Andrews D, (2003) Does family intervention work for delinquents? Results of a meta-analysis. *Canadian Journal of Criminology and Criminal Justice* 45(3): 327-342.
- Dowden C, Andrews D, (1999) What works in young offender treatment: A meta-analysis. *Forum on Corrections Research* 11(2): 21-24.
- Dowden C, Antonowicz D and Andrews DA (2003) The effectiveness of relapse prevention with offenders: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology* 47(5): 516-528.
- Durlauf SN, Nagin D, (2011) Imprisonment and crime: Can both be reduced? *Criminology and Public Policy* 10(1): 13-54.
- Egg R (1992) Therapy versus penalty: An evaluation study. In: Lösel FA, Bender D and Bliesener T (eds) *Psychology and Law: International Perspectives*. New York: de Gruyter, 174-181.
- Egli N, Pina M, Christensen PS, et al. (2009) *Effects of Drug Substitution Programs on Offending among Drug Addicts*. Campbell Systematic Reviews 3.
- Eisner M, (2009) No effects in independent prevention trials: can we reject the cynical view? *Journal of Experimental Criminology* 5:163-183.
- EMCDDA (2001) *Reviewing Current Practice in Drug-Substitution Treatment in the European Union*. Luxembourg: Office for Official Publications of the European Communities. Accessed Aug 15, 2011.
<http://www.emcdda.europa.eu/publications/insights/substitution-treatment>
- EMCDDA (2011) Examples of evaluated practices: EDDRA-Interventions in the Criminal Justice System. EMCDDA. Accessed August 9, 2011
<http://www.emcdda.europa.eu/html.cfm/index52006EN.html?by=262&value=829>
- Eysenck H, (1995) Meta-analysis squared. Does it make sense? *American Psychologist* 50(2): 110-111.
- Farrington DP, (2003) Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Science* 587(1): 49-68.

- Fazel S, Bains P, and Doll H (2006) Substance abuse and dependence in prisoners: A systematic review. *Addiction* 101(2): 181-191.
- Feder L, Wilson DB, and Austin S (2008) *Court-mandated interventions for individuals convicted of domestic violence*. Campbell Collaboration Systematic Reviews 12.
- Fernández-Montalvo J, López-Goñi JJ, Illescas C, et al. (2008) Evaluation of a therapeutic community treatment program: A long-term follow-up study in Spain. *Substance Use and Misuse* 43(10): 1362-1377.
- Garner J, Fagan, J, and Maxwell C (1995) Published findings from the Spouse Assault Replication Program: A critical review. *Journal of Quantitative Criminology* 11(1): 3-28.
- Gensheimer L, Mayer J, Gottschalk R, et al. (1986) Diverting youth from the juvenile justice system: A meta-analysis of intervention efficacy. In: Apter S, Goldstein A, (eds) *Youth Violence: Programmes and Prospects*. Elmsford: Pergamon, 39-57.
- Golding JM (1999) Intimate partner violence as a risk factor for mental disorders: A meta-analysis. *Journal of Family Violence* 14(2): 99-110.
- Gondolf E (2002) *Batterer Intervention Systems: Issues, Outcomes, and Recommendations*. Thousand Oaks, CA: Sage.
- Gossop M, Trakada K, Stewart D, et al. (2006) *Levels of Conviction Following Drug Treatment: Linking Data from the National Treatment Outcome Research Study and the Offenders Index*. Home Office Research Findings 275. London: Home Office.
- Gottschalk R, Davidson II W, Gensheimer L, et al. (1987) Community-based interventions. In: Quay H (ed.) *Handbook of Juvenile Delinquency*. New York: Wiley, 266-289.
- Gracia E, Herrero J, (2006) Acceptability of domestic violence against women in the European Union: A multilevel analysis. *Journal of Epidemiology and Community Health* 60(2): 123-129.
- Graham-Kevan N, (2007) Domestic violence: Research and implications for batterer programmes in Europe. *European Journal on Criminal Policy and Research* 13(3): 213-225.
- Hamberger KL, Hastings JE, (1993) Court mandated treatment of men who assault their partner: Issues, controversies, outcomes. In: N.Z. Hilton, (ed.) *Legal Responses to Wife Assault: Current Trends and Evaluation*. London: Sage, 188-229.
- Hamilton, L., Koehler, J.A., and Lösel, F.A. (2011) Programmes to reduce reoffending throughout Europe: Three surveys on current practice. *STARR Probation Final Report: Appendix D*.

- Hammersley R, (2011) Pathways through drugs and crime: Desistance, trauma, and resilience. *Journal of Criminal Justice* 39(3): 268-272.
- Hanson RK, Bourgon G, Helmus L, et al. (2009) The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior* 36(9): 865-891.
- Hanson RK, Gordon A, Harris AJR, et al. (2002) First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment* 14(2): 169-194.
- Harper G. and C. Chitty (eds.) (2005) *The Impact of Corrections on Reoffending: A Review of 'What Works'*, 2nd ed. Home Office Research Study 291. London: Home Office.
- Hedges LV, (1981) Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6(2): 107-128.
- Higgins J, Thompson SG, Deeks JJ, et al. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal* 327:557-560.
- Hollin CR (2002) Risk-needs assessment and allocation to offender programmes. In: McGuire J, (eds) *Offender Rehabilitation and Treatment: Effective programmes and Policies to Reduce Reoffending*. Chichester, UK: Wiley, 309-332.
- Hollin CR, (2008) Evaluating offending behaviour programmes: Does only randomization glister? *Criminology and Criminal Justice* 8: 89-106.
- Hollin C, Palmer E.J, (2009) Cognitive skills programmes for offenders. *Psychology Crime & Law* 15: 147-164.
- Holloway K, Bennett TH (2004) *The Results of the First Two Years of the NEW-ADAM Programme*. Home Office Online Report 19/04. Accessed July 12th, 2011. <http://library.npia.police.uk/docs/hordsolr/rdsolr1904.pdf>
- Holloway K, Bennett TH, and Farrington DP (2008) *Effectiveness of Treatment in Reducing Drug-Related Crime*. Stockholm, Sweden: Swedish National Council for Crime Prevention (Brottsförebyggande rådet – Brå).
- Hughes C, Stevens A (2007) *The Effects of Decriminalization of Drug Use in Portugal*. Briefing Paper 14. The Beckley Foundation Drug Policy Programme. Accessed Aug 15, 2011: <http://kar.kent.ac.uk/13325/>
- Irvin JE, Bowers CA, Dunn ME, et al. (1999) Efficacy of relapse prevention: A meta-analytic review. *Journal of Consulting and Clinical Psychology* 67(4): 563-570.
- Janson, C-G (2004) Youth justice in Sweden. In: Tonry M, Doob A (eds) *Youth Crime and Youth Justice: Comparative and Cross-National Perspectives*. *Crime and Justice: A Review of Research* vol. 31. Chicago: University of Chicago, 391-441.

- Jolliffe D, Farrington D (2008) *The Influence of Mentoring on Reoffending*. Stockholm, Sweden: Swedish National Council for Crime Prevention (Brottsförebyggande rådet – Brå).
- Kleiman MAR, Caulkins JP, and Hawken A (2011) *Drugs and Drug Policy: What Everyone Needs to Know*. New York: Oxford University Press.
- Koehler JA, Lösel FA, Akoensi TD and Humphreys DK (2011) A systematic review and meta-analysis on the effects of young offender treatment programmes in Europe *STARR Probation Final Report: Appendix A*.
- Koehler, JA, Humphreys DK, Akoensi TD, Sánchez de Ribera O, and Lösel, FA (2011) A systematic review and meta-analysis of the effectiveness of drug treatment programmes to reduce reoffending in Europe. *STARR Probation Final Report: Appendix B*.
- Landenberger NA, Lipsey MW (2005) The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology* 1(4): 451-476.
- Latimer J, Dowden C, and Morton-Bourgon K (2003) *Treating Youth in Conflict with the Law: A New Meta-Analysis*. Rep. RR03YJ-3e. Ottawa, Canada: Department of Justice.
- Latimer J, Morton-Bourgon K, and Chrétien J-A (2006) *A Meta-Analytic Examination of Drug Treatment Courts: Do They Reduce Recidivism?* Ottawa, Canada: Canadian Department of Justice, Research & Statistics Division.
- Lipsey M (1992) Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In: Cook T, Cooper H, Cordray D, et al. (eds) *Meta-Analysis for Explanation: A Casebook*. New York, NY: Russell Sage, 83-127.
- Lipsey M, (1999) Can intervention rehabilitation serious delinquents? *Annals of the American Academy of Political and Social Science* 564: 142-166.
- Lipsey M, (2003) Those confounded moderators in meta-Analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science* 587(1): 69-81.
- Lipsey M, Chapman G, and Landenberger N (2001) Cognitive-behavioral programs for offenders. *Annals of the American Academy of Political and Social Science* 578(1): 144-157.
- Lipsey M, Cullen F, (2007) The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science* 3: 297-320.
- Lipsey M, Petrie C, Weisburd D, et al. (2006) Improving Evaluation of Anti-Crime Programs: Summary of a National Research Council Report. *Journal of Experimental Criminology* 2(3): 271-307.

- Lipsey MW, Wilson DB (1998) Effective intervention for serious juvenile offenders. In: Lober R, Farrington DP (eds) *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*. London: Sage.
- Lipsey MW, Wilson DB (2001) *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- Lloyd C, Mair G, and Hough M (1994) *Explaining Reconviction Rates: A Critical Analysis*. Home Office Research Study 136. London, UK: HMSO.
- Lösel F (1995) The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In: McGuire J (ed.) *What Works: Reducing Reoffending. Guidelines from Research and Practice*. Chichester, UK: Wiley, 79-111.
- Lösel F (1996) Working with young offenders: The impact of meta-analyses. In: Hollin C, Howells K (eds) *Clinical approaches to working with young offenders* Chichester, UK: Wiley, 57-82.
- Lösel F (2000) The efficacy of sexual offender treatment: A review of German and international evaluations. In: van Koppen PJ, Roos NHM, (eds), *Rationality, Information and Progress in Psychology and Law*. Maastricht, NL: Metajuridica Publications, 145-170.
- Lösel F (2007) Doing evaluation in criminology: Balancing scientific and practical demands. In: King RD, Wincup E (eds) *Doing Research on Crime and Justice, 2nd ed.* Oxford, UK: Oxford University Press, 141-170.
- Lösel F (2011) What works in correctional treatment and rehabilitation for young adults? In: Lösel F, Bottoms AE and Farrington DP (eds) *Young Adult Offenders: Lost in Transition?* Milton Park, UK: Routledge (in press).
- Lösel F, Beelmann A, (2003) Effects of child skills training in preventing antisocial behaviour: A systematic review of randomized evaluations. *Annals of the American Academy of Political and Social Science* 587(1): 84-109.
- Lösel F, Egg, R (1997) Social-therapeutic institutions in Germany: Description and evaluation. In: Cullen E, Jones L and Woodward R (eds) *Therapeutic Communities in Prisons*. Chichester: John Wiley, 181-203.
- Lösel F, Köferl P (1989) Evaluation research on correctional treatment in West Germany: A meta-analysis. In: Wegener H, Lösel F, and Haisch J (eds) *Criminal Behavior and the Justice System*. New York: Springer, 334-355.
- Lösel F, Köferl P, and Weber F (1987) *Meta-Evaluation der Sozialtherapie*. Stuttgart: Enke.
- Lösel F, Schmucker M (2002) Assessor's biases. In: Fernandez-Ballesteros (Ed.) *Encyclopedia of Psychological Assessment vol. 1*. London: Sage, 98-101.

- Lösel F, Schmucker M, (2005) The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology* 1(1): 117-146.
- Lösel FA, Wittmann WW, (1989) The relationship of treatment integrity and intensity to outcome criteria. *New Directions for Program Evaluation* 42: 97-108.
- Lowenkamp CT, Holsinger AM, and Latessa EJ (2005) Are drug courts effective: A meta-analytic review. *Journal of Community Corrections* 28: 5-10.
- MacKenzie D (2006) *What Works in Corrections: Reducing the Criminal Activities of Offenders and Delinquents*. New York: Cambridge.
- MacKenzie DL, Brame R, McDowall D et al. (1995) Boot camp prisons and recidivism in eight states. *Criminology* 33: 327-358.
- Maguire M, Grubin D, Lösel F, et al. (2010) 'What works' and the correctional services accreditation panel: Taking stock from an inside perspective. *Criminology and Criminal Justice* 10: 37-58.
- Marsch LA, (1998) The efficacy of methadone maintenance interventions in reducing illicit opiate use, HIV risk behavior and criminality: A meta-analysis. *Addiction* 93(4): 515-532.
- Martinson R, (1974) What works?—Questions and answers about prison reform. *The Public Interest* 35: 22-54.
- McCord J, (2003) Cures that harm: unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Science* 587:16-30.
- McMurran M (2002) Motivation to change: selection criterion or treatment need? In: M. McMurran (Ed.) *Motivating Offenders to Change*. Chichester, UK: Wiley, 3-13.
- Mears D, (2003) Research and interventions to reduce domestic violence revictimization. *Trauma, Violence, and Abuse* 4(2): 127-147.
- Ministry of Justice (UK) (2010) *Compendium of Reoffending Statistics and Analysis*. Ministry of Justice Statistics Bulletin. London, UK: Ministry of Justice.
- Mitchell O, Wilson DB, and MacKenzie DL (2006) *The Effectiveness of Incarceration-Based Drug Treatment on Criminal Behavior*. Campbell Collaboration Systematic Reviews 11.
- National Treatment Agency (2009) *Breaking the Link: The Role of Drug Treatment in Tackling Crime*. London: National Health Service. Accessed July 11th, 2011: http://www.nta.nhs.uk/uploads/nta_criminaljustice_0809.pdf
- Pawson R, Tilley N (1997) *Realistic Evaluation*. Thousand Oaks, CA: Sage.

- Pearson FS, Lipton DS, (1999) A meta-analytic review of the effectiveness of corrections-based treatment for drug abuse. *The Prison Journal* 79(4): 384-410.
- Pearson FS, Lipton DS, Cleland CM, et al. (2002) The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime & Delinquency* 48: 476-496.
- Perry A, Coulton S, Glanville J, et al. (2008) *Interventions for Drug-Using Offenders in the Courts, Secure Establishments and the Community*. Cochrane Collaboration Reviews 3.
- 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.
- Pettersson T, Sundin-Osborne A, and Bishop N (1986) *Results of the Drug Misuser Treatment Programme at Österåker Prison*. Report 1986:2. Norrköping, Sweden: National Prison and Probation Administration.
- Prendergast ML, Podus D, Chang E, et al. (2002) The effectiveness of drug abuse treatment: A meta-analysis of comparison group studies. *Drug and Alcohol Dependence* 67: 53-72.
- Redondo S, Sánchez-Meca J, and Garrido V (1999) The influence of treatment programmes on the recidivism of juvenile and adult offenders: An European meta-analytic Review. *Psychology, Crime and Law* 5(3): 251-278.
- Redondo S, Sánchez-Meca J, and Garrido V (2001) Treatment of offenders and recidivism: assessment of the effectiveness of programmes applied in Europe. *Psychology in Spain* 5(1): 47-62.
- Rosenbaum A, (1988) Methodological issues in marital violence research. *Journal of Family Violence* 3(2): 91-104.
- Rosenfeld BD, (1992) Court-ordered treatment of spouse abuse. *Clinical Psychology Review* 12(2): 205-266.
- Rosenthal R (1991) *Meta-Analytic Procedures for Social Research*. Newbury, CA: Sage.
- Rothman EF, Butchart A, and Cerdá M (2003) *Intervening with Perpetrators of Intimate Partner Violence; a Global Perspective*. Geneva: World Health Organization.
- Roxburgh M, Donmall M, White M, and Jones A (2010) *Statistics from the National Drug Treatment Monitoring System, 1 April 2009-31 March 2010*. London: Department of Health. Accessed July 11th, 2011
<http://www.nta.nhs.uk/uploads/ndtmsannualreport2009-10finalversion.pdf>
- Scourfield JB, Dobash RP, (1999) Programmes for violent men: Recent developments in the U.K. *The Howard Journal* 38(2): 128-143.

- Shadish WR, Cook TD, and Campbell DT (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shepard M, (1992) Predicting batterer recidivism five years after intervention. *Journal of Family Violence* 7(3): 167-178.
- Sherman L, Farrington D, Welsh B, et al. (eds) (2002) *Evidence-Based Crime Prevention*. New York: Routledge.
- Sherman LW, Weisburd DL, (1995) General deterrent effects of police patrol in crime 'hot spots': A randomized, controlled trial. *Justice Quarterly* 12(4): 625-648.
- Stevens A (2011) *Drugs, Crime and Public Health: The Political Economy of Drug Policy*. London: Routledge.
- Stevens A, McSweeney T, van Ooyen M et al. (2005) Editorial: On coercion. *The International Journal of Drug Policy* 16: 207-209.
- Stevens A, Trace M and Bewley-Taylor D (2005) *Reducing Drug Related Crime: An Overview of the Global Evidence. Report 5*. The Beckley Foundation Drug Policy Programme.
- Stover CS, (2005) Domestic violence research: What have we learned and where do we go from here? *Journal of Interpersonal Violence* 20(4): 448-454.
- Straus M (1991) Conceptualization and measurement of battering: Implications for public policy. In: Steinman M (ed.) *Woman battering: Policy responses*. Cincinnati: Anderson, 19-47.
- Sullivan C, Rumpitz M, Campbell R, et al. (1996) Retaining participants in longitudinal community research. *Journal of Applied Behavioral Science* 32(3): 262-276.
- Tong LSJ, Farrington DP, (2006) How effective is the Reasoning and Rehabilitation programme in reducing offending? A meta-analysis of evaluations in four countries. *Psychology, Crime and Law* 12: 3-24.
- Tournier P, Barre M, (1990) Enquête sur les systèmes pénitentiaires dans les membres du Conseil de l'Europe: Démographie carcérale comparée. *Bulletin d'Information Pénitentiaire* 15: 4-44.
- Ttofi M, Farrington D, and Baldry A (2008) *Effectiveness of Programmes to Reduce School Bullying: A Systematic Review*. Stockholm, Sweden: Swedish National Council for Crime Prevention (Brottsförebyggande rådet – Brå).
- Uchtenhagen A (2002) *Drug abuse treatment in the prison milieu: A review of the evidence*. In: Conference proceedings of Prisons, Drugs and Society, 20th-22nd September 2001, Bern, Switzerland. Strasbourg: Council of Europe, 79-95.

- Uchtenhagen A, Schaaf S, Bock I, et al. (2006) *Quasi-Compulsory and Compulsory Treatment of Drug-Dependent Offenders in Europe: Final Report on Quantitative Evaluation*. Zurich: Research Institute for Public Health and Addiction, Zurich University.
- UK Drug Policy Commission (2008) *Reducing Drug Use, Reducing Reoffending: Are Programmes for Problem Drug-Using Offenders in the UK Supported by the Evidence?* London: UKDPC.
- United Nations (2009) *The UN Secretary General's database on violence against women: Country Pages*. United Nations Division for the Advancement of Women. New York: Accessed June 12, 2011. <http://webapps01.un.org/vawdatabase/country.action>
- United Nations (2008) *Good Practices in Legislation on Violence Against Women*. Report of the expert group meeting of the United Nations Division for the Advancement of Women, United Nations Office on Drugs and Crime, 26th-28th May, 2008. Vienna, Accessed August 4, 2011: [http://www.un.org/womenwatch/daw/egm/vaw_legislation_2008/Report%20EGMGPL_VAW%20\(final%2011.11.08\).pdf](http://www.un.org/womenwatch/daw/egm/vaw_legislation_2008/Report%20EGMGPL_VAW%20(final%2011.11.08).pdf)
- Ward T, Brown M, (2004) The Good Lives model and conceptual issues in offender rehabilitation. *Psychology, Crime and Law* 10(3): 243-257.
- Ward T, Maruna S (2007) *Rehabilitation*. New York: Routledge.
- Wartna BSJ, Nijssen LTJ (2006) *National Studies on Recidivism: An Inventory of Large-Scale Recidivism Research in 33 European Countries*. WODC-studies on recidivism: Fact sheet 2006-11. Netherlands: Justitie Research and Documentation Centre.
- Weermann F (2007) Juvenile offending. In: Tonry M, Bijleveld C, (eds) *Crime and Justice in the Netherlands. Crime and Justice: A Review of Research*, vol. 35. Chicago: University of Chicago, 261-318.
- Weis R, Whitemarsh S, and Wilson N (2005) Military-style residential treatment for disruptive adolescents: Effective for some girls, all girls, when, and why?" *Psychological Services* 2(1): 105-122.
- Weisburd D, Lum C, and Petrosino A (2001) Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science* 578(1): 50-70.
- Weisburd D, Petrosino AJ, and Mason G (1993) Design sensitivity in criminal justice experiments. In: Tonry M (ed) *Crime and Justice, A Review of Research* vol. 17. Chicago: University of Chicago Press, 337-379.
- Weisburd D, (2003) Ethical practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials. *Evaluation Review* 27(3): 336-354.

- Welsh B, Farrington D (2001) A review of research on the monetary value of preventing crime. In: Welsh B, Farrington D, and Sherman L (eds.) *Costs and Benefits of Preventing Crime*. Oxford, UK: Westview, 87-122.
- Welsh B, Farrington D (2002) *Crime Prevention Effects of Closed Circuit Television: A Systematic Review*. Home Office Research Study 252. London: Home Office.
- Welsh BC, Sullivan CJ, and Olds DL (2010) When Early Crime Prevention goes to scale: A New Look at the Evidence. *Prevention Science* 11: 115-125.
- Westmarland N, Kelly L, and Chalder-Mills J (2010) *What Counts as Success?* London: Respect.
- Wilson DB, (2001) Meta-analytic methods for criminology. *Annals of the American Academy of Political and Social Science* 578(1): 71-89.
- Wilson D, (2009) Missing a critical piece of the pie: Simple document search strategies inadequate for systematic reviews. *Journal of Experimental Criminology* 5(4): 429-440.
- Wilson D, Bouffard, L, and MacKenzie D (2005) A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior* 32(2): 172-204.
- Work with Perpetrators of Domestic Violence in Europe (2008) *Country summaries*. Dissens, Germany: Work with Perpetrators. Accessed July 15, 2011: http://www.work-with-perpetrators.eu/en/country_summaries.php
- Zarkin GA, Dunlap LJ, Bray JW, et al. (2002) The effect of treatment completion and length of stay on employment and crime in outpatient drug-free treatment. *Journal of Substance Abuse Treatment* 23(4): 261-271.