1. Background of the Review

By 2011 there are very few jurisdictions throughout the western world that do not have some form of electronic monitoring (EM) to supervise offenders. EM was developed to replace custody or imprisonment because the surveillance and control over offenders in the community is believed to prevent criminal activities, by reducing both their capacity and their opportunity to commit crimes (Mackenzie 2006:305). In addition, EM was intended to reduce costs and provide a cheaper alternative to custody and face-to-face supervision (Garland 2002). By Governments subcontracting EM out to the commercial sector, it was meant to allow crime control to be increased without increasing the costs of implementing it (Paterson 2007).

In the early days of electronic tagging, a bracelet was placed on the ankle or wrist of the offender. A signal was then sent from a base station from within the home of the offender. If the offender was not within a stipulated range from the base station then an alert signal would be sent to the firm monitoring the offender. More recent innovations in technology have led to voice recognition and the use of satellite tracking schemes (Black and Smith 2003).

EM has flourished in Great Britain with the Home Detention Curfew Scheme and the Intensive Surveillance and Supervision Programmes. Between 1999 and 2006, 137,000 people were placed on the Home Detention Curfew Scheme at a cost of £342 million pounds. (Shapps 2006). Likewise, by 1992 in the United States, EM was being used in 50 United States with approximately 7,000 offenders being monitored (Renzema 1992) and 15,000 EM projects running across the nation (Haverkamp, Mayer and Levy 2004).

Today, “electronic monitoring” has become a generic term which encompasses a range of different technologies, rather than a single kind of penal measure. These include radio frequency EM, GPS satellite tracking devices, voice recognition, etc. On the one hand, these are all meant to “monitor” offenders remotely. On the other hand, they impact on offenders’ lives in different ways, impose different kinds of restrictive regime, of greater or lesser degrees of intrusiveness. Therefore, these

---

1 We thank Anonymous Reviewers for the following insightful comments.
differences make it difficult to speak of a single effect of EM on recidivism and practically no “read-across” from one type of EM to another should be assumed (e.g., the impact of radio frequency EM overnight-only curfews for 3 months should not be assumed to be a comparable intervention to GPS tracking 24/7 for 6 months), and claims of effects of EM on recidivism ought to specify the way in which EM was being used in that scheme or study.

It was once customary to speak of “front door” and “back door” schemes of EM, positioning it in relation to imprisonment as either a) a means of keeping someone out of prison in the first place, or b) a means of post-release supervision, and possibly, a means of facilitating release earlier than would otherwise have been allowed. This distinction is too crude a way of conceptualizing EM. It is nowadays used, worldwide, in a range of legal contexts which may or may not have different expectations in terms of “recidivism”, and different impacts upon it. While similarities of consequence might well be expected, EM’s impact in one legal context cannot necessarily be “read across” to another legal context, as if the technology had an autonomous effect independent of the legal measure in which it was embedded:

- Community sentence - stand-alone or (more rarely) integrated with probation
- Conditional prison sentence
- Pre-trial detention
- Restraining orders (in domestic violence)
- Early release of short-term prisoners
- Temporary release from prison
- Parole of longer-term prisoners

One systematic review of the evidence was already conducted by Renzema and Mayo-Wilson (2005), covering evidence through 2002. The reviewers concluded that the available evidence was too limited to support any conclusions about the effectiveness of EM. Since then, however, many developments in both technology and research have emerged. Greater emphasis is now given to the possibilities that are likely to be linked to supervising offenders, especially sex offenders who are released back to the community and are subject to geographic supervision (e.g., staying away from schools or playgrounds). We know of at least one recent RCT in the field that can provide stronger evidence for a systematic exploration (Killias, Gillieron, Kissling and Villettaz 2010), and we suspect that since 2002 more RCTs were conducted on EM. This question has been addressed by the literature and we would like to review it in a systematic way.
1. OBJECTIVE OF THE REVIEW

This review has one objective: This is, to assess the magnitude of and direction of the effect of EM on recidivism rates among offenders, both while being monitored and following release from monitoring. We are also interested in knowing whether the effect varies by the type of offence or on the type of offender.

2. METHODS OF THE REVIEW

2.1 CRITERIA FOR CONSIDERING STUDIES FOR THE REVIEW

2.1.1 TYPES OF INTERVENTION

This systematic review will focus on electronic monitoring as a criminal justice intervention. However, there is no one operational definition for EM, as the technology as well as the intention behind using EM can change from one study to the next. The common theme will therefore be on a broad category of EM intervention, however the analysis, if possible, will also be categorized based on the technology:

2.1.1.1 Radio frequency EM used for curfews house arrest

2.1.1.2 Voice verification – used either to monitor/enforce house arrest, or presence at several different locations.

2.1.1.3 Remote Alcohol Monitoring: a) breathalyser-based, linked to house arrest and b) transdermal alcohol monitoring – can be used on a mobile offender

2.1.1.4 Global Positioning System (GPS) satellite tracking – retrospective and real-time, with or without designated exclusion zones

2.1.1.5 Inmate Monitoring - use of EM in prisons, to pinpoint individual locations within buildings and/or monitor perimeters

2.1.1.6 Victim protection (usually in cases of domestic violence) - the victim has a device in her home or, if she is mobile, on her person, which warns her of the offender’s proximity.
We should note that EM can be used in conjunction with other treatments, such as mandatory drug testing, periodic appearances at the local police station, or commitment to employment. Such studies will be included as well, however we will assess mixed treatments separately, in order to learn how EM interacts with other treatments.

2.1.2 TYPES OF PARTICIPANTS

Both male and female participants, juvenile and adult participants, from any country around the world, who had committed any type or number of criminal offenses, will be included in the review.

2.1.3 TYPES OF STUDIES

We will include the following study designs:

2.1.3.1 Randomized controlled trials/true experiments that randomly allocate participants to an EM condition and an alternative condition, such as traditional probation.

2.1.3.2 Quasi-experimental designs fail to randomly allocate participants to conditions. These designs must have the following features:

2.1.3.2.1 A comparison group that does not receive EM monitoring. This may be designed based on historical an historical comparison group design or that used offenders for a comparable jurisdiction that does not use EM. Excluded comparison groups will include those that use offenders who refuse EM or those who were ineligible for EM.

2.1.3.2.2 Baseline assessment of the comparability of the EM and comparison condition.

2.1.3.2.3 An estimate of the effect of EM that adjusts for baseline difference.
2.1.4 TYPES OF OUTCOME MEASURES

2.1.4.1 Studies must report at least one measure of recidivism, such as an arrest or a re-conviction for a new offence.

We note, that and recognition given to the fact that “effect on recidivism” is not the sole basis on which EM has been introduced or evaluated, however we focus herein on this aspect only for the sake of this systematic review.

We further note that some outcomes may be measured at the pre-trial stage as well, in which case will analyse violations these conditions as well, separately (e.g., assess the available literature in three groups: pre-trial phase, sentencing, and following incarceration)

2.1.4.2 Outcomes may be measured either during the period of supervision or for a period following supervision, and will be measured separately.

2.2 SEARCH STRATEGY FOR IDENTIFICATION OF RELEVANT STUDIES

2.2.1 The search strategy will include the following sources:

2.2.1.1 Extensive search of online databases (see 4.4 below).

2.2.1.2 Searches of narrative and empirical reviews of literature that examined the effectiveness of EM interventions.

2.2.1.3 Registers of randomised controlled trials: the Registry of Randomized Experiments in Criminal Sanctions, 1950 – 1983 (Weisburd, Sherman and Petrosino 1990) and the Social Psychological, Educational and Criminological Trials Register (SPECTR) developed by the Cochrane Centre.
2.2.1.4 As studies will be located, their references will be examined for details on other relevant studies. These will then be examined with accompanying notes being made to explain where the document was originally cited.

2.2.2 Each title and abstract will then be screened to establish if it meets the criteria for this review.

2.2.3 An eligibility criterion will be completed in respect of all studies that will pass an initial screening and are retrieved for further examination as with 2.1.1 above. The checklist can be found in Appendix II.

2.2.4 Studies eligible according to 2.1.3 above will be carefully assessed using the Eligibility Criterion at Appendix II. The criterion will be completed for each of the studies to determine a host of relevant questions, such as the type of programme involved, the sanction imposed on the control group if any at all, who delivered the treatment/intervention, the length of time involved and details regarding the comparison groups and the similarities between them, etc.

2.2.5 No limitations are made on the nature of publication (i.e., published or and unpublished material and ‘gray literature’), year publication or language of publication

2.3 CODING OF STUDIES

2.3.1 The two independent reviewers will extract information from hard copies of eligible studies using the eligibility criterion at Appendix II. If both reviewers agree on the rating results of each article and the coding of the data from each article, then the data will be entered into Comprehensive Meta Analysis 2.0, mutatis mutandis. The Coding Protocol is contained at Appendix I.

2.4 SEARCH TERMS

2.4.1 Cursory review of the databases reveals that there is a very large body of evidence on electronic monitoring in the biomedical discipline, particularly
around compliance or adherence with medical care – which is an area outside the boundaries of this review.

2.4.2 We therefore restricted our search strategy to three sets of terms. The first related to the condition of interest “arrest”, “crime” and “offender”, the second related to the intervention i.e. “electronic monitoring” or “electronic tagging”. The final term related to the outcomes of the interventions i.e. “recidivism”, “reoffending” and “control group”.

2.4.3 We therefore developed three sets of keywords in the search: EM and crime-related data. Boolean combination of the term EM:

2.4.3.1 [“electronic monitor” OR “monitoring electronically” OR “EM”, OR “electronic tagging” OR “satellite monitoring system” OR “GPS System” OR “radio” OR “voice recognition technology” OR “intensive surveillance” OR “house arrest” OR “Remote alcohol monitoring”];

2.4.3.2 [“arrest” OR “rearrest” OR “crime” OR “offender” OR “prosecution” OR “police” or “Court” OR “parole officer” OR “victim”]

2.4.3.3 [“intervention” OR “program” OR “outcome” OR “evaluation” OR “effect*”, “experiment” OR “study” OR “control*” OR “comparison”]

2.5 ELECTRONIC SOURCES

The following databases will be searched for eligible studies (appears in alphabetical order), followed by a search in Google Scholar, Google Books and “regular” Google:

2.5.1 Academic Search Premier
2.5.2 C2 SPECTR
2.5.3 CINCH
2.5.4 Criminal Justice Abstracts
2.5.5 Criminal Justice Periodical Index
2.5.6 Dissertation Abstracts (Proquest.com)
2.5.7 ERIC
2.5.8 ESDS
2.5.9 Government Data
2.5.10 Government Publications Office – Monthly Catalogue
2.5.11 Government Publications Reference File
1.1 DATA COLLECTION AND ANALYSIS

1.1.1 ASSESSMENT OF RISK OF BIAS IN INCLUDED STUDIES

The extent to which we can draw conclusions about EM depends on the validity of the outcomes of the primary studies. We are particularly concerned about internal and external validities, given the arguably non-comparability of studies. The reliability of the results may also be at risk, given the methodologies used in primary studies, should low-level studies be included as well.

We plan on using critical assessment for various risk domains in a checklist format, proposed by Juni (2001). This list appears in a tabular format in Appendix III and it contains five types of biases:

1.1.1.1 selection bias (systematic differences between baseline variables that define the groups before using the EM);
1.1.1.2 attrition bias (systematic differences between participants who completed the EM programme and those who have not);

1.1.1.3 performance bias (exposure to factors other than EM);

1.1.1.4 detection bias (i.e., systemic measurement differences); and

1.1.1.5 reporting bias (e.g., selective outcome reporting).

Some items are objective and quite apparent (e.g., the participants were selected and allocated in non-random procedures, precise exclusion criteria were not always used in the selection of the participants, and studies have not incorporated power calculations), and some are subjective (e.g., can the design address the studied question in a comprehensive way?). We will take a robust approach by stating whether within each study there is a “low” “medium” or “high” level of bias on every risk domain, which we will score independently. We will then review these scores to obtain a measurement of each bias across the studies, in order to assess whether their plausible impact on the outcomes.

1.1.2 ASSESSMENT OF PUBLICATION BIAS

Publication bias can lead to systematic bias in our review. We will estimate the reporting bias in published versus unpublished works using funnel plots (Rothstein, Sutton and Borenstein 2005). Funnel plots can be used to assess whether a systematic review is likely to be vulnerable to publication bias, by plotting EM treatment effect (i.e. mean difference between intervention group and control) against the inverse of the variance or the sample size.

However, we will only explore this option should enough studies meet our eligibility criteria.

1.1.3 MEASURES OF TREATMENT EFFECT

1.1.3.1 Effect Size Calculations

The odds-ratio will be the effect size of choice for all outcomes of a dichotomous or binary nature, calculated by comparing EM and non-EM conditions on crime data. Standardised mean difference (SMD) effect size (ES) in the form of Cohen’s \( d \) (Cohen 1988) will be used for
continuous or count data. When before-and-after data are reported, we will compute the ES by adjusting for the before-data only.

We will then convert all scores to SMD using Comprehensive Meta-Analysis 2.0, in order to standardise the ES scores. We will interpret the results using Cohen (1988) criteria for assessing the magnitude of the effects, whereas ES=.2 is considered a small effect size, effect sizes of about .5 to be medium, and effect sizes of .8 or higher to be large.

In addition to this, we are concerned that baseline differences could characterize at least some of the studies, particularly if the process of allocating cases to treatments has not been random. Therefore, for each study, we will first inspect if the groups are comparable at baseline and whether any statistical adjustments are therefore required.

1.1.3.2 Small Sample Size Bias Adjustment

In the case of small sample sizes, we will use the following formula to recomputed Cohen’s $d$, suggested by the Campbell Collaboration\(^2\):

\[
d' = \left(1 - \frac{3}{4N - 9}\right)d
\]

1.1.3.3 Adjusting for Baseline Imbalances

In order to adjust for baseline differences, we will use the difference-in-differences (post-test mean minus the post-test mean for each group) in the numerator of each effect-size (while the denominator continues to be the raw within-group standard deviations, not the standard deviation of differences).

1.1.3.4 Heterogeneity

Homogeneity test (Cochrane-Q) will be calculated to determine if variability across ESs is greater than would be expected from sampling error alone. Q is useful in testing the deviation of the effect size from

---

each study from the combined mean effect size, by using Chi-square
distribution to measure the probability that the combined variation is
zero. However, we will also measure \( I^2 \) statistic, which describes the
percentage of variation across studies that are due to heterogeneity
rather than chance. We will implement this heterogeneity test, given
criticism against the traditional Q test in relation to statistical power in
relatively small n meta-analyses (Higgins et al. 2003).

\[
I^2 = 100\% \times \frac{(Q - df)}{Q}
\]

\( I^2 \) is expressed as a percentage of the total variance in all the data, with
25%, 50% and 75% considered as low, moderate or high level of heterogeneity.

1.2 DATA SYNTHESIS

1.2.1 CRIME DATA

1.2.1.1 Overall Effect Size

Standardised methods for synthesising the data will be used, using
inverse-variance weighted random effects model.

1.2.1.2 Combing Multiple Outcomes

Should we detect that there are several independent outcomes (e.g.,
arrest and reconviction) or time-points within each study (e.g., arrest
within 6 months or 12 months), we will treat them in two separate ways.
First, we will report and synthesize each separately (for example, a
separate meta-analysis for arrest data and a separate for reconviction
data). We hope that sufficient data will be available to cluster the
available information within such homogeneous outcomes.

Second, we will to combine effect sizes using a technique suggested by
Borenstein et al (2009: 227-229), whereas the effect size is computed as
the mean of the multiple outcomes,
\[
\bar{Y} = \frac{1}{m} \left( \sum_{j} Y_j \right)
\]

with \( m \) representing the number of outcomes within a study, and \( Y \) representing the outcome of the \( j \)th outcome (or time point). Variance of the composite, again as suggested by Borenstein et al (2009: 230), is given by the following formula, with \( V \) being the variance, \( r \) as the correlation coefficient between the outcomes (assuming equal variances), and the weight of each is reciprocal of the variance

\[
V_{\bar{Y}} = \frac{1}{m} V(1 + (m - 1)r)
\]

We note that if \( r \) is unknown or cannot be computed from the data available in the primary studies, we will perform a sensitivity analysis with a range of correlation (.00, .25, .50, and .75). In this regards, Borenstein et al. (2009: 232-233) suggest that “treating the [multiple outcomes] as independent of each other yields the same precision as setting the correlation as 0.00…[furthermore,] if the correlations between multiple outcomes are highest in some studies than in others, [than] this variation will affect the relative weights assigned to different studies, with more weight going to the study with a lower correlation.”

1.2.1.3 Combining Multiple Subgroups

As we plan to do for multiple outcomes, we also plan to do for multiple subgroups within each study: we will synthesize them separately, for each class or category of offenders (list of potential variables appears hereunder) and then we will try to combine the subgroups into summary data.
First, if sufficient data are reported in the primary studies, we will explore the impact of potential covariates on the outcomes as well, using meta-regression or analogue-to-the-ANOVA moderators’ analysis. We want to be able to look into the various treatment components as well as extraneous elements that may influence the results, such as:

- Features of the participants (age, gender, ethnicity, type of offence)
- EM components (electronic, radio-signal, privately-operated vs. state operated; degree of EM implementation)
- Methodological quality of the study (e.g., experimental, non-experimental)
- Design features

Second, we will use the summary data from the subgroups to recreate the data for the study as a whole, and then use this summary data to compute the effect size and variance (Borenstein et al 2009:221-222). Thus, the combined sample size across subgroup, the combined weighted mean (by sample size) across groups, and the combined standard deviation, will be computed using the following formulas:

\[
n_1 = n_{11} + n_{12}
\]

\[
\bar{X}_1 = \frac{n_{11}\bar{X}_{11} + n_{12}\bar{X}_{12}}{n_{12} + n_{12}}
\]

\[
s_1 = \sqrt{\frac{(n_{11} - 1)s_{11}^2 + (n_{12} - 1)s_{12}^2 + \frac{n_{11}n_{12}}{n_{11} + n_{12}}(\bar{X}_{11} - \bar{X}_{12})^2}{n_{11} + n_{12} - 1}}
\]

2. TIMEFRAME

Given the extensive work already completed, we envisage completion of the Campbell review by 1st January 2013.
3. PLANS FOR UPDATING THE REVIEW
The review will be updated on a three year basis. As part of this update we will need to code any new studies identified and rerun the analyses.

4. ACKNOWLEDGEMENTS
We would like to acknowledge the assistance of Professor Marc Renzema from Kutztown University, PA.
We would particularly like express our gratitude to Professor David Wilson for his tremendous help and guidance.

5. STATEMENT CONCERNING CONFLICT OF INTEREST
None.
6. REFERENCES


Shapps, G. (2006) *The Tagging Game*


7. APPENDICES

Appendix I  Coding Protocol

Appendix II  Eligibility Criteria

Appendix III  Assessment of Biases
Appendix I

Coding Protocol

Use one coding sheet for each distinct research project that is included in the systematic review.

Throughout, zero = not known, 1 = Yes, 2 = No

Name of coder     Coder__________________________________
Date of coding     Date____________________________________

Study

Study no. identifier    Studyid _________________________________
Author(s) name    Author__________________________________
Author(s) affiliation    Affil __________________________________
Date(s) research was conducted  Dateres ________________________________
Date of primary publication  Datepub________________________________
Place of research    Place___________________________________
Country of research   Country_________________________________
Source of funding

Details of primary report

__________________________________________________________________________
__________________________________________________________________________

Publication type:

1 book
2 book chapter
3 Journal article
4 technical report
5 unpublished conference paper
6 dissertation

Pub type___________

Details of other reports

__________________________________________________________________________
__________________________________________________________________________
2. Eligibility

Eligibility criteria
Are the following inclusion criteria present? (if not clear, attempt to find out from the author)

The report deals with EM including electronic tags, satellite monitoring systems or voice recognition technologies.

Type of EM measured

The report analyzes participants under the supervision of the Criminal Justice System

The study has reported a post program measure of recidivism.

The report indicates which methods were adopted to gather information on EM.

Before data on EM as defined above (specifically) are reported.

After data on EM as defined above (specifically) are reported

The effectiveness of the program was measured by comparing participants who received it (the experimental group) with students who did not receive it (the control group).

The Study had some control of extraneous variables (establishing the prior equivalence of groups) by (i) randomization, or (ii) pretest measures of EM, or (iii) matching, or (iv) pretest measurement of risk factors or risk scores for EM.

Numbers of participants who reoffended are reported.

Scores on EM are reported.
3. Sample

Mean age of participants:  
Experimental: Eage________
Control: Cage________

Gender composition:  
Experimental: % female _____ Epcf_______
Control: % female _____ Cpcf_______

Ethnic composition:  
Experimental: __________________
Control: __________________

Initial sample size:  
Experimental: En1_____________
Control: Cn1_____________

Sample size in short follow-up:  
Experimental: En2_____________
Control: Cn2_____________

Sample size in long follow-up:  
Experimental: En3_____________
Control: Cn3_____________

Number of programmes:  
Experimental: Eprogramme__________
Control: Cprogramme__________

Number of separate programmes:  
Experimental: Esep.progs________________
Control: Esept.progs________________

4. Research design

1. randomized
2. before-after with control condition
3. before-after with age cohort comparison
4. only after with control condition  
Design_______

Who were randomized? 1. Programmes 2. Separate Programmes 3. Participants  
Random_______

How were units allocated to experimental and control conditions?  
____________________
Alloc__________

Was Control condition comparable? 1 = Yes, 2 = No  
Comp___________

Variables measured to establish matching or comparability  
____________________
To what population can the results be generalized? _________________________________
__________________________________________________________________________

Is there a potential generalizability threat from overall attrition?

1 = Yes, 2 = No  
Attrito_____________

Is there a potential threat to internal validity from differential attrition?  
1 = Yes, 2 = No  
Attritd_____________

5. Pretest Measures

What measures were used?

Probation    Yes = 1 No = 2    Pro ________
Community Programme    Yes = 1 No = 2    CP__________
Prison    Yes = 1 No = 2    PR__________
Diverted Sentence    Yes = 1 No = 2    DS__________
Parole    Yes = 1 No = 2    PAR__________
Probation & Parole    Yes = 1 No = 2    ProP__________

Where available, effect size measures will be based on the measures outlined above.

Scale:

What scale was used to measure recidivism?    Rescale_____________

What reference period was used for recidivism? (in days)  Reref___________________

Was there any information about reliability or validity of measures? Rel_____________
Val_________________

6. Intervention (experimental condition)

Type and components of intervention: (Yes = 1, No = 2)

<table>
<thead>
<tr>
<th></th>
<th>Int</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probation</td>
<td>IntPr</td>
</tr>
<tr>
<td>Community Programme</td>
<td>IntCP</td>
</tr>
<tr>
<td>Prison</td>
<td>IntPr</td>
</tr>
<tr>
<td>Diverted Sentence</td>
<td>IntDS</td>
</tr>
<tr>
<td>Parole</td>
<td>IntPar</td>
</tr>
<tr>
<td>Probation and Parole</td>
<td>IntProP</td>
</tr>
</tbody>
</table>
Was the intervention highly structured, that is did it follow a protocol or manual? (Yes = 1, No = 2) Struct________

Duration of program delivery (in months): (Yes = 1, No = 2)

Min _____
Max _____
Mean _____
Fixed _____

What time of the year (month) did the program start? Start ______________

What time of the year (month) did the program end? End ______________

Who delivered the intervention? (Yes = 1, No = 2)

Mental Health Professional DelMen_____________

Criminal Justice Professionals DelCJ___________

Professional educator Deled______________

Non Professional DelNP______________

other (specify ______________________) Deloth_____________

Were there any problems of implementation?
(specify) ___________________________________________

Was there a measure of treatment integrity? Integ______________

What happened to the control group?
1. Probation
2. Community Programme
3. Prison
4. Diverted Sentence
5. Parole
6. Probation and Parole Contgp___________

7. Post-test measures

Follow-up time period (No. of months): Short FU___________

Long FU___________
Scales:

a) What scale was used to measure recidivism in the short follow-up?  Ssrec _____
b) What scale was used to measure recidivism in the long follow-up?  Lsrec _____

What reference period was used for recidivism?  Short FU:  Srrec _____
Long FU:  Lrrec _____

8. Effect Size Measures

Prevalence of recidivism

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of offences before</td>
<td>EOBef</td>
<td>COBef</td>
</tr>
<tr>
<td>No. of non-offences before</td>
<td>ENOBef</td>
<td>CNOBef</td>
</tr>
<tr>
<td>Short: No. of offences after</td>
<td>EOS</td>
<td>COS</td>
</tr>
<tr>
<td>Short: No. of non-offences after</td>
<td>ENOS</td>
<td>CNOS</td>
</tr>
<tr>
<td>Long: No. of offences after</td>
<td>EOL</td>
<td>COL</td>
</tr>
<tr>
<td>Long: No. of non-offences after</td>
<td>ENOL</td>
<td>CNOL</td>
</tr>
</tbody>
</table>

Short: Odds Ratio ORBS ________________  Confidence Interval  CIBS ________________
Long: Odds Ratio ORBL ________________  Confidence Interval  CIBL ________________
Long: Odds Ratio ORVL ________________  Confidence Interval  CIVL ________________
Recidivism scores

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>EMOBef</td>
<td>CMOBef</td>
</tr>
<tr>
<td>SD</td>
<td>ESDOBef</td>
<td>CSDOBef</td>
</tr>
<tr>
<td>N</td>
<td>ENOOBef</td>
<td>CNOOBef</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short: Mean</td>
<td>EMOS</td>
<td>CMOS</td>
</tr>
<tr>
<td>SD</td>
<td>ESDOS</td>
<td>CSDOS</td>
</tr>
<tr>
<td>N</td>
<td>ENOOS</td>
<td>CNOOS</td>
</tr>
<tr>
<td>Long: Mean</td>
<td>EMOL</td>
<td>CMOL</td>
</tr>
<tr>
<td>SD</td>
<td>ESDOL</td>
<td>CSDOL</td>
</tr>
<tr>
<td>N</td>
<td>ENDOOL</td>
<td>CNOOL</td>
</tr>
</tbody>
</table>

Short: dBS_________________________      SEBS________________________________

Long: dBL____________________________ SEBL________________________________
## ASSESSMENT OF BIASES

<table>
<thead>
<tr>
<th>Risk of Bias</th>
<th>Number of Studies who are characterised by the bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Bias</strong></td>
<td></td>
</tr>
<tr>
<td>Attrition bias</td>
<td></td>
</tr>
<tr>
<td>Detection bias</td>
<td></td>
</tr>
<tr>
<td>Performance bias</td>
<td></td>
</tr>
<tr>
<td>Reporting bias</td>
<td></td>
</tr>
<tr>
<td>Selection bias</td>
<td></td>
</tr>
</tbody>
</table>

Approach for summary assessments of the risk of bias for each important outcome (across domains) within and across studies

<table>
<thead>
<tr>
<th>Risk of bias</th>
<th>Interpretation</th>
<th>Within a study</th>
<th>Across studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low risk of bias</strong></td>
<td>Plausible bias unlikely to seriously alter the results</td>
<td>Low risk of bias for all key domains</td>
<td>Most information is from studies at low risk of bias</td>
</tr>
<tr>
<td><strong>Unclear risk of bias.</strong></td>
<td>Plausible bias that raises some doubt about the results</td>
<td>Unclear risk of bias for one or more key domains</td>
<td>Most information is from studies at low or unclear risk of bias</td>
</tr>
<tr>
<td><strong>High risk of bias</strong></td>
<td>Plausible bias that seriously weakens confidence in the results</td>
<td>High risk of bias for one or more key domains</td>
<td>The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results</td>
</tr>
</tbody>
</table>