

# The Empirical Basis of Sex Offender Treatment Effectiveness

**Conor Duggan**

University of Nottingham, UK

[Sexual Offender Treatment, Volume 9 (2014), Issue 2]

## Abstract

*Whether or not sex offenders can be successfully treated remains a matter a controversy. This is due, in part, to the quality of evidence that one might regard as acceptable. Evidence-based practice for treatment efficacy lays great weight on the evidence from randomised controlled trials but is this the only or indeed the most appropriate approach for this type of disorder?*

*I this review the findings I summarize my presentation at the IATSO conference 2014 in Porto including results of two recently completed Cochrane Reviews that assessed efficacy for both psychological and pharmacological interventions for sex offenders. Also, I examine the recent concern that psychological treatments fail to consider adequately the harm that might arise from such interventions. This latter concern, together with the uncertainty as to which intervention ought to be offered to whom leaves the practitioner in an uncomfortable position in deciding which treatment ought to be recommended for a convicted sex offender.*

*Keywords: sexual offenders, treatment, efficacy, effectiveness, Cochrane, review*

Examining the literature over the past 20 years on whether sex offenders can be effectively treated is likely to leave the reader with a sense of uncertainty as each review appears to contradict its predecessor. For instance, the first major review by Furby et al (1989) concluded that treatment did not reduce recidivism among sex offenders. Nagayama Hall (1995), in addition to criticising the methodology adopted by Furby and colleagues, updated their earlier review and found almost equivalent positive effects for hormonal and cognitive behavioural treatments. White et al. (2000) in the first of a series of Cochrane Reviews could identify only three studies that satisfied their inclusion criteria and found no evidence of effectiveness. Kenworthy et al (2003) in a further Cochrane Review that updated the psychological component of the earlier White review concluded that '...some evidence indicating positive effects of psychological interventions ... (had) begun to emerge.' This was further supported by Losel & Schmucker (2005) who in reviewing all treatments found evidence of effectiveness - especially for surgical castration. This positive assessment was in turn contradicted by Rice and Harris (2013) who claimed that there was insufficient evidence to reject the null hypothesis that treatment had any effect.

Fortunately, recent methodological and conceptual advances have led to a consensus on the priority given to different types of evidence that might emerge from various designs. This hierarchy of evidence (see fig 1) gives priority to that derived from systematic reviews of randomised controlled trials as offering the most compelling evidence for effectiveness. This is captured by the following quotation on treatment evaluation from Sackett et al (1996) where they advise to '...avoid non-experimental approaches ... since these routinely lead to *false positive conclusions about efficacy*... (so that) ... the systematic review of several randomised trials ... has become the "gold standard" for judging whether a treatment does more good than harm.' (my italics). Their concern then is that 'non-experimental approaches' (i.e. non-RCTs) are prone to inflate the benefit of the

intervention being studied.

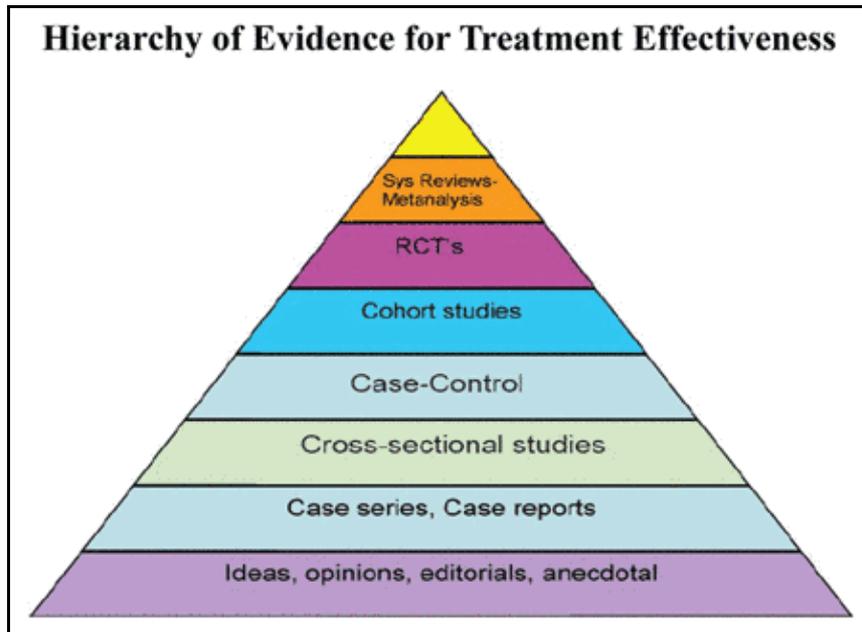


Figure 1

This belief on the superiority of RCTs needs to be tempered by an understanding that treatments evaluated in RCTs do not arise out of the blue, rather they generally have a long gestation of several years, often beginning with case reports that are then subject to more stringent evaluations as one rises in the hierarchy before being subjected to a RCT. Thus, these different layers of the hierarchy are complementary rather than contradictory to one another.

A further point to make about the evidence layers within the hierarchy is less favourable and this is that the further one goes down the hierarchy, the credence one can place in the evidence available at that level diminishes. It may come as a surprise that the opinions of experts when they publish reviews or textbooks carry the least credibility. This may seem strange as these are the leaders in the field and so might be regarded to have views on which one could rely. Unfortunately, clinical experts often lead one astray. An example of how expert clinicians can mislead is provided by a much cited paper by Antman et al (1992). When this group compared the recommendations from experts for the treatment of heart attack with the already established evidence by a meta-analysis of randomised controlled trials, they found that (a) the experts had not only ignored beneficial treatments but (b) were continuing to recommend treatments that were either of no benefit or in some cases actually harmful.

## What is a Systematic Review?

For those who are not familiar with the process of systematic reviewing, it is worth emphasising two points in particular from the process of conducting such a review. A systematic review ought to be seen as a scientific experiment where the methodology is clearly set out at the beginning and then followed rigorously to its conclusion. Thus, it ought to be possible to replicate the findings from a

systematic review by following the same methodology if there is agreement on the formulation of the question. The fact that there is such a lack of agreement on whether or not sex offenders can be satisfactorily treated as described above must therefore stem from a lack of agreement on either the formulation of the problem or the methodology employed or both.

Central to the conduct of a systematic review is clarity on the formulation of the clinical problem that is captured by the acronym PICO so that the **Population**, the **Intervention**, the **Comparator** and the **Outcome** are each specifically denoted before the citations are examined and analysed. Thus, for the treatment of sex offenders, one would need to specify whether the review includes all sex offenders or just a subgroup (e.g. rapists, child molesters etc.) as the population being studied. Similarly, is the intervention to include all types of intervention (as in the White et al. 1998 review) or restricted to psychological interventions only (as in the Kenworthy et al. 2003 review)? For the comparator, one needs to specify if the comparison group is a no treatment control or is the experimental treatment being measured against another active control treatment or both? Finally, the main (or primary) outcome needs to be specified - this being the main point of the experiment - although a number of secondary outcomes are also usually considered. While the primary outcome in sex offending treatment ought to be the rate of recidivism, this outcome unfortunately is problematic as (a) the base rate is low - so that a very large number of subjects needs to be entered into the trial to demonstrate an effect and (b) the follow-up needs to include a lengthy time at risk in the community to capture a reasonable number who might reoffend. As we shall see shortly, few of the trials considered in the recent reviews measure re-offending - certainly for the necessary period at risk - and this diminishes their value. So-called surrogate outcomes (e.g. anxiety, anger etc) were often measured but these may be poor proxies for reoffending.

There is one further recent challenge on the choice of which outcome to measure that can be seen in the general literature. This is the need to involve patients in deciding on which outcome is important to them as consumers of the intervention. It is now recognised, for instance, that the outcomes assumed by researchers to be the most important are not necessarily those that are considered to be most important by patients. Given the disenfranchisement of sex offenders, it may seem odd to involve them in any decisions about their life and future but who is to say that that they may not be surprisingly informative about this if they were asked, as has been the case elsewhere.

The other aspect that I wish to consider is the need to constantly improve and update extant reviews. The need for this will be obvious on two counts. First, the field moves forward with new trials being published whose inclusion that may alter or confirm the findings from earlier reviews. Second, and this may seem less obvious, previous reviews may not be as comprehensive as originally believed and for this reason may be seriously misleading. An example is provided by comparing the findings from a succession of Cochrane Reviews that essentially used the same methodology. Here we find that the more recent reviews have identified several trials that ought to have been evaluated by the earlier reviews but were omitted. While the reasons for these omissions could have been due to inadequate searching of the literature at the time or faulty processing is less important than the need for secondary reviews to revise the extant literature comprehensively with the modern technology. This implies an equal obligation on secondary commentators to cite the most up to date review. For instance, although we have identified several omissions in the White et al 1998 review, this continues to be cited in several subsequent reviews while the findings of its later superior reviews are omitted.

One further issue to examine in this preliminary discussion is the place of meta-analysis. While some regard this as being independent and separate from a systematic review per se, I believe that it is better to see it as an optional part of a systematic review. Briefly, there are three points to be made. First, a meta-analysis calculates a 'common' or 'average' treatment effect based on pooled

data from two or more studies. Second, by so doing, it improves the precision of the point estimate by using all the available data. An example of this meta-analysis being used to good effect is Petrosino's 2013 meta-analysis of an intervention for juvenile delinquency where he showed that the intervention had the reversed effect from that which was intended so that by being able to incorporate the findings from all seven RCTs (five of which were statistically insignificant), and show an odds ratio of 1.72 in favour of the control. This meant that juveniles exposed to this intervention were nearly twice as likely to engage in criminal activity as adults - the opposite of that which was intended. Third, meta-analysis should only be done when studies are sufficiently similar as regards their design, intervention and outcome that it makes sense to combine their results. As with much else in science, this involves a judgement call so that one is not combining 'apples with oranges'.

## Randomised Controlled Trials

As high quality systematic reviews are dependent on properly conducted randomised controlled trials, it is important to consider some of the criteria that define a properly conducted trial. Before going into details, it is first of all important to recognise the simplicity and purpose of an RCT. This consists of a random allocation of eligible subjects such that chance alone determines whether the individual is allocated to one treatment or the other. Having being allocated to the intervention being evaluated or the comparator, the subjects are given one or other intervention and are then compared to determine any differences and, ideally, followed-up to see if any differences observed are sustained. The purpose of the design is to adjust for any possible confounding - known or unknown - that might otherwise account for any differences that are found. An example of a well-conducted RCT is the International Study of Infarct Survival (ISIS) in which over 17,000 individuals who had suffered or believed to have suffered a heart attack were randomised within 24 hours to one of four conditions: streptokinase - a clot busting drug, aspirin - which prevents clots forming, a combination of streptokinase and aspirin or a placebo. The investigators found a significant reduction in mortality for both of the active treatments (i.e. streptokinase or aspirin) over the placebo, a reduction that was further enhanced when the two drugs were combined. Moreover, the difference in mortality persisted to 10 years of follow-up. While the fact that this trial had recruited patients from over 17 countries meant that it was organisationally complex, nonetheless suggested that the results were generalizable.

While the design and recruitment to the ISIS trial because of its scope and rigour is exceptional, it, together with any other RCT is subject to several biases that need to be considered when evaluating the results from any RCT. These can occur at several stages through the conduct of a trial and each needs to be evaluated when assessing its quality. There are six possibilities of bias which all Cochrane Reviews are now required to report on explicitly in a Risk of Bias Table. Here, I will simply list each of these with a brief description but the interested reader should consult Greenhaigh (1997) for further information. The following biases need to be considered when evaluating a RCT: (a) Selection bias - how participants are entered into the trial, (b) Allocation bias - whether the allocation of participants has been properly concealed? (c) Performance bias - how participants are exposed or not exposed to the intervention, (d) Attrition bias - how completely the participants are followed up? (e) Detection bias - how participant outcomes are assessed? (f) Reporting bias - whether a study is published and if published whether all outcomes are published or only the significant ones? The point to underscore here is that just because a trial describes itself as being randomised is no guarantee as to its quality.

## Cochrane Systematic Reviews

The Cochrane Collaboration was founded in 1993 comprising 13 centres covering 52 specialities. The Cochrane Library contains the Cochrane Database of Systematic Reviews (CDSR), the Cochrane Controlled Trials Register (Central) in addition to several other registers and databases. The Cochrane Database of Systematic Reviews provides full text if completed reviews carried out by the Cochrane Collaboration together with protocols for reviews currently in preparation. It restricts its evidence to that from randomised controlled trials. Named after Archie Cochrane who was an epidemiologist and health service methodologist whose most famous work was 'Effectiveness and Efficiency: random reflections on health services' (Cochrane 1972) was influential in promoting the use of randomised controlled trials in the evaluation of health service interventions. On reviewing the book of the British Medical Journal in 1972, for instance, CT Dollery wrote '...the hero of the book is the randomised controlled trial, and the villains are the clinicians in the 'care' part of the National Health Service (NHS) who either fail to carry out such trials or succeed in ignoring the results if they do not fit in with their preconceived ideas.'

## **Cochrane Systematic Reviews of the Treatment of Sex Offenders**

I will now describe briefly the findings from two recent reviews into the psychological and pharmacological reviews of sex offenders. As indicated earlier, these are updates on the earlier White et al (1998) and Kenworthy et al. (2003) Cochrane Reviews. A similar search strategy was developed for both reviews that involved searching 20 databases up to Sept 2010. For each of these reviews, two authors working independently, selected the studies, extracted the data and assessed the studies' risk of bias. The search strategy produced 36,704 citations, of which 36,308 were excluded as not being relevant. The full text of the remaining 396 studies were then examined and a further 251 citations were excluded as not being treatment trials. From the remaining 144 studies, a further 128 studies were excluded as not being RCTs leaving 16 RCTs. There were 10 trials of psychological interventions, 7 of pharmacological interventions and one which was a three armed trial of psychological vs pharmacological vs placebo so that it appeared in both reviews.

### **Psychological Interventions for Sex Offenders (Dennis et al., 2012)**

Ten studies were included with 944 eligible participants. Four of the studies ( $n = 70$ ) involved behavioural interventions, 5 involved CBT-type intervention ( $n = 665$ ) and one study involved a psychodynamic intervention ( $n = 231$ ). Few of the studies provided information of the review's primary outcome (i.e. re-offending) and in the largest and best-designed study (SOTEP) where this was examined, no difference was found in the rate of re-offending between the treated and the control group ( $RR = 1.10$ , 95%  $Ci$  0.78-1.56).

### **Pharmacological Interventions for Sex Offenders (Khan et al., in press)**

Seven studies were included with 123 eligible participants. Three testosterone suppressing drugs (i.e. cyproterone acetate, medroxyprogesterone and ethinyl oestradiol) were assessed in six studies in which two used medroxyprogesterone as an adjunctive treatment to a psychological therapy. The other study used two antipsychotic drugs (benperidol and chlorpromazine) versus a placebo. The results of these small studies ( $n = 19.2$ ), while encouraging, did not provide good evidence for the use of any of the pharmacological interventions. Furthermore, all of the studies were at least 20 years old and none had tested newer drugs (e.g. SSRIs or GnHR analogues).

In conclusion, the evidence of efficacy for sex offending interventions from RCTs is weak. Most trials were too small, of insufficient duration, assessed outdated treatments and used surrogate

outcomes for re-offending to enable the practitioner to decide on what works best for whom'. Furthermore, this situation is unlikely to change in the near future as there are very few pending trials for sex offenders on the Trials Register.

## Barriers to Conducting RCTs

Before considering why there are so few RCTs for the treatment of sex offenders, I want to mention some of the barriers to conducting RCTs in general. The first of these is that, although simple in design, running a large RCT is very expensive - often running into millions of euro/dollars/pounds. They are also organisationally complex and require the researchers to battle with various ethical committees and other regulatory agencies whose function appears to be to inhibit rather than to facilitate research. They are therefore time consuming - often taking years to commence and recruit a sufficient numbers of patients - and this is quite apart from any follow-up that may be involved. For all of these reasons, funders tend to act conservatively and are reluctant to provide resources to fund novel treatments unless these (and their implementers) have some track record in the area. The second reason is that trials usually provide an answer to a single question so that the tighter the protocol of the trial, which thereby increases the likelihood of an unequivocal answer, the more likely it is that there will be unanswered questions such as 'Would the result be the same if we had given a larger/smaller dose at an earlier/later time to a different type of patient? The answer clearly is that one does not know and the only way to find out is to conduct yet another trial under those conditions. Unfortunately, those who have completed a successful trial are more likely to want to refine their intervention further rather than expanding into the unknown, while those who have failed to find any effect are equally unlikely to wish to waste energy in a variation of where they had previously failed. A final reason is that RCTs are only indicated where there is uncertainty as to the outcome (what is known as 'equipoise') between the competing conditions. If there are very good reasons for believing that some intervention is likely to be a success (as for instance, with the introduction of penicillin in the 1940s) then it would be inappropriate (and unethical) to conduct a RCT. These situations, however, are the exception rather than the rule and do not appear to apply to the treatment of sex offenders where all the evidence points to a modest gain at the margins - if indeed there is any gain at all.

## Is the rejection of RCTs for sex offenders based on sex offenders being a Special Case or is this just Special Pleading?

I have already mentioned some of the difficulties in conducting sex offender trials. Some of these are real, for instance, the low base rate and long period of follow-up meaning that very large numbers require to be recruited in order to demonstrate an effect; other objections, however, I believe carry less weight and I shall now consider some of these briefly and any interested reader will find these arguments elaborated more thoroughly in Duggan & Dennis (2014).

**Objection 1.** There is a belief that, while a RCT may be a suitable design for a medical intervention such as aspirin, it is not an appropriate design for complex psychosocial interventions. Although it is agreed that psychosocial interventions may need some alterations as regards how they are reported (Montgomery et al, 2013) I fear that this objection carries little weight as several trials of psychosocial interventions have already been successfully carried out. Even in the field of sexual offender treatment, the successful evaluation of multisystemic therapy for juvenile sex offenders using an RCT design (Borduin, Schaeffer & Heilblum, 2009) indicates that this objection to be false. This objection, however, is helpful as it leads to a more general discussion of different types of

RCTs (See Fig 2). There is a basic distinction, for instance, between Explanatory and Pragmatic Trials. The former - which tests efficacy - seeks to answer the question: 'Can this intervention work?' Hence, it is usually conducted under ideal conditions with carefully selected patients, well trained therapists and measuring outcomes that are often not relevant to routine clinical practice. Conversely, pragmatic trials - which assess effectiveness - are designed to answer the question: 'Does it work?' They are conducted as part of normal service delivery with unselected patients, staff with less expertise, and have outcomes that are patient related. While explanatory and pragmatic trials are on a continuum, it is sensible to establish efficacy first before assessing effectiveness.

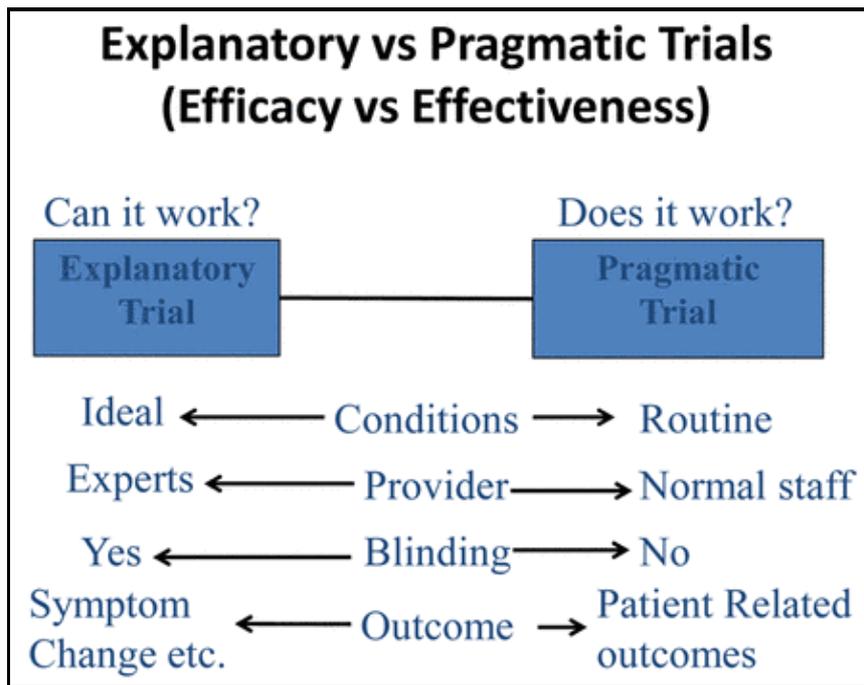


Figure 2

This explanatory/pragmatic distinction is itself a part of a larger continuum in treatment assessment that comprises four phases (see Fig 3). While these phases have particular applicability to drug assessment and approval which I shall describe, the same scheme also broadly applies to the evaluation of the psychological therapies. In this scheme, Phase I involves giving the new drug to normal volunteers to study its safety. In Phase II, feasibility trials test out whether larger trials are possible and under what conditions they might be carried out. Phase III comprises the explanatory/pragmatic distinction which I have described above and Phase IV involves a roll out to the population at large with post-marketing surveillance for the cost implications, adverse effects etc. This figure allows us to draw two broad conclusions: (a) most interventions fail as the failure rate at each of these stages is substantial and (b) the length of time required to complete this process is considerable, probably a minimum of eight years and often considerably longer. As it therefore takes a considerable time to identify a successful intervention and when this is coupled with the time needed to influence policy and be translated into practice, the challenge in transferring research into practice is challenging. Echoing the findings from the Antman et al. 1992 paper on the

treatment of heart attack, it has been estimated that translating the findings from a good meta-analysis to routine practice takes about 15 years, one can see the enormity of trying to change practice even when good evidence is available.

Phase I	Phase II	Phase III	Phase IV
20-80 participants	100-300 participants	1,000-3,000 participants	Thousands of participants
Up to several months	Up to (2) years	One (1) - Four (4) years	One (1) year +
Studies the safety of medication/treatment	Studies the efficacy	Studies the safety, efficacy and dosing	Studies the long-term effectiveness; cost effectiveness
70% success rate	33% success rate	25-30% success rate	70-90% success rate

Figure 3

**Objection 2:** This refers to the RCTs demand standardisation of therapy delivery so that this reduces the flexibility and responsiveness of a competent therapist. This has a particular resonance with the need for the 'responsivity' element within the well- established RNR approach in criminology and requires answering. The first point to make is that while RCTs do require standardisation of the therapeutic input through manualisation, this is not necessarily a bad thing as it reduces therapeutic drift. Indeed, the rigour for fidelity in many of the programmes delivered in correctional settings would put to shame the delivery in health settings. The second point is that manualisation ought not necessarily affect therapeutic qualities such as empathy or warmth as alleged by the critics. Most importantly, however, and this is a point that I will expand further below when I consider harm, the standardisation and recording within a therapeutic trial offers an opportunity to identify the characteristics of good and bad therapists and hence can be an opportunity of advancing the field further.

**Objection 3:** This involves the claim that those offenders randomised to the 'no treatment control' will be interpreted as being at greater risk so that if they were subsequently to re-offend, the public would be outraged and officials blamed. This is clearly illogical as it assumes that the treatment is effective whereas the point of conducting an RCT is that there is uncertainty as to whether the intervention works or not. Moreover, it also assumes that the intervention is unlikely to do harm; an assumption which I shall challenge using the findings from the Cambridge Somerville Youth Trial as an example.

**The Cambridge-Somerville Youth Trial.** This intervention, which was designed to reduce delinquency, sought to remedy the multiple deficiencies faced by disadvantaged youth and their families by providing '...support, friendship and timely guidance' by a trained counsellor. Each boy (with a median age of 10.5 years ) was matched with another across a range of variables to ensure

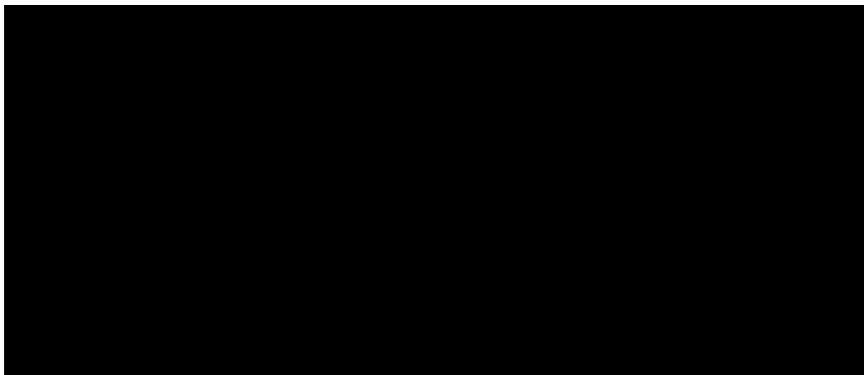
that they (and their families) were similar and then randomised to the treatment or control condition. Those in the interventions group were supported in several ways by the counsellors who visited the families at least twice a month and the trial continued for 5 years.

The trial ended in 1945 and it appeared that many of the boys in the treatment group had improved in their level of adjustment. They and their parents also reported being satisfied with the intervention. However, compared to their controls, there was no difference in their outcome across a series of measures. So, it was proposed to conduct a longer term follow-up as treatment effects were expected to appear as the youths matured.

This led Joan McCord conducted a 30 year follow-up between 1975-1981 achieving a 98% ascertainment. Among the 253 pairs, she found no difference in their outcome in 150. But in the 103 where there was a paired difference, those in the treatment group were more likely (a) to have died prematurely, (b) to suffer from major mental disorder and (c) to have committed two or more crimes (d) show signs of alcoholism (f) to have lower levels of occupational attainment etc, **(i.e. to have fared worse!)** (McCord, 2003). It is worth quoting McCord's own conclusion: 'Let me emphasise again the fact that the Cambridge-Somerville Youth Study was **effective**. The intervention had lasting effects. **These effects were not beneficial**. The important legacy of the program, however, is its contribution to the science of prevention.'

Two observations pointed to the fact that intervention was harmful: (a) Adverse effects increased with increased intensity and duration of treatment (i.e. it reflected a dose response relationship). (b) Adverse effects occurred only among boys whose families had cooperated with the programme. Although the full explanation for 'why' this harm occurred has proved to be elusive, one hypothesis is that the more deviant youth influenced their more vulnerable peers while they were at Summer camp. This form of 'Deviancy Training' (Dishion & Dodge 2005) ought to concern therapists as sex offender programmes are often provided in a group context to individuals with high levels of deviancy.

There is one further general implication from this trial that requires to be noted and this refers to 'The Deterioration Effect' a term coined by Bergin in 1966. In answering the Eysenckian critique that psychotherapy interventions achieved no more than what might occur with spontaneous remission, Bergin observed that that the variance in the active intervention was generally greater than in the no treatment control. This implied that, when there was no difference in the mean effect in the outcome between the two groups, this increase in the variance in the active treatment condition meant that while many more were profiting from the intervention than in the no treatment control, these were being compensated for by those who had deteriorated or harmed (see fig 4).



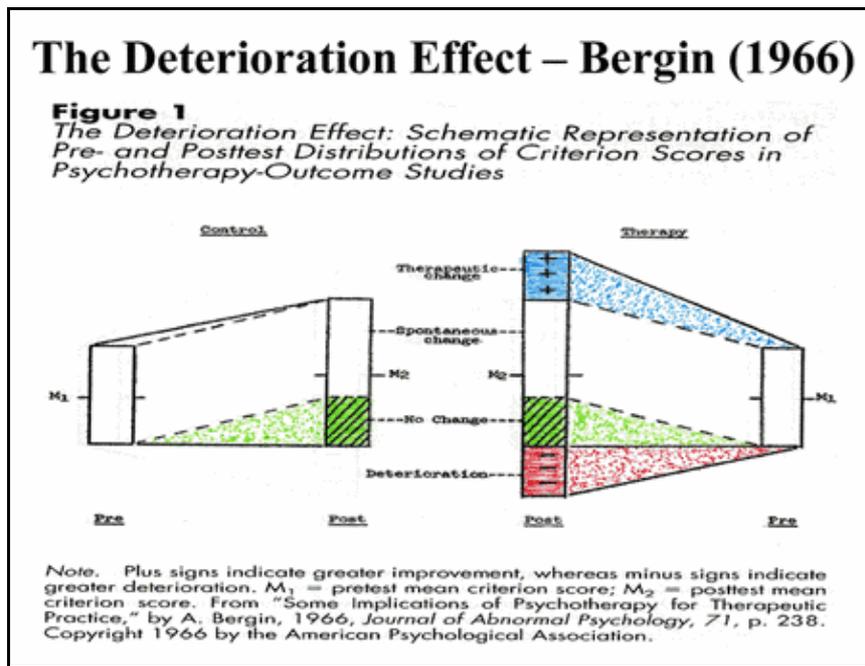


Figure 4

A randomised controlled trial - because it can control for the natural history of the disorder - therefore provides us with the most valid estimate of harmful (and beneficial) effects. This is an opportunity which we have not made the most of because trialists tend to report on the mean effect rather than the full range of scores by individual. Were such scores available, inspection of the outliers in the fourth quartile might identify those who are being harmed by the intervention as well as those who benefit from it (from the first quartile) (Duggan et al. 2014). For instance, it would have been very interesting in the SOTEP trial to know if (a) the variance in the treated group was greater than that in the control and, if it was, (b) to learn something of the characteristics of those at either end of the outcome continuum. This identification would be only the first step, what is then required is a mechanism to explain that effect. Successful application of this process would, not only lead to the avoidance of harm but identify which psychological intervention '...works best for whom'.

**A final question:** Is it worth it? Having considered whether an intervention can and does work, a further question remains: Is it worth it? (Haynes 1999). This distinction between effectiveness of an intervention and its value relates to the cost of the intervention as  $\text{Value} = \text{Effectiveness}/\text{Costs}$ . Muir-Gray (2011) has remarked on a paradigm shift by commissioners of services in the past decade who will now be looking for value while regarding quality (i.e. effectiveness) as necessary but not sufficient. Within the 3x3 matrix in figure 5, one can see that decisions on which services to fund will be based on both effectiveness and costs with the triangle on the upper right being supported and that on the lower left being rejected. Regrettably, none of the trials in the two recent Cochrane Reviews gave any information on the costs of the intervention which is an unfortunate omission given that sex offenders are expensive individuals to manage (especially within the custodial system) so that any effective interventions are likely to be cost effective.

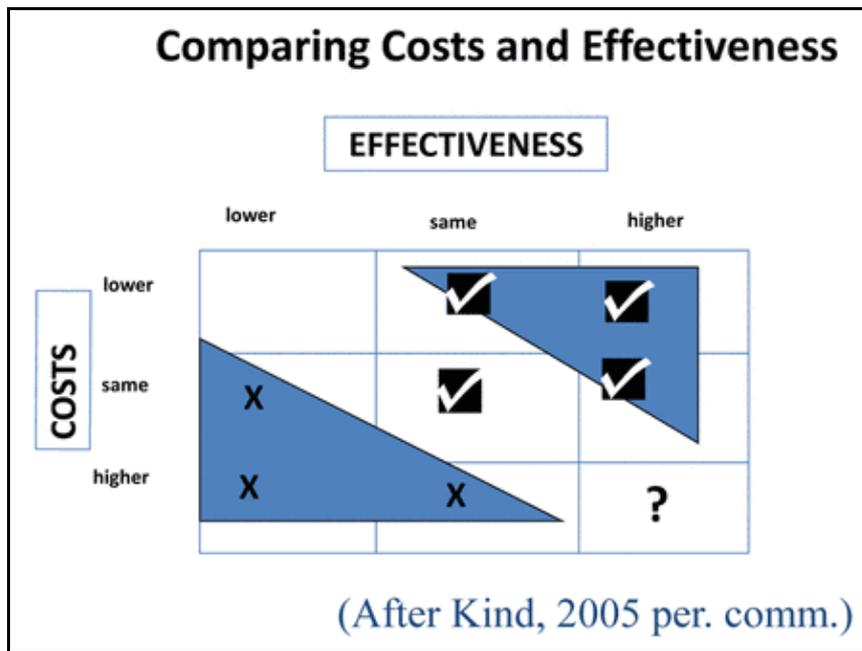


Figure 5

## Summary

There are very few good quality trials for the treatment of sex offenders and many of the extant trials are dated. Although RCTs are difficult to carry out, the 16 trials included in the two Cochrane Reviews contrast markedly with the 13,290 trials on the Cochrane Database for schizophrenia and 16,483 trials on the Cochrane Depression, Anxiety and Neurosis Register. But, does this lack of evidence matter for the (a) sex offender, (b) the public (c) the policy makers?

Here, I shall leave the last word to Muir-Gray who makes a distinction between policy-making and policy-taking (Muir-Gray, 2001). He argues that it is the responsibility of the scientist to produce evidence that contributes to policy-making. When, however, a policy needs to be taken, this involves, not only an assessment of the evidence but integrating that with competing values and limited resources. This review into the treatment of sex offenders suggests that there is some way to go if the necessary evidence for this group is to emerge.

## References

1. Antman, E.A., Lau, J., Kopelnick, B., Mosteller, F. & Chalmers T. (1992). A comparison of Results of Meta-analysis of Randomised Controlled Trials and Recommendations of Clinical Experts. *Journal of the American Medical Association*, 268 (2), 240-248.
2. Bergin, A.E (1966). Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology*, 71, 235-246.
3. Borduin, C.M., Scheffer, C.M. & Heilblum, N. (2009). A randomized controlled trial of multisystemic therapy with juvenile sex offenders: effects of youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology*, 77(1), 26-37.
4. Cochrane, A.L. (1972). Effectiveness and efficiency: random reflections on health services. The Nuffield Provincial Hospital Trust.
5. Dennis, J., Huband, N., Khan, O., Ferriter, M., Jones, H. & Powney, M. (2012). Psychological interventions for adults who have sexually offended or at a risk of sexually

- offending. Cochrane Database of Systematic Reviews. Issue 12 Art No. CD007507.
6. Dishion, T.J. & Dodge, K.A. (2005). Peer Contagion in Interventions for Children and Adolescents: Moving Towards an Understanding of the Ecology and Dynamics of Change. *Journal of Abnormal Child Psychology*, 33 (3), 395-400.
  7. Dollery, C.T. (1972). "Constructive Attack: Effectiveness and Efficiency. *Random Reflections on Health Services*. (AL Cochrane). *Book Reviews British Medical Journal*. 56.
  8. Duggan, C & Dennis, J. (2014). The place of evidence in the treatment of sex offenders. *Criminal Behaviour and Mental Health*, 24, 153-162.
  9. Duggan, C., Parry, G., McMurrin, M., Davidson, K. & Dennis, J. (2014). The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR funded trials. *Trials*, 15 (335) ISSN 1745-6215.
  10. Furby, I., Weinrott, M.R. & Blackshaw, L. (1989). Sex offender recidivism: a review. *Psychological Bulletin*, 105(1), 744-59.
  11. Greenhalgh, T. (1997). How to read a paper: Assessing the Methodological Quality of Published Papers. *British Medical Journal*, 315, 305-308.
  12. Hall, G.C.N. (1995). Sex offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Clinical and Consulting Psychology*, 63 (5), 802-809.
  13. Haynes (1999). Can it work, Does it work? Is it worth it? *British Medical Journal*, 319, 652-653.
  14. Khan, O. et al. Pharmacological interventions for those who have sexually offended or are at risk of offending. *Cochrane Database of Systematic Reviews* (in press).
  15. Kenworthy, T. et al. (2003). Psychological interventions for those who have sexually offended or at risk of sexually offending. *Cochrane Database of Systematic Reviews*. Issue 4. CD004858.
  16. Losel, F. & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: a comprehensive meta-analysis. *Journal of Experimental Criminology*, 1 (1), 117-46.
  17. McCord, J. (2003). Cures that Harm: Unanticipated Outcomes of Crime Prevention Programs. *The Annals of the American Academy of Political and Social Science*, 587, 16-30.
  18. Montgomery, P., Grant, S., Hopewell, S., MacDonald, G., Moher, D., Michie, S. & Mayo-Wilson, E. (2013). Protocol for CONSORT - SPI: an extension for social and psychological interventions. *Implementation Science*, 8, 99 doi:10.1186/1748-5908-8-99.
  19. Muir-Gray, J.A. (2001). Using systematic reviews for evidence based policymaking. (in *Systematic Reviews in Health Care: Meta-analysis in context*. Eds M. Egger, G. Davey Smith & D.A Altman. pp410-418) London BMJ Publications.
  20. Muir-Gray, J.A. (2011). *How to get better value healthcare*. Offox Press Ltd.
  21. Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M.E., Lavenberg, J.G. (2013). 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database of Systematic Reviews*, 4, CD002796. doi: 10.1002/14651858.CD002796.pub2.
  22. Rice M.E, Harris G.T. (2013). Treatment for adult sex offenders: May we reject the null hypothesis? In: In K. Harrison & B. Rainey (Eds), editor(s). *Handbook of Legal & Ethical Aspects of Sex Offender Treatment & Management*. London: Wiley-Blackwell.
  23. Sackett, D.L., Rosenberg, M.C., Muir-Gray, J.A., Haynes, B.R. & Scott-Richardson, W. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, 71.
  24. White, P., Bradley, C., Ferriter, M. & Hatzipetrou, L. (2000). Managements for people with disorders of sexual preference and for convicted sexual offenders. *Cochrane Database Syst Rev*. (2): CD000251.

## Author address

***Conor Duggan***

*University of Nottingham, UK*

*Conor.Duggan@nottingham.ac.uk*