CHAPTER 17

Descriptive Validity and Transparent Reporting in Randomised Controlled Trials

AMANDA E. PERRY

INTRODUCTION

The validity of any research is key in addressing the inferences or conclusions that stem from a study (Campbell 1957). Such issues of validity have been central to the development of research design within criminology since the early 1950s, when two famous social scientists researched the validity of experiments in social settings. This work conducted by Campbell (1957) and Campbell and Stanley (1963) proposed four distinct categories of validity. These included internal validity, (i.e., whether the intervention did cause a change in the outcome) statistical conclusion validity, (i.e., whether the cause and effect were related), construct validity (i.e., the measurement of the theoretical constructs that underlay the intervention and outcome), and external validity (i.e., the representative or generalisability of causal relationships across populations, type of treatment, or intervention).

While these forms of validity have been around for a number of decades, more recent forms of validity, such as descriptive validity, have been developed to enhance the reporting of trial information (e.g., Farrington 2003). In criminology, descriptive validity was first noted as important because of its role in systematic reviews. Most commonly reported studies in systematic reviews are those using either quasi-experimental or experimental study designs, which are often combined to provide an overall effect of treatment. The combining of studies through meta-analytical techniques is often completed on the assumption that such studies are similar on a number of different measures. Such measures include the inferences that can be drawn with respect to the internal and external validity of the study. In all studies, the judgement upon which this inference is drawn is based upon the availability of information reported by the authors of the study, or what we refer to as descriptive validity.

Descriptive validity is, therefore, inextricably linked to the concept of internal and external validity and relies upon the transparent reporting of information. As a result, a well-conducted study will only have high descriptive validity if the authors of the study fully describe the details of the study at all stages (e.g., identification and selection of participants

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_17,

through to interpretation of the results). Authors who conduct sound studies, but who provide poor descriptions about how the experiment was conducted, would, therefore, have low descriptive validity.

In healthcare, such concerns about the transparent reporting of Randomised Controlled Trials (RCTs) brought about the development of the Consolidated Standards of Reporting Trials (CONSORT) Statement, which was specifically devised to assess the extent to which authors reported transparent information within RCTs (e.g., Altman 2001). For these reasons, we are particularly interested in the issue of descriptive validity and its importance for criminologists with RCTs. This is not to say that descriptive validity is not important with other types of study designs, but the preponderance of evidence in other discipline areas has concentrated on RCTs because of their use and influence in conclusions drawn from meta-analyses.

As a result, this chapter will consider the emergence of descriptive validity in criminology and the application of the CONSORT Statement to criminal justice trials. In particular, the chapter refers to two studies, which help us to identify poor descriptive validity in criminology and provide some examples of best practice guidance.

Descriptive Validity in Criminology

General validity issues within criminology have been debated by researchers for many decades and include the development of mechanisms for reporting the internal and external validity of study information (e.g., Farrington 2003; Lösel and Köferl 1989; Sherman et al. 2002). However, descriptive validity in criminology is rarely reported, but has shown to be of great importance in RCTs conducted in healthcare. (e.g. Moher et al. 1998). Researchers have been keen to emphasize the extent to which a report provides information about the design, conduct, and analysis of the trial and its methodological quality (Moher et al. 1995). Equally as important are findings, which demonstrate the link between poor reporting of information and biased trial results (e.g. Hewitt et al. 2005).

In criminology, descriptive validity was first formally recognised by Farrington (2003), and more recently by Perry and Johnson (2008) and Perry et al. (in press). Descriptive validity in criminology is a critical concern for criminologists as they seek to develop evaluation studies of the highest standard (Perry et al. in press). For criminologists, the explicit reporting of study details are important because such information helps to maintain the methodological rigour and quality of reporting trial information. Farrington's recognition of descriptive validity was noted in his methodological evaluation of criminal justice research where he defines descriptive validity as "the adequacy of the presentation of key features of an evaluation in a research report" (p. 56). Farrington in his assessment of descriptive validity maintains that descriptive validity is important because without such key information it is difficult to assess the appropriateness of combining the results of trials or quasi-experimental studies in systematic reviews or meta-analyses where studies may differ in their methodological quality (see also Boruch 1997; chap. 10).

One other important element of descriptive validity is how descriptive validity is measured to provide an accurate picture of the adequacy of reporting within a particular study. Farrington suggests his 14 item checklist as one mechanism for studies within criminology (Farrington 2003). The 14 different elements include the design of the study, a description of

the characteristics of the experimental units and setting, information on sample sizes and attrition rates, hypotheses to be tested and the theories from which it was derived. Other factors for consideration also include operational definitions of the intervention and details about how it was implemented and a description of what the control and treatment group received. The final elements noted relate to the definition, reliability and validity of the outcome, the study follow up period, the reporting of effect sizes, confidence intervals and statistical significance levels, a description of how independent and extraneous variables were controlled for knowledge of the intervention and finally any conflict of interest issues, including the independence of the researchers and how the research was funded.

In healthcare, a widely accepted measurement of descriptive validity is the CONSORT Statement.

Measuring Descriptive Validity

In healthcare, the CONSORT Statement was published in 1996 to aid the transparent reporting of information in trials. Since its first publication, the CONSORT Statement has evolved with a number of subsequent revisions in 2001 and the Statement currently consists of two sections. Section one contains 22 items (Altman 2001; Moher et al. 2001b), which cover all aspects of reporting, including information across the title and abstract, introduction and background, methods (participants, interventions, objectives, outcomes, sample size, randomisation, blinding, and statistical methods), results (participant flow, recruitment, baseline data, numbers analysed, outcomes and estimation, ancillary analyses, and adverse events), interpretation, generalisability and overall evidence of a trial (see http://www.consort-statement.org/). Table 17.1 provides a detailed description of each checklist item.

The second section of the CONSORT Statement is a flow diagram. The CONSORT flow diagram is used to show the progression of participants through a trial and offers the authors an opportunity to standardise the methodology for the preparation of trial findings. At each stage in the trial (e.g., enrolment, intervention allocation, follow-up, and analysis) information is gathered for the purpose of transparency. Figure 17.1 shows the flow diagram outline, which allows readers to critically appraise and interpret the trial findings allowing the reader to assess the number of participants who did and did not complete the study enabling an assessment of the trial analysis (i.e., intention to treat or per protocol).

Although this concept is familiar to researchers in healthcare, there has been little use of the CONSORT Statement in criminology. As a consequence, we know little about the descriptive validity of criminal justice trials and can identify only two recent studies addressing this issue (Perry and Johnson 2008; Perry et al. in press). These two studies applied the CONSORT Statement to a number of RCTs reporting on aspects of good and poor descriptive validity in criminology. We use the findings of these two studies to identify areas of weak descriptive validity and examples of best practice from a range of criminal justice trials, which demonstrate how descriptive validity can be improved.

The first study of these two studies reports on the descriptive validity of a selection of RCTs focusing on the provision of mental health services for juvenile offenders (Perry and Johnson 2008). The second provides the results of a comprehensive evaluation of 83 criminal justice trials across a range of different areas of crime and justice (Perry et al. in press).

TABLE 17.1. The CONSORT checklist

CONSORT checklist item	CONSORT item	CONSORT checklist description
Title and abstract	1	How participants were allocated to interventions
Introduction and Background	2	Scientific background and explanation of the rationale
Methods Participants	3	Eligibility criteria for participants and settings and locations of the data collection.
Interventions	4	Details of the interventions intended for each group, and how and when they were administered.
Objectives	5	Specific objectives and hypotheses
Outcomes	6	Clearly defined primary and secondary outcome measures and methods used to enhance the quality of measurement.
Sample size	7	How the sample size was determined and when applicable an explanation of any interim analyses and stopping rules.
Randomisation	8	Method used to generate the random allocation sequence including any details of restrictions.
Sequence allocation and allocation concealment	9	Methods used to implement the random allocation sequence.
Implementation	10	Who enrolled participants and who assigned participants to their groups
Blinding	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.
Statistical methods	12	Statistical methods used to compare groups for primary outcome. Methods for additional analyses such as subgroup analyses and adjusted analyses
Results Participants flow	13	Flow of participants through each stage of the trial from enrolment to follow up
Recruitment	14	Dates defining the periods of recruitment and follow up
Baseline data	15	Baseline demographics and clinical characteristics of each group
Numbers analysed	16	Number of participants in each group included in each analysis and whether the analysis was by intention to treat.
Outcome and estimation	17	For each primary and secondary outcome a summary of results for each group and the estimated effect size and its precision
Anicillary analyses	18	Reports on any other adjustments performed, including subgroup analyses and adjusted analyses indicating those pre-determined and those exploratory analyses.
Adverse events	19	All important adverse events or side effects in each intervention group
Discussion Interpretation	20	Interpretation of results, taking into account study hypotheses, potential sources of bias and dangers associated with multiplicity of analyses and outcomes.
Generalisability	21	The external validity and generalisability of the trial results
Overall evidence	22	General interpretation of the results in the current context of evidence.

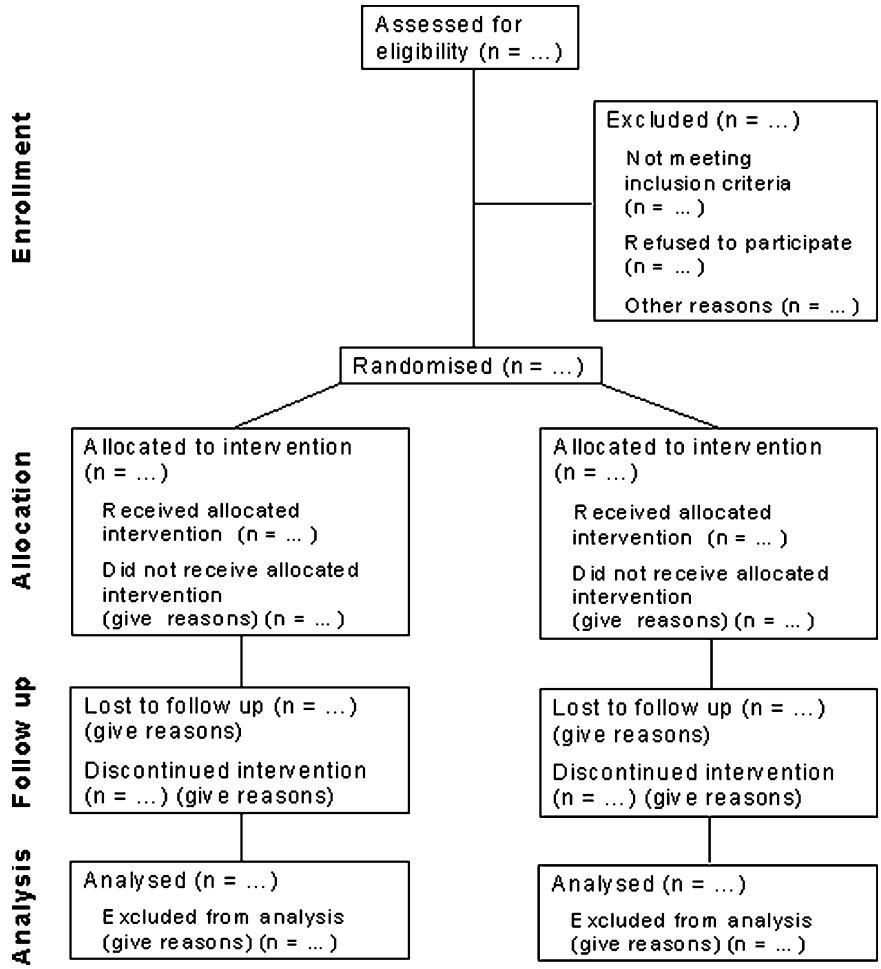

FIGURE 17.1. CONSORT flow diagram. Source: (see http://www.consort-statement.org/).

Using CONSORT to Measure Descriptive Validity in Criminology

The first study examined the application of the CONSORT Statement in a group of RCTs identified from a systematic review of mental health provision for juvenile offenders (Perry and Johnson 2008). This study applied the CONSORT Statement to a selection of RCTs to assess their descriptive validity, and the extent to which journal editors referenced the CONSORT Statement in the "instructions to authors" section when submitting articles to journals for publication.

The study identified a selection of 17 trials published between January 2001 and December 2006. These were identified using a total of 13 electronic databases,[1] grey literature databases

[1] PsychINFO, Applied Social Sciences Index and Abstracts, International Bibliography of the Social Sciences, Social Policy and Practice, Social Services Abstracts, Sociological Abstracts, ERIC, Criminal Justice Abstracts, PAIS International, C2 SPECTR, Swets-Wise, DARE and Cochrane Database of Systematic Reviews.

(SIGLE, Index of Theses and ZETOC) and ten key websites.[2] In addition, the Journal of Adolescence, and the Child and Adolescent Mental Health Journal were hand-searched for relevant papers which formed the basis of a larger systematic review, focusing on the provision of mental health services for juvenile offenders (Perry et al. 2008).

The study used two independent reviewers to extract relevant data systematically from each study using a modified version of the CONSORT Statement. The modified version was used to ensure that multiple aspects of single items on the CONSORT Statement were listed separately for the purpose of data extraction. This process resulted in a 42 item checklist. Each item on the CONSORT Statement was then rated as either "yes" (information reported) or "no" (no information reported).

The findings of the study identified studies with both high and low descriptive validity. Items of high descriptive validity generally included the less technical aspects of the study for example, reporting of the sample demographics at baseline. Other more technical aspects of descriptive validity were poorly reported. These included sample size estimations, outcome measures, effect sizes, and blinding. The study also revealed that 18% of the editors' instructions for authors contained details about the CONSORT Statement. This somewhat encouraging finding is particularly important especially considering that the CONSORT Statement has been rarely mentioned in criminology, and comparable findings have been identified in healthcare (Altman 2005).

The authors of the study do note a number of limitations with the study findings, which have an impact on how descriptive validity may be presented within criminology. Firstly, the selection of 17 trials represent only a small percentage of all trials conducted in the criminal justice system. Because of this reason, we cannot be sure that the trials are representative of other criminal justice trials conducted in the area. Secondly, the endorsement of the CONSORT Statement by journal editors in this specific sample of trials may be an artefact of the journals themselves. For example, many of the journals reported on both health and criminal justice issues and such journal may not be representative of trials published in criminology more broadly (e.g., in the courts, police, and prisons). As a consequence, these recognised limitations could affect how we draw conclusions about what this tells us about descriptive validity in criminal justice trials (e.g., external validity). In order to address these difficulties, the second study reports on the findings from a large-scale evaluation of descriptive validity in a representative sample of criminal justice trials (Perry et al. in press).

Descriptive Validity Across a Representative Sample of Criminology Research

The second study is aimed to explore the application of descriptive validity using the CONSORT Statement in a representative sample of RCTs. Three further objectives of the study assessed; (1) changes in reporting of transparent information over time, (2) the impact of sample size in relation to the number of CONSORT items reported, and (3) an assessment of the differences between the transparent reporting of information in trials with different types of interventions.

[2] Department of Health, Department of Education and Skills, Ministry of Justice, Joseph Rowntree Foundation, Royal College of Psychiatrists, Youth Justice Board, Policy Studies Institute, Mental Health Foundation, Young Minds and NACRO.

These three areas of interest were chosen because of previous evidence demonstrating that these factors have an impact on descriptive validity (e.g., Emerson et al. 1990). For our first issue, the idea of progression is thought to stem from the fact that methodological advances in writing and reporting of scientific manuscripts have improved naturally over time as standards of journals demand the highest quality research. As such, researchers anticipated that some improvement in the transparent reporting of trial information would be identified through the evaluation. (Graf et al. 2002; Prady et al. 2008; Shea et al. 2006).

Our second issue of sample size suggests that trials of larger sample sizes demonstrate greater descriptive validity with a general progression over time (e.g., Ioannidis et al. 1998; Kjaergard et al. 2001). This finding may be an artefaxt of the article length which will vary depending upon the type of publication (e.g., journal article or government report). Other differences have been found with different types of interventions. Reasons for such differences point towards the suggestion that some interventions may be easier to blind than others, and as a consequence, the reporting of some items (e.g., blinding and how it was achieved) may be easier to achieve in some studies than others. The final aspect of the study compared the reporting of descriptive validity in criminal justice and healthcare trials. This element of the study was used to assess whether the descriptive validity of criminal justice trials faired better or worse than those conducted in healthcare.

The study used a group of RCTs identified from a previous study conducted by Farrington and Welsh (Farrington and Welsh 2005). The 83 trials were published between 1982 and 2004 and were identified using four key criteria. The key criteria included: (1) studies where units (persons or places) were randomly assigned, (2) studies with conditions with at least 100 units assigned to two experimental groups, (3) studies reporting on measures of offending, and (4) studies published in English. Trials were grouped into different sections of the criminal justice system and included policing, prevention, corrections, courts and, community interventions.

Using similar methodology to the first study, each RCT was assessed using two independent reviewers and the CONSORT Statement was modified so that multiple items were listed separately. In this second study, the modification process also incorporated some additional aspects of Farrington's 14 item validity checklist (Farrington 2003). This was to ensure that any absent information (not present in CONSORT), but covered in the 14 questions was added to the tool. This process resulted in a 54 item coding tool. Additionally, the scoring key in the first study was expanded to include the "partial" reporting of information. This addition was necessary because some, but not all, information was often reported on each item of the CONSORT Statement. The additional "partial" scoring key therefore recorded instances where some aspects of the CONSORT Statement were reported.

In preparation for the analysis of the findings the items were divided into a number of different domains. These domains included a description of the study background and introduction (items, 1, 2a, 2b, 8 and 9), participant details (items, 3a, 3b, 13, 26, 31 and 32), data collection information (items, 4a, 4b, 4c, 29, 30 and 12), intervention and control group descriptions (items, 5a, 5b, 6a, 6b, 7a and 7b), the randomisation methodology (items, 16, 17, 18, 19, 20, 21, 22 and 23), information on outcome measures (items, 10, 11, 24, 25, 35, 36 and 38), the statistical analysis (items, 14, 15, 27, 28, 33, 34a, 34b, 37a and 37b), and study findings (items, 39, 40, 41, 42, 43, 44 and 45).

Figure 17.2 shows the overall results of each item reported for the relevant domain. Overall, the results of the study found similar findings to those presented in the first study with some areas being relatively well reported. For the background and introduction domain, all but two items scored above 50%. A similar pattern is shown in the participant and data collection domains. The remaining five domains demonstrate generally poor descriptive validity

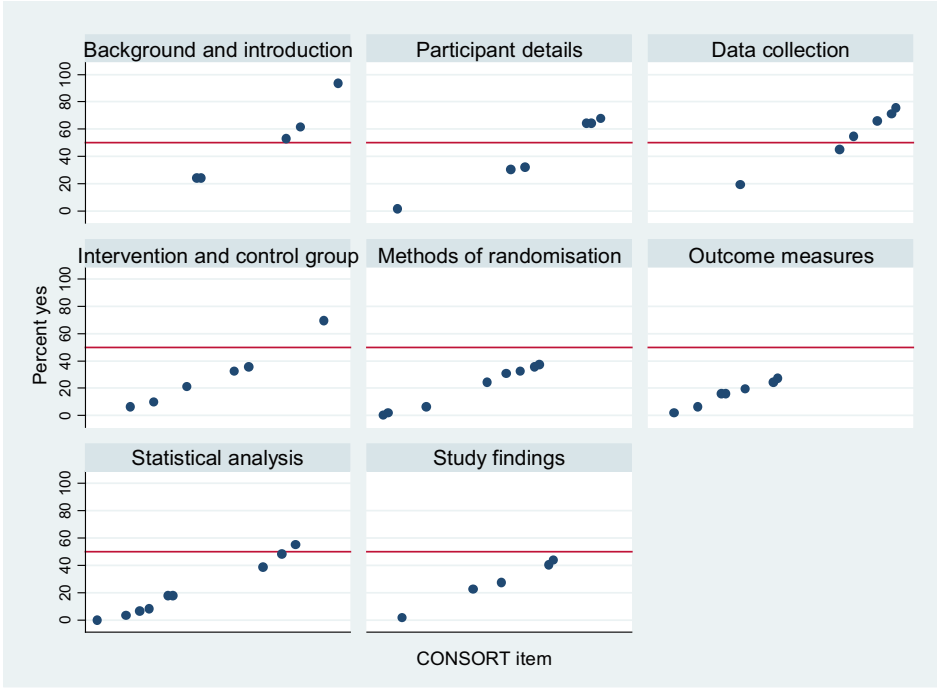

FIGURE 17.2. Findings of CONSORT in themed areas.

with nearly all items being reported in less than 50% of cases. Only two items on the intervention and control domain and one item on the statistical analysis domain report above the 50% threshold.

Contrary to other research evidence, the results of this study did not show a significant improvement in reporting over time (Prady et al. 2008; Shea et al. 2006). This is perhaps surprising given the expectation that methodological advances in other scientific discipline areas have shown improvement. (e.g., Graf et al. 2002).

The second evaluation assessed the impact of sample size and trial outcome. Previous research in this area has shown mixed results with some larger trials showing better reporting of trial information (e.g., Kjaergard et al. 1999), whilst other researchers have shown no association between sample size and trial outcome (e.g., Jüni et al. 2002). The findings from our evaluation of criminal justice trials therefore supports the latter of these two findings suggesting that there is no evidence to substantiate the relationship between the sample size of the trial and reporting on the CONSORT Statement.

The third evaluation assessed the impact of intervention differences on the application of the CONSORT Statement. The findings of our study revealed no significant differences between the different intervention types (although the policing interventions do far better than the others). The authors note this finding could be due to the fact that the same authors tend to report on their areas of research interest, and have a particular writing style and type of publication (e.g., journal article publications vs. reports for government). This effect may consistently have an impact on the quality and amount of information reported within a study design.

The final analysis of this second study compared the application of the CONSORT Statement to a range of other healthcare trials. Because the CONSORT Statement has been used for a number of decades in healthcare, the authors anticipated that the criminal justice trials may fair worse in comparison. Surprisingly, the results of the study showed that the reporting of descriptive validity was comparable with some items in criminal justice trials fairing better than healthcare (e.g., random allocation concealment). Other items were, however, significantly worse (e.g., blinding and the reporting of attrition). In conclusion, the authors of this study suggested that there was still room for overall improvement in the transparent reporting of information in criminal justice trials. Generally, the reporting of descriptive validity in criminal justice studies were found to be behind those conducted in healthcare in most areas of the CONSORT Statement, though in some aspects (e.g., random allocation) they seemed to fair better. The results from this study demonstrate how the reporting of descriptive validity in criminal justice trials information varies across the application of the CONSORT Statement. Such variability may be due to a number of different reasons. Firstly, the CONSORT Statement was initially developed for use with medical trials, and therefore aspects of its endorsement may not be applicable for research conducted in criminology. Secondly, these results may reflect a lack of attention paid by criminologists to the issue of descriptive validity. As a result, the quality of reporting information and hence the descriptive validity fall significantly behind those researchers who conduct medical trials in healthcare.

In summary, the evidence from these two studies seemed to suggest that criminologists report fairly well on some aspects of descriptive validity. Such areas tend to be the less technical aspects of the paper (e.g., study recruitment and aim of the study). Where criminologists faired significantly worse were on aspects of statistical information. The following section of this Chapter extends our review of this evidence to concentrate on examples of poor descriptive validity identified by the two previous studies (Perry and Johnson 2008; Perry et al. in press). Furthermore, it considers whether this lack of reporting is due to issues relating specifically to the study of criminology or whether such reporting is because the issue of descriptive validity is relatively new to criminologists.

Application of the CONSORT Statement in Criminal Justice Trials

The findings from the two criminology studies of descriptive validity provide some indication of which elements of descriptive validity are poorly endorsed by criminologists on the CONSORT Statement. As noted earlier, this may be for a number of different reasons. Firstly, the CONSORT Statement was originally developed for use in healthcare trials; it is, therefore, perhaps not surprising that none of the trials reviewed in the two studies reported on all items on the Statement. This concern leads to the question of whether the CONSORT Statement needs to be refined for use with evaluations of descriptive validity in criminal justice trials and/or whether criminologists need to improve the standard of reporting information to enable the accurate assessment of descriptive validity.

In both studies similar poorly endorsed items of descriptive validity were identified. These included the reporting of adverse events, blinding, reporting of primary and secondary outcome measures, descriptions about how the results were analysed (including intention to treat and per protocol approaches), reporting of information relating to the number of participants processed throughout the trial and details about the control group.

The following section of this chapter considers these poorly reported items of descriptive validity in more detail and evaluates whether these findings are an artefact of the type of

research conducted in criminology or whether the lack of emphasis on descriptive validity in criminology has resulted in researchers whose reporting of descriptive validity should be improved.

Descriptive Validity in Criminology and Reporting of Adverse Events

Unintended and often undesirable effects are rarely noted in criminology, but form an important element of the CONSORT Statement. The Statement advocates that authors examining the results of trial information need to be aware that an intervention can harm participants as well as providing benefits. In healthcare, anticipated adverse effects are most commonly associated with the side effects of taking a newly prescribed drug. In criminology, the concept of an adverse event is therefore somewhat different as few trials are interested in reporting on the impact of prescribed drugs. Instead, an adverse event in criminological research may focus on an increase (as apposed to a decrease) in recidivism rates following the delivery of an intervention.

Examples of such adverse events can be identified from the criminological literature with one of the most famous examples being the "Scared Straight" evaluation. The findings of this study resulted in an increase in the likelihood of criminal activity following young people being exposed to the program (Petrosino et al. 2002). Other considerations of adverse effects in criminology may also consider whether a particular intervention is deemed ethical, acceptable, and is of benefit to those undergoing the intervention.

Within our discussion of the two papers, few examples of adverse events were noted. This could be because the majority of these trials showed positive results, or it may suggest that adverse events are either not widely published by journal editors and/or considered by criminology researchers as important. There is, of course, the tendency for publishers to consider trials with positive results more favourably than those with negative results, and this has certainly been evidenced in healthcare (e.g., Egger and Smith 1998). This phenomenon is often referred to as publication bias where the published literature presents the findings of a particular perspective. Although we have no evidence for this in criminology, it could explain why a few researchers have addressed the issue of adverse events.

We did, however, find one trial which describes the unexpected findings of an intervention group experiencing higher recidivism rates than the control (see Table 17.2: Peters et al. 1997). An acceptance of such important negative findings should encourage criminologists and journal editors alike to present the findings of their research, which acknowledge the presence of any adverse events. The consequence of not reporting such information increases the risk of wasting valuable resources on unethical interventions, which have undesired effects on society.

Descriptive Validity in Criminology and the Reporting of Blinding and Masking

Our second area of poor descriptive validity in criminology is that of blinding or masking participants. In RCTs the term "blinding" refers to keeping the study participants, administrators delivering the intervention, and those collecting and analysing the data unaware of the assigned intervention (Boutron et al. 2005). This ensures that individuals are not unduly influenced by the knowledge of who is receiving the intervention. For descriptive validity, the reporting of blinding should recognise who was blinded (for example participants,

TABLE 17.2. Examples of descriptive validity and adverse events in criminology

CONSORT item number	CONSORT item description	Example from the literature
19	All important adverse events or side effects in each intervention group.	"In Cleveland, the question that must be asked is, why did the experimental groups do worse? This question calls for further investigation and leads to the second point: the findings presented here are from interim reports. The groups included in this study must be tracked for a longer period of time. It is entirely plausible that the control group in Cleveland will, over time, experience a recidivism rate comparable to that of the experimental group". (Peters et al. 1997).

TABLE 17.3. Examples of descriptive validity and blinding in criminology

CONSORT item number	CONSORT item description	Example from the literature
11a	Blinding or Masking: whether or not participants, those administering the interventions and those assessing the outcomes were blinded to group assignment	"The research design is a double-blinded baseline study. The program participants were divided into three groups; an experimental group which received acupuncture on a regular basis, a control group which did not receive acupuncture, and a placebo group which received an acupuncture-like simulation. It should be noted that initially all cases referred to the program were given the acupuncture treatment. Once the acupuncture study began, offenders were randomly placed into one of the three groups listed above. Neither staff nor clients were aware of group placement" (Latessa and Moon 1992).
11b	If blinding was done, how was the success of the blinding evaluated	"To evaluate participant blinding, a questionnaire asked participants which treatment they believed they had received. . .they were asked to indicate what led to this belief" (Adapted from Lao et al. 1999).

evaluators, or data analysts), and the mechanism of blinding. Blinding is, however, not always possible, but, a lack of blinding does not prevent authors from explaining why participants, administrators of the intervention or evaluators were not blinded.

Where blinding has been implemented the reporting of descriptive validity should also include information about whether such blinding was successful. In principle, if blinding was successful, the ability of participants to accurately guess their group assignment should be not better than chance. Our examples (see Table 17.3) provide a description of a double-blind study of acupuncture treatment for a group of offenders (Latessa and Moon 1992) and an evaluation of blinding using a questionnaire where participants were asked which treatment they believed they had received (Lao et al. 1999).

The standard expectation of "double blinding" occurs when both the evaluator and the participant are blind to the intervention, allocation, and assessment. Evidence from a systematic review assessing methods of blinding across different trials reports that blinding in the literature was most commonly represented in three different categories (Boutron et al. 2006). These categories included; (1) blinding with patients and care providers, (2) the maintenance of blinding, and (3) whether assessors were blind to the main outcomes. The results of the study revealed that blinding of outcome assessors depended upon the primary outcome. For

example, where mortality was the primary outcome, there was little point in blinding the outcome assessor. In other situations, they argue that blinding the outcome assessor should be possible with a centralized assessment of complementary investigation. Some examples of this blinding show how this is possible to achieve in different types of interventions (Boutron et al. 2006; Francis et al. 2003; Fiorucci et al. 2004).

In summary, blinding in criminal justice trials at least at the level of the outcome assessor should be feasible in all trial situations. This was certainly not something that was reported by criminologists. Other levels of blinding may be more difficult (that is not to say impossible) and innovative ideas about blinding in criminology research should be encouraged. This is particularly pertinent given that research from other areas shows how an absence of blinding can threaten the internal validity of a trial and its results (Boutron et al. 2005).

Descriptive Validity in Criminology and the Reporting of Primary and Secondary Outcome Measures

In criminological research most trials have several outcomes. Specifying which of these outcomes is the primary outcome of interest is not common place. However having more than one or two outcomes can incur problems of interpretation associated with multiplicity of analyses (see also CONSORT items 18 and 20).

Other information recorded about the outcome measures should include how they were measured and whether any steps were taken to increase the reliability of the measurements. This is often referred to as "enhancing the quality of the outcome measures". Such measures are more likely to be free from bias if the participant and assessor are blinded to group assignment, or where additional data is collected to support or corroborate other evidence (e.g., official data from criminal records is obtained to check self-report measurement of crime). Examples from two trials (see Table 17.4) provide some evidence to demonstrate the use of primary outcome measures and the enhancement of their quality in two criminal justice trials (Rhodes and Gross 1997; Robinson 1995).

Whilst these examples are found in the criminological literature, such examples are not in plentiful supply. In comparison, medical trials have well established protocols that require

TABLE 17.4. Examples of descriptive validity and primary and secondary outcome measures in criminology

CONSORT item number	CONSORT item description	Example from the literature
6a	Clearly defined primary and secondary outcome measures	"While recidivism remains the primary outcome measure in studies of program effectiveness, the impact of program participation on the granting of conditional release is also of interest" (Robinson 1995)
6b	When applicable enhance the quality of outcome measures	"Outcomes were evaluated using formal assessment instruments to measure self-reported behaviour at baseline and again at 3 and 6 months. Independent data from criminal justice and drug treatment systems were analysed to gauge the validity of these self reports" (Rhodes and Gross 1997).

pre-determined primary outcome measures to be specified in the trial protocol prior to commencement of the analysis (Freemantle 2001). Such specification seeks to ensure a reduction in a number of potential biases. Type I error is one particular problem for trials which include many outcome measures because there is an increased possibility of the results occurring simply by chance. Pre-determined outcome measures also reduce the temptation for researchers to potentially "trawl" for positive results when these are not indicated by the initial analysis (see Freemantle 2001 for further discussion).

In contrast, criminological studies tend to clearly state the outcome measures, but do not make the distinction between primary and secondary outcomes. For example, a trial conducted by Needles and colleagues reports on a range of outcome measures (including self-reported drug use, cocaine and hair test results, marijuana hair test results, physical and social problems relating to alcohol or drug use, participation in drug treatment programmes and sexual behaviour) without specifically listing the primary and secondary measures (Needles et al. 2005). Such reporting of outcomes in criminology is common and many studies evaluate the impact of multiple outcome measures and analyses running the risk of introducing Type I error.

In conjunction with the primary outcome measure is the calculation of the sample size. Many trials in healthcare and education have been noted as being too small to be able to detect differences between the intervention and control groups (Torgerson and Elbourne 2002). In order to estimate the sample size in a trial, a number of different elements are required. These include (1) the estimated outcomes in each group, (2) the alpha (type I) error level, (3) the statistical power (type II) error, and (4) for continuous outcomes (e.g. mean number of offences), the standard deviation or standardised effect size of the measurements.

Sample size estimates are important in trial design because a study needs to be large enough to have high probability (power) of detecting a statistically significant difference of a given size (e.g., the anticipated percentage reduction in the number of reconvictions). The size of the effect is related to the size of the sample, so large samples are necessary to detect small differences and vice versa. In criminology, the reporting of sample size calculations are rare. This may be because such calculations are not conducted by researchers or that such methodological requirements are not commonly considered important enough by criminologists to report. Either way, criminologists should include sample size calculations within their study descriptions. Such calculations when conducted prior to the start of the study should help identify those studies, which may be inadequately powered from the outset.

Descriptive Validity in Criminology and the Reporting of Statistical Analyses

The reporting of statistical analyses was generally poor across the two criminological studies. Basic information about the number of participants in each group or arm of a trial provides important information about why individuals may fail to complete an intervention and what may bias the results of the trial. Primarily, two approaches to the analysis of trial information are taken. Intention To Treat (ITT) analysis includes all the study participants regardless of whether they "crossed over" into another intervention or did not complete the intervention. Supporters of this view argue that this methodology provides a conservative estimate of the results by truly reflecting what happens to individuals receiving a treatment or intervention in practice (Torgerson and Torgerson 2008). The logic of this argument suggests that by including all participants (and not just necessarily those who are most likely or most highly

TABLE 17.5. Examples of descriptive validity and intention to treat analyses in criminology

CONSORT item number	CONSORT item description	Example from the literature
16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat".	"An intent-to-treat analysis plan was followed in which each individual who participated in the follow up was analyzed according to his or her original per-school ($N = 104$) or school-age group ($N = 95$) random assignment, regardless of the length of exposure. This has the advantage of increased stringency while increasing detection power by increasing the number of individuals available for analysis. Post hoc analyses were then conducted with data from five individuals originally assigned to the pre-school treatment group removed from the sample." (Campbell et al. 2002)

motivated to complete treatment) in the analysis this method reduces any possibility of bias. In contrast, some trialists try to correct for non-adherence by using a per protocol analysis. This method can produce a biased estimate of effect as participants not complying with the treatment protocol are excluded from the analysis. An example of an ITT description from criminology (see Table 17.5) shows a trial of school children (Campbell et al. 2002).

The fundamentals of this analysis are central to the design of a trial and are equally apparent regardless of whether the trial is conducted in healthcare or criminology. This lack of transparent reporting in criminological trials may therefore suggest an inherent flaw of this important element of trial design. This is therefore something that criminologists need to pay greater attention to in the reporting of trial information.

Descriptive Validity in Criminology and the Reporting of Participant Numbers

The number of participants who started and finished a trial in each arm of the design provides important information on how many people were used in the data analysis. This information is important because participants who were excluded after allocation are unlikely to be representative of all participants in the study. For example, participants who do not complete all the data evaluation points may do so because they may be less motivated or could be distinctly different (i.e., commit different sorts of crimes) than those that do not. For healthcare, this is usually represented by using the CONSORT flow chart (see Fig. 17.1). The Statement advocates that for each group authors should report the number of participants randomly assigned, receiving intended treatment, completing the study protocol and analysed for the primary outcome.

Only 2 of our 83 criminal justice trials used any form of this flow chart. This is perhaps not surprising since the CONSORT Statement in criminology has not been widely disseminated, nor has it been adopted by the majority of journal editors as a requirement of authors' submission of research reports for publication. Researchers reporting on the findings of trials in criminology should therefore be encouraged to be more transparent about the reporting of participants throughout all stages of their trials.

Descriptive Validity in Criminology and the Reporting of Control Group Information

Our final area of concern in the reporting of descriptive validity in criminology relates to the description of the intervention and control group(s). In both the control and intervention group(s) it is important to provide information on the details of the intervention received. These details should include information about who administered the intervention. For example, with cognitive behavioural interventions it may be necessary to describe the number, training and experiences of the staff. Additionally, the details of the timing and duration of both the intervention and control groups are required. This is particularly important if the intervention has multiple components as it allows the study to be replicated (e.g. a therapeutic community, followed by aftercare in the community and a work release program). The four examples (see Table 17.6) provide information about an intervention group receiving multi-systemic therapy in comparison to a group of offenders receiving individual therapy (Borduin et al. 1995). The final example provides information about the timing and duration of a Big Brother Big Sister evaluation (Grossman and Tierney 1998).

TABLE 17.6. Examples of descriptive validity and the reporting of control and intervention information in criminology

CONSORT item number	CONSORT item description	Example from the literature
4	The precise details of the interventions intended for each group and how and when they were actually administered	"Therapeutic interventions were based on the multisystemic approach to the treatment and prevention of behavior problems. . . Using interventions that are present-focused and action orientated, MST directly addresses intrapersonal and systemic factors that are known to be associated with adolescent antisocial behavior. . .." (Borduin et al. 1995). "All offenders in this condition received individual therapy that focused on personal, family and academic issues. . . . Their theoretical orientations were an eclectic blend of" . . .. (Borduin et al. 1995). "MST (multisystemic therapy) was provided by three female and three male graduate students (ages ranged from 23 to 31 years, $M = 26$) in clinical psychology. One of the therapists was Native American, and the others were white. Each had approximately 1.5 years of direct clinical experience with children or adolescents before the study. The six therapists served in the study for an average of 16 months (range 12–24 months). Therapist supervision was provided by Charles M. Borduin in a 3-h weekly group meeting and continued throughout the course of the investigation. During these meetings..." (Borduin et al. 1995). "The average length of match for those treatments who had been matched was almost 12 months, with white girls having met with a Big Sister for the longest period (12.4 months) . . ..Little Brothers and little Sisters met with their Big Brother and Big Sister on a regular basis. More than 70% of the youths met. . . at least three times a month, and approximately 45% met one or more times per week. An average meeting lasted 3.6 h" (Grossman and Tierney 1998).

These examples demonstrate best practice evidence giving the amount of detail that should be reported on both the intervention and control groups. In the majority of criminal justice trial information was lacking about the control group. Often, for example, phrases were used that included "care as usual", without describing what this meant and how much was received by the control group.

SUMMARY

Descriptive validity and validity more broadly is one of the most important aspects of research design as it ensures the implementation of sound methodology from which conclusions can be drawn and generalised to other research findings (Campbell 1957; Perry et al. in press). In criminology, the concept of descriptive validity has been around for a few years (e.g., Farrington 2003), but has received greater recognition in healthcare (e.g., Altman 2005). The key driving force for this interest in healthcare was the development of the CONSORT group and subsequently the CONSORT Statement in 1996 (Altman 2001). Since then, the CONSORT Statement has been internationally adopted by many journal editors and evaluations of the CONSORT Statement have shown how it has improved the quality of reporting in RCTs over time (e.g., Plint et al. 2006).

The impact of poor descriptive validity has been conducted on three key areas of interest. These include an assessment of whether the introduction of the CONSORT Statement has increased the reporting of transparent information (e.g., Moher et al. 2001a), the extent to which journal editors have adopted the CONSORT Statement (e.g., Devereaux et al. 2002), the relationship between the impact of transparent reporting on allocation concealment in the randomization process (Hewitt et al. 2005), selection bias (Moher et al. 1998), performance bias (Jüni et al. 2000), blinding (Schulz et al. 1995), sample size (Ioannidis et al. 1998), and different types of interventions (Emerson et al. 1990). The CONSORT Statement in healthcare is therefore an established fundamental and credible tool that is used to assess the extent of transparent reporting in trials.

In criminal justice trials, the reporting of descriptive validity varies across the CONSORT Statement. The majority of CONSORT items cover the essential requirements of trial design, methodology, results, and trial findings common to all RCTs regardless of the subject matter. Other aspects of the CONSORT Statement may need some adaptation (e.g. the impact of adverse events). Indeed adoption of the CONSORT Statement in discipline areas has occurred in education (Torgerson et al. 2005) and non-pharmocological trials (Boutron et al. 2005). There is therefore no reason to suggest that the CONSORT Statement cannot be used (with some minor amendments) with trials conducted in the criminal justice field.

The other assumption that can be drawn from these findings is that criminology and the reporting of descriptive validity has not been acknowledged and as a consequence, the transparent reporting of trial information appears to be worse than in other disciplines such as healthcare. Use of a CONSORT like Statement in criminology would provide a useful framework for researchers not only in the reporting of good descriptive validity, but also in the best practice methodology for trial design. Any such adoption of the CONSORT Statement by criminal justice researchers is likely to enhance the quality of reporting information of trials and should be used to aid the interpretation of evidence.

Our discussions surrounding the application of the CONSORT Statement in criminology raise a number of issues. The central question to the adoption of a CONSORT-like Statement

in criminology is whether such a tool would increase the transparent reporting of trial information. Evaluations to assess whether the CONSORT Statement has improved the reporting of trial information over time have been systematically reviewed in an evaluation of studies that compared CONSORT-adopting and non-adopting journals after the publication of CONSORT; CONSORT adopters before and after the publication of CONSORT, or a combination of both the listed criteria (Plint et al. 2006). The results of the systematic review showed that journals endorsing the CONSORT Statement reported significantly more information about methods of sequence allocation, allocation concealment, and the overall number of CONSORT items than did journal editors who had not endorsed the Statement.

Overall, the results of the study concluded that the CONSORT Statement was associated with some improvements in the reporting of transparent information although, clear endorsement of the Statement by journal editors in the "instructions to authors" has shown more varied results (Altman 2005; Hopewell et al. 2008).

Other researchers have also investigated the impact of reporting transparent information on different aspects of the CONSORT Statement. Research evaluating specific aspects of descriptive validity have found mixed results. For example, researchers evaluating the impact of reporting inadequate details of allocation concealment found exaggerated intervention effects compared with trials reporting adequate allocation concealment (Moher et al. 1998, 1999; Schulz et al. 1995).

Further work on this issue was also conducted by Juni and colleagues (2001) who showed that trials with inadequate or unclear concealment produced odds ratios that were, on average, 30% lower (more beneficial) than those with adequate methodology (combined odds ratio 0.70, 95% CI: 0.62–0.80). However, other researchers have found no association between reported allocation concealment and intervention effects (Emerson et al. 1990).

Other debates in the literature around validity and reporting bias include the impact of trial size and trends in the improvement of reporting methodology over time. The size of the trial has shown discrepancies between the results of several small trials in meta-analyses (e.g., Ioannidis et al. 1998). Reasons for such differences point towards the fact that small trials are more likely to be published if they show a statistically significant intervention effect. In such cases, these discrepancies may be due to publication bias (e.g., Egger and Smith 1998). Other researchers have shown how trials with greater sample sizes show greater descriptive validity (Kjaergard et al. 1999; Prady et al. 2008; Shea et al. 2006). This effect does, however, seem to be specific to the type of intervention. For example, Jürgen and colleagues in 2002 found improvement in trials with mortality outcomes whereas surrogate outcome trials did not demonstrate an improvement in methodological quality over time.

To summarise, use of the CONSORT Statement in healthcare has generally improved the descriptive validity of trials. Specific evaluations of different aspects of the CONSORT Statement have revealed mixed results which may or may not demonstrate specific links between descriptive validity and bias within trial interpretation and design. Several aspects of the CONSORT Statement have, however, been used in groups of educational trials and has recently been referred to in criminology (Perry et al. in press; Torgerson et al. 2005).

For criminologists, descriptive validity is an important concept because it ensures that the methodological rigour and reporting of trial information is of the highest standard. As a consequence, one of the limitations of descriptive validity is its reliance upon an accurate and full description of the trial details (Moher et al. 2005). This is particularly important with the increasing use of meta-analytical techniques and systematic review evidence which seeks to synthesise different studies together to ascertain an overall effect. Without this information, a

well conducted RCT or quasi-experimental study may have low descriptive validity because the authors have provided a poor description.

Research investigating the use of the CONSORT Statement in criminology found mixed results with some aspects of reporting highly endorsed but others severely lacking in the criminological literature. To some extent, it is difficult to ascertain why such differences might occur and suggestions for these differences might include an inability to acknowledge the importance of descriptive validity in criminological literature. This lack of acknowledgement does not, however, excuse the poor aspects of reporting on many of the more technical aspects of study design (e.g., calculation of sample sizes). What faces criminologists is a challenge to raise their game in the reporting of full transparent and accurate information, thus increasing the standard of research within the discipline. The transparent reporting of information does not stop at studies only using RCT designs, and such issues are equally of concern for quasi-experimental and observational study designs. Whilst the CONSORT Statement has been specifically designed to focus on aspects of trial design (e.g., random allocation), other aspects of the CONSORT Statement are also applicable to other types of study designs (e.g., the aim and recruitment of participants into the study).

It is crucial that RCTs are reported in a transparent manner to ensure that judgements can be made about the internal and external validity of the study. We would therefore recommend that the concept of descriptive validity be supported by journal editors and researchers within criminology to ensure that methodological rigour is upheld to the highest standards in RCTs and other quasi-experimental study designs.

REFERENCES

Altman DG (2001) The revised CONSORT Statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 134:663–694

Altman DG (2005) Endorsement of the CONSORT statement by high impact medical journal: a survey of instructions for authors. Br Med J 330:1056–1057

Borduin CM, Mann BJ, Cone LT, Henggeler SW, Ficci BR, Blaske DM et al. (1995) Multisystemic treatment of serious juvenile offenders: long term prevention of criminality and violence. J Consult Clin Psychol 63:569–578

Boutron I, Moher D, Tugwell P, Giraudeau B, Poiraudeau S, Nizard R, Ravaud P (2005) A checklist to evaluate a report of a nonpharmocological trial (CLEAR NPT) was developed using consensus. J Clin Epidemiol 58(10):1233–12240

Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A et al. (2006) Methods of blinding in reports of randomised controlled trials assessing pharmacologic treatments: a systematic review. PLoS Med 3(10):1931–1939

Boruch RF (1997) Randomized field experiments for planning and evaluation: a practical guide. Sage Publications, Newbury Park, CA

Campbell DT (1957) Factors relevant to the validity of experiments in social settings. Psychol Bull 54:297–312

Campbell DT & Stanley JC (1963) Experimental and quasi-experimental designs for research on teaching. In: Gage NL (ed) Handbook of research on teaching. Chicago, Rand McNally, 171–246

Campbell FA, Ramey CT, Pungello E, Sparling J, Miller-Johnson S (2002) Early childhood education: young adult outcomes from the Abercedarian project. Appl Dev Sci 6:42–57

Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH (2002) The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. Control Clin Trials 23:380–388

Egger M, Smith GD (1998) Bias in location and selection of studies. Br Med J 316:61–66

Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC (1990) An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Control Clin Trials 11:339–352

Farrington DP (2003) Methodological quality standards for evaluation research. Ann Am Acad Pol Soc Sci 587:49–68

Farrington DP, Welsh BC (2005) Randomized experiments in criminology: what have we learned in the last two decades? J Exp Criminol 1(1):9–38

Fiorucci S, Mencarelli A, Lechi A, Renga B et al. (2004) Co administration of nitric oxide aspirin prevents platelet and monocyte activation and protects against gastric damage induced by aspirin in humans. J Am Coll Cardiol 44:635–641

Francis CW, Berkowitz SD, Comp PC, Lieberman JR Ginsberg JS et al. (2003) Comparison of ximelagatran and warfarin for the prevention of venous thromboembolism after total knee replacement. N Engl J Med 349: 1703–1712

Freemantle N (2001) Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? Br Med J 322(7292):989–991

Graf J, Doig GS, Cook DJ, Vincent JL, Sibbald WJ (2002) Randomized controlled trials in sepsis: has methodological quality improved over time? Crit Care Med 30(2):461–472

Grossman JB, Tierney JP (1998) Does mentoring work? An impact study of the Big Brothers Big Sisters Program. Eval Rev 22:403–426

Hewitt C, Hahn S, Torgerson DJ, Watson J, Bland MJ (2005) Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. Br Med J 330:1057–1058

Hopewell S, Altman DG, Moher D, Schulz KF (2008) Endorsement of the CONSORT statement by high impact factor medical journals: a survey of journal editors and journals 'Instructions to Authors'. Trials 9:20

Ioannidis JP, Cappelleri JC, Lau J (1998) Issues in comparisons between meta-analyses and large trials. JAMA 279:1089–1093

Jüni P, Altman DG, Egger M (2001) Assessing the quality of controlled clinical trials. In: Egger M, Davey-Smith G, Altman DG (eds) Systematic reviews in health care: meta-analysis in context, 2nd ed. BMJ Books, London

Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M (2002) Direction and impact of languate bias in meta analyses of controlled trials: empirical study. International Journal of Epidemiology 31:115–123

Jüni P, Tallon D, Egger M (2000) 'Garbage in – garbage out'? Assessment of the quality of controlled trials in meta-analyses published in leading journals. In: Proceedings of the 3rd symposium on systematic reviews: beyond the basics. St. Catherines College, Centre for Statistics in Medicine, Oxford

Kjaergard LL, Villumsen J, Gluud C (1999) Quality of randomised clinical trials affects estimates of intervention efficacy. In: Proceedings of the 7th Cochrane Colloquium. Universitia S. Tommaso D'Aquino Rome. Milan, Centrro Cochrane Italiano, poster B10

Kjaergard LL, Nikolova J, Gluud C (1999) Reported methodological quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 135:982–989

Kjaergard LL, Villumsen J, Gluud C (2001) Reported methodological quality and discrepancies between large and small randomized trials in meta-analyses. Annals of Intern Medicine 135:982–989

Lao L, Bergman S, Hamilton GR, Langenberg P, Berman B (1999) Evaluation of acupuncture for pain control after oral surgery. Arch Otolaryngol Head Neck Surg 125:567–572

Latessa E, Moon (1992) The effectiveness of acupuncture in an outpatient drug treatment program, Journal of Contemporary Criminal Justice 8(4):317–331

Lösel F, Köferl P (1989) Evaluation research on correctional treatment in West Germany: a meta-analysis. In: Wegener H, Lösel F, Haisch J (Eds) Criminal behavior and the justice system: psychological perspectives. Springer-Verlag, New York, pp 334–355

Moher D, Jadad AR, Nichol G Penman M, Tugwell M, Wash S (1995) Assessing the quality of RCT: an annotated bibliography of scales and checklists. Control Clin Trials 16:62–73

Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP (1998) Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 352:609–613

Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. (1999) Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 3(i–iv):1–98

Moher D, Schulz KF, Altman D, for the CONSORT Group (2001a) The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 285:1987–1991

Moher D, Jones A, Lepage L (2001b) Use of the CONSORT statement and quality of reports of randomized trials: a comparative before and after evaluation. JAMA 285:1992–1995

Moher D, Altman D, Schulz K, for the CONSORT Group. The CONSORT statement: revised recommendation for improving the quality of reports of parallel-group randomised trials [Chinese]. Chin J Evid-Based Med 2005; 5(9):702–707

Needles K, James-Burdumy S, Burghardt J (2005) Community case management for former jail inmates: its impacts on re-arrest, drug use and HIV risk. J Urban Health 82(3):420–431

Perry AE, Johnson M (2008) Applying the Consolidated Standards of Reporting Trials (CONSORT) to studies of mental health provision for juvenile offenders: a research note. J Exp Criminol 4:165–185

Perry AE, Gilbody S, Akers J, Light K (2008) Access and provision of services for young people with mental health needs. Youth Justice Board, England and Wales

Perry AE, Weisburd D, Hewitt C (in press) Are criminologists reporting experiments in ways that allow us to assess? J Exp Criminol

Peters M, Thomas D, Zanberlon C (1997) Boot camps for juvenile offenders. Office of Juvenile Justice and Delinquency Prevention (Program Summary, Washington, DC

Petrosino A, Turpin-Petrosino C, Buehler J (2002) Scared straight and other juvenile awareness programs for preventing juvenile delinquency: a systematic review of randomized experimental evidence. Campbell Collaboration, USA

Plint AC, Moher D, Morrison A, Schulz K, Altman D, Hill C, Gaboury I (2006) Does the CONSORT checklist improve the quality of reporting randomized controlled trials? A systematic review. Med J Aust 185(5):263–267

Prady SL, Richmond SJ, Morton VM, MacPherson H (2008) A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. PLos One 3(2):1–10

Rhodes W, Gross M (1997) Case management reduces drug use and criminality among drug-involved arrestees: an experimental study of an HIV prevention intervention (PDF Version). Washington, D.C.: National Institute of Justice, U.S. Department of Justice

Robinson D (1995) The impact of cognitive skills training on post release recidvisim among Canadian federal offenders. Correctional Service of Canada, Ottowa, Research Report R-41

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273:408–412

Shea B, Boers M, Grimshaw JM, Hamel C, Bouter LM (2006) Does updating improve the methodological and reporting quality of systematic reviews? BMC Med Res Methodol 6:27

Sherman LW, Farrington DP, Welsh BC, Mackenzie DL (eds) (2002) Evidence-Based Crime Prevention. Routledge, London

Torgerson CJ & Elbourne D (2002) A systematic review and meta analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. J Res Read 35:129–143

Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J (2005) A comparison of RCTs in health and education. Br Educ Res J 31(6):761–785

Torgerson CJ, Torgerson DJ (2008) Designing and running randomised trials in health and the social sciences, Palgrave, MacMillan