YOUTH
ENDOWMENT
FUND

# Technical Guide

Version 4-1 (December 2021)

Campbell Collaboration

UNIVERSITY OF CAMBRIDGE
Institute of Criminology
Violence Research Centre

CLARE HALL
CAMBRIDGE

# Table of contents

## Changes in the December update

The December update included some minor changes to the evidence security rating. These changes are designed to better communicate the security of the evidence at the lower end of the rating.

- Previously, the evidence rating used a five-point scale. Topics could receive a rating of between one and five magnifying glasses.
- Topics can now receive a rating of zero magnifying glasses. This means the rating now uses a six-point scale, from zero out of five to five out of five. The zero rating is applied to topics where we lack a meta-analysis.

# Overview of the Toolkit

## Aims

The YEF Toolkit is a free-to-access, online summary of research on interventions which could prevent children and young people (CYP) getting involved in violent crime. It aims to make research findings accessible and easy-to-understand for a non-academic audience. It gives an overview of the 'best bets' for reducing violence – an overall picture of the approaches that are most likely to succeed, given the available research.

The Toolkit will ensure that research findings are:

1. Available. The research is spread over different disciplines and journals. The Toolkit brings it together in one place and make it easily available.
2. Accessible. It presents findings without jargon and in plain English.
3. Actionable. It focuses on the practical implications of research findings.

## Toolkit structure

The Toolkit has three levels (see Figure 1).

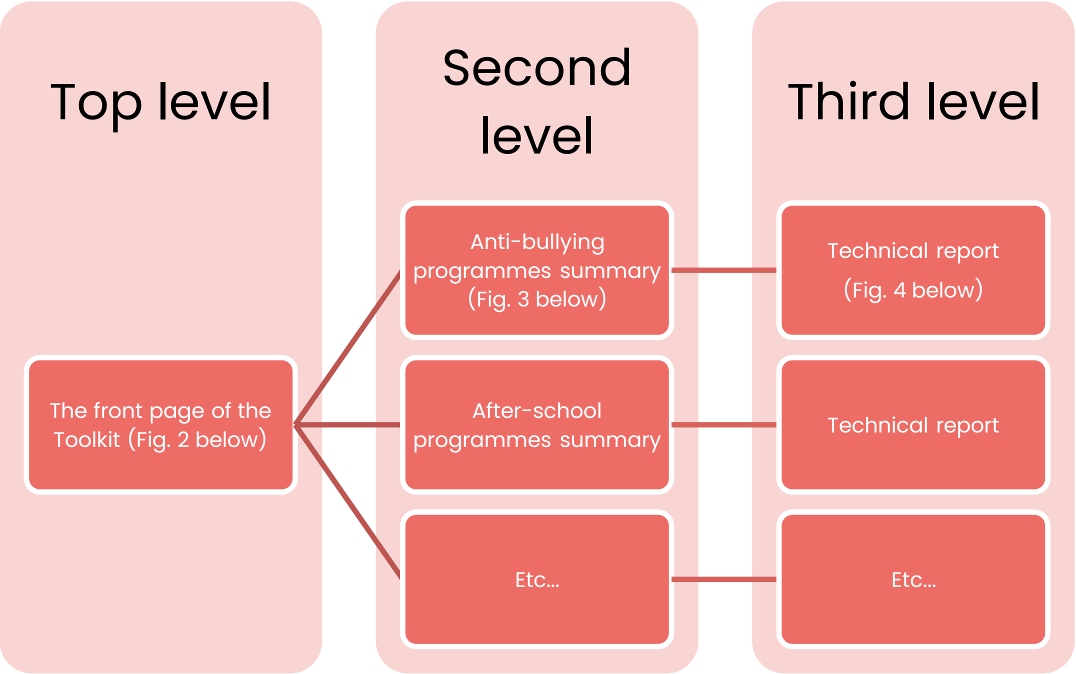
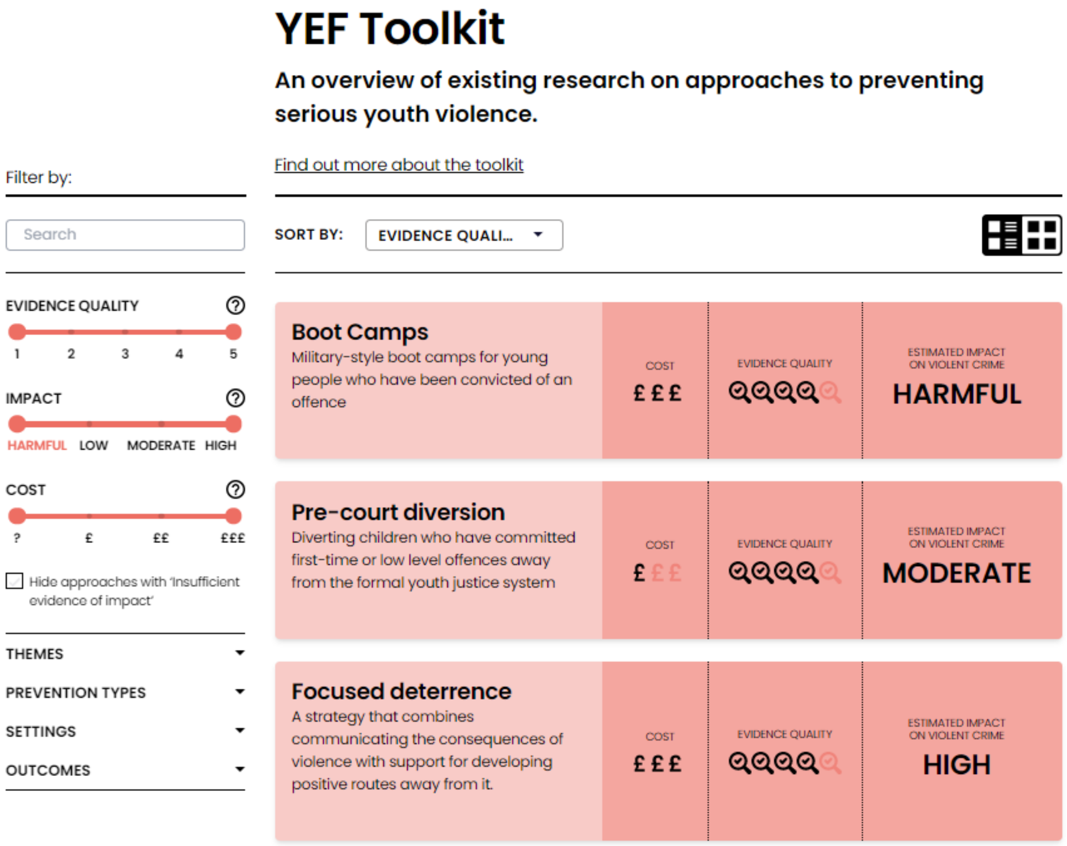

Top level
Second level
Third level
The front page of the Toolkit (Fig. 2 below)
Anti-bullying programmes summary (Fig. 3 below)
After-school programmes summary
Etc...
Technical report (Fig. 4 below)
Technical report
Etc...

***Figure 1 The Toolkit's structure***

## The top level

The top level or 'front page' of the Toolkit provides an overview of the impact and cost of approaches, and the security of the relevant research. The top-level lists different approaches to preventing children becoming involved in violent crime. For each approach, the Toolkit displays three ratings.

- An impact rating. The impact rating conveys the average impact of each approach, according to the research. Our impact estimate comes from the estimated effect from a systematic review of the global evidence for the approach. That is, the estimate is based on what we call the 'most appropriate effect size estimate from the most relevant and reliable review'. YEF will be commissioning additional systematic reviews in order to add additional topics to the Toolkit.
- An evidence rating. This rating conveys our confidence in the security of the research used to arrive at the impact rating. It uses a five-point scale (very low to very high).
- A cost rating. This rating provides an initial indication of how much the approach might cost, on average, compared to other approaches in the Toolkit. It uses a three-point scale (low, medium, high).



YEF Toolkit

An overview of existing research on approaches to preventing serious youth violence.

Find out more about the toolkit

□ Hide approaches with 'Insufficient evidence of impact'

| Approach | COST | EVIDENCE QUALITY | ESTIMATED IMPACT ON VIOLENT CRIME |
|---|---|---|---|
| **Boot Camps**<br>Military-style boot camps for young people who have been convicted of an offence | £££ | 4 of 5 | HARMFUL |
| **Pre-court diversion**<br>Diverting children who have committed first-time or low level offences away from the formal youth justice system | £ | 4 of 5 | MODERATE |
| **Focused deterrence**<br>A strategy that combines communicating the consequences of violence with support for developing positive routes away from it. | £££ | 4 of 5 | HIGH |

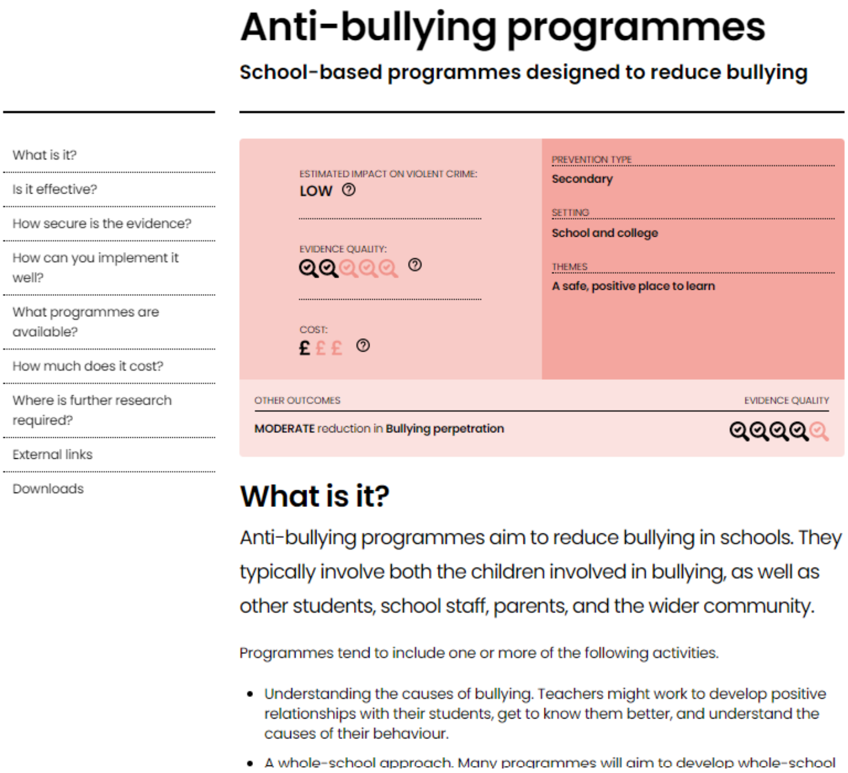
***Figure 2 The top level of the Toolkit***

## The second level

Clicking on any of the approaches displayed on the top level of the Toolkit will take the user to the second level. The second level provides a more detailed summary of the research for each Toolkit topic. For example, Figure 3 shows a screenshot of the summary for Anti-bullying programmes. The summaries present the following information:

- What is it? A detailed description of the approach, its core components and how the approach activities might vary.
- Is it effective? A more detailed description of the average impact of the approach and how the impact might vary.
- How secure is the evidence? A brief justification for the evidence rating and description of the security of the evidence base.
- How can you implement it well? A summary of the evidence from narrative thematic synthesis of process evaluations, preferably from the UK and Ireland.

- What programmes are available? Links to relevant programmes in the EIF Guidebook.
- How much does it cost? A description of the data used to create the cost rating and how the cost will vary.
- Topic summary. An overall summary of the research on the approach.



# Anti-bullying programmes

**School-based programmes designed to reduce bullying**

- What is it?
- Is it effective?
- How secure is the evidence?
- How can you implement it well?
- What programmes are available?
- How much does it cost?
- Where is further research required?
- External links
- Downloads

ESTIMATED IMPACT ON VIOLENT CRIME:
LOW

EVIDENCE QUALITY:
●●○○○

COST:
£££

PREVENTION TYPE
**Secondary**

SETTING
**School and college**

THEMES
**A safe, positive place to learn**

| OTHER OUTCOMES | EVIDENCE QUALITY |
|---|---|
| **MODERATE** reduction in **Bullying perpetration** | ●●●●○ |

## What is it?

Anti-bullying programmes aim to reduce bullying in schools. They typically involve both the children involved in bullying, as well as other students, school staff, parents, and the wider community.

Programmes tend to include one or more of the following activities.

- Understanding the causes of bullying. Teachers might work to develop positive relationships with their students, get to know them better, and understand the causes of their behaviour.
- A whole-school approach. Many programmes will aim to develop whole-school

***Figure 3 An example of a summary from the second level of the Toolkit***

## The third level

The third level of the Toolkit consists of 'technical reports' written by the Toolkit evidence review team. The team wrote a separate technical report for each topic included in the Toolkit. Technical reports can be downloaded from the bottom of any summary in the second level. The technical report describes the research used to write the summary for the second level and create the ratings shown on the top level.

## About this technical guide

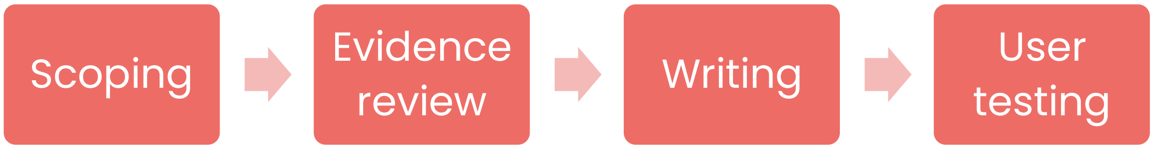
This technical guide outlines our approach to developing the Toolkit. It describes:

- An overview of the development process.
- How we created the technical reports, which are the foundation for the YEF Toolkit.
- How we created the headline ratings for impact, evidence and cost.

The guide documents the choices we made in how the Toolkit is constructed and presented. We describe how we select the evidence which is presented in the Toolkit, how we rate that evidence, and how we calculate and report effectiveness and cost on the front page (top-level) of the Toolkit.

## How did we create the Toolkit?

The Toolkit was developed over a four-stage process.



### Scoping

The project began with a scoping phase which aimed to answer some initial questions:

- Who is the primary audience for the Toolkit?
- What is the audience looking for in an evidence tool? How can we ensure the Toolkit is useful?
- What topics should the Toolkit cover?

During the scoping phase the YEF spoke to hundreds of stakeholders to better understand their needs and views.

### Evidence review

In the next phase the YEF conducted a competitive tender to appoint a team to carry out an evidence review. The YEF appointed a team from the Campbell Collaboration (led by Howard White) and the University of Cambridge (Prof. David Farrington and Dr Hannah Gaffney).

The evidence review team searched for and summarised existing systematic reviews in the Evidence and Gap Map. The selection of systematic reviews is covered in detail below. The output from this work was summarised in technical reports (level 3 of the Toolkit). Each technical report summarises the research on one topic in the Toolkit.

### Writing

The YEF team wrote up the findings from the evidence review phase into a series of accessible summaries (level 2 of the Toolkit). These summaries were reviewed by the evidence review team, external academics and members of the Toolkit audience. The summaries were also shared with a panel of young advisors – children and young people who provided advice and perspectives on the Toolkit content.

**User testing and design**

The YEF worked with a design agency, DIAS Creative, to design a website to host the Toolkit. We put this website through two phases of user testing:

1. In the initial phase of alpha testing, DIAS Creative developed different options for presenting the information in the Toolkit. These were presented to Toolkit users for feedback.
2. In the beta phase, DIAS Creative created a prototype online Toolkit which was also shown to users for feedback.

**Updating the Toolkit**

We will update the Toolkit twice a year. Initially, updates will expand the number of topics. We will then focus on updating existing topics to ensure they reflect the latest research.

## Selection of the most appropriate estimate of effectiveness

For each approach topic, the Toolkit presents a headline impact rating. This rating is based on the best available and most appropriate impact estimate from a meta-analysis. We identified the most appropriate estimate for this purpose using a three-stage process.

**Stage 1: Identify relevant reviews**

The evidence review team prepared a scoping note for each approach topic. This scoping note set out the eligible population, intervention, comparisons, outcomes and study designs (PICOS) for the topic. We used the PICOS to screen possibly relevant reviews to assess if they matched the scope. These reviews are identified from the reviews in the YEF Evidence and Gap Map, with approach-specific supplementary searches in Google Scholar and suggestions from Professor David Farrington and Dr Hannah Gaffney. We excluded reviews for which an update of that review is already in our list of eligible reviews. We also exclude reviews published before 2010.

**Stage 2: Eligibility screening**

All relevant reviews are screened for eligibility against the following criteria:

i) It must be a systematic review, that is based on systematic searching, screening, coding and reporting

ii) The review must contain a meta-analysis

Systematic reviews not including a meta-analysis may be drawn on in the report for information on the intervention, implementation issues and cost, but not for the impact estimate.

**Stage 3: Identify the most relevant effect size estimate**

If two or more reviews meet the eligibility criteria, then we identify the most relevant effect size from those in the eligible reviews, using the following criteria: (i) the effect size is a measure of a crime or violence outcome, (ii) the reported outcome has the strongest link to violent crime; (iii) the effect size based on the largest number of studies; (iv) the effect size has the highest evidence rating; (v) the most recent estimate; and (vi) the most UK and Ireland studies used to calculate the effect size.

These criteria are applied using a ‘knockout principle’. That is, if we can pick an effect size based on the first criterion then we stop there. If we cannot then we move to the next criterion, and so on. Often a single review will report several eligible effect sizes. These same principles apply to picking an effect size from several effect sizes within a single eligible review.

The technical report may report effect sizes from other eligible reviews, or effect sizes from sub-group analysis. But the effect size used to produce the headline impact rating is always clearly identified.

## Measures of offending

Individual primary studies may use different measures of offending, such as arrests or self-reported data. Systematic reviews typically combine these different measures, relying on the standardization of effect sizes to handle the fact that, whilst they measure the same underlying construct, different measures may give different numbers.

Similarly, in the technical reports we take offending outcomes reported in reviews at face value, rather than trying to distinguish different approaches to measuring offending. Our selection of effect size gives preference to reviews with a measure of serious violence or offending, but makes no distinction based on how these outcomes have been measured.

One of the key assumptions of the Toolkit is that variation in measures of offending is relatively evenly distributed across the studies included. So the risk of variation in approaches to measuring offending threatening the relative assessment of approach topics is low. The YEF is commissioning new systematic reviews with consistent methodologies, which will enable us to test this assumption.

# Calculation and reporting of the effectiveness (impact) estimate

## How effectiveness (the impact estimate) is reported

### Front page (top level): effectiveness categories

The front page (top level) of the Toolkit reports the average effectiveness of each approach in reducing serious violence. Effectiveness for each approach is reported by effectiveness category: harmful, low, moderate and high. The thresholds for these categories are based on the standardized mean difference (smd or d) for serious violence for that approach. The bands are shown in Table 1. Selection of the bands was based on analysis of the distribution of the d statistics in the included reviews, dividing the effect sizes into groups of roughly one-third for low, moderate and high (there is only one approach which is harmful).

**Table 1 Thresholds for effectiveness categories**

| Standardized mean difference (d) | Categories |
|---|---|
| **d < 0** | **Harmful** |
| **0 <= d < 0.10** | **Low (small or no effects)** |
| **0.10 <= d < 0.25** | **Moderate** |
| **d => 0.25** | **High** |

Only a minority of reviews report the effect on violence. If a review does not report an effect for violence, there are two issues to be dealt with in order to obtain a serious violence estimate:

- Where the review reports an offending outcome, we assume that the reduction in serious violence is at the same rate as general offending. This position is justified by the fact that children and young people typically do not specialise in a particular type of crime (see Box 1 and Annex 1).

- Where the review does not report an offending outcome but does report an intermediate outcome or behaviour such as bullying or disruptive behaviour, we use the following method to estimate what we call ‘indirect effects’: (i) calculate the reduction in intermediate outcome as a result of the intervention, (ii) use an estimate of the association between the intermediate outcome and offending to calculate the prevalence of offending for children and young people (CYP) with and without the behaviour; (iii) using the information from steps (i) and (ii) we calculate the number of CYP who do and don’t offend with and without the intervention; and (iv) using the figures from step (iv) we calculate the odds ratio for the effect of the intervention on offending from which we can derive the d statistic.

Our approach to calculating indirect effect sizes in these cases is laid out in more detail below.

**Second-level: percentage reduction in serious violence**

On the second level of the Toolkit we report the percentage reduction in serious violence resulting from the intervention. The percentage reduction is the relative reduction, not the absolute percentage point reduction. That is, if the offending rate is 40% and the intervention reduces it to 30%, then the relative reduction is (40-30)/40=10/40=25%. The relative reduction is the standard approach to reporting the effect of interventions (rather than the absolute reduction).

**Box 1: To What Extent is Violent Offending by Young People Versatile or Specialized?**

***By David P. Farrington, Institute of Criminology, Cambridge University***

The issue of versatility or specialization in offending by young people has important theoretical implications. Should criminological theories assume that all types of offending reflect the same underlying theoretical construct (e.g. an antisocial potential) or should they assume that violent offending reflects an underlying violent potential, that theft reflects an underlying thieving potential, etc? Some theories assume that a general underlying antisocial potential develops over time (influenced by long-term individual, family and socio-economic factors), but that the actual occurrence of specific types of offences depends on short-term immediate situational factors such as criminal opportunities (see e.g. Farrington, 2020). Most theories assume a general underlying potential, which would indicate versatility rather than specialization in criminal careers.

It has long been known that people who commit violent offences tend to have committed frequent offences. This was pointed out by Farrington (1978, 1982), who concluded that the probability of committing a violent offence in a criminal 'career' increased with the number of offences committed, and that specialization in violent offending was uncommon. Farrington (1991) analysed this in more detail in the Cambridge Study in Delinquent Development (CSDD), which is a prospective longitudinal study of 411 London boys born mostly in 1953. Up to age 32, there were 50 convicted violent offenders who committed a total of 85 violent crimes (on different days leading to convictions) and 263 non-violent crimes. Only 7 out of 50 had no convictions for non-violent crimes. The probability of committing a violent crime increased from 12.2% of those with one conviction to 65.2% of those with 9 or more convictions. The actual numbers of participants convicted of violent offences in each frequency category was not significantly different from the expected number based on the assumption that violent crimes were committed at random in criminal careers. Furthermore, participants who committed violent offences (each committing an average of 7.0 crimes) were not significantly different from equally frequent non-violent offenders (each committing an average of 6.5 crimes) on childhood (age 8-10), adolescent (age 12-16), teenage (age 18) and adult (age 32) factors.

In the latest review of the development of violent offending, Farrington (2018) analysed CSDD convictions up to age 56. The probability of committing a violent offence increased from 10% of those with one offence to 37% of those with 2-3 offences, 63% of those with 4-10 offences, and 78% of those with more than 10 offences. Farrington (2019) analysed CSDD convictions up to age 61, and reported that 92.1% of 76 violent offenders also committed non-violent offences, and that, for convictions up to age 20, this was true of

87.2% of 39 violent offenders. Incidentally, the prevalence of convictions for violence was greater for the sons of the CSDD males (born on average in 1981) than for the CSDD males themselves (17.5% compared with 14.4%). For comparability, both samples were studied up to the same age (29 on average), and overall more of the CSDD males were convicted of any offence (39.1% compared with 27.7% of their sons). Serious theft, minor theft and fraud/receiving were less common among the sons, unlike violence.

Farrington et al. (1988) then studied the complete juvenile court careers of nearly 70,000 offenders in Arizona and Utah. They classified offences into 21 types. For those with 2 or more referrals, the FSC values were .12 for robbery, .06 for aggravated assault, .07 for simple assault, and .03 for weapons offences. The FSCs were lower for those with 10 or more referrals, lower for females than for males, and lower for younger (age 13 or less) than older (age 16-17) offenders.

Since these pioneering studies, there has been a great deal of research on specialization versus versatility in criminal careers (reviewed by Mazerolle and McPhedran, 2019). While the FSC has been widely used, the other main measure has been the diversity index. The FSC yields information about the extent to which particular crimes are specialized, while the diversity index yields information about the extent to which particular offenders are specialized (see Piquero et al., 1999; Mazerolle et al., 2000). The minimum value of the diversity index is zero, indicating total specialization, while its maximum value is (k – 1)/k, where k is the number of types of offences, indicating maximum diversity. However, as Sullivan et al (2009) pointed out, the diversity index does not tell us about the type(s) of crimes in which an offender specializes. Therefore, it has limited relevance in this paper. Farrington et al. (1988) used a different method of studying specialization in criminal careers, by comparing actual numbers of offences of a particular type in a career (for those with at least 10 offences) with expected numbers (on the assumption that offences were committed at random). They found very few specialists. They concluded that 11 out of 209 people who committed robbery specialized in robbery, only 1 out of 350 people who committed aggravated assault specialized in aggravated assault, only 6 out of 525 people who committed simple assault specialized in simple assault, and only 2 out of 168 people with weapons offences specialized in weapons offences.

Based on existing research, it can be concluded that there is a small amount of specialization superimposed upon a large amount of versatility in criminal careers, and that most young people who commit violent offences are not specialists. However, specialization is more common at older ages and among sex offenders. Generally, people who commit violent offences tend to commit frequent offences, and it is not too implausible to suggest that violent offences are almost committed at random in criminal careers. However, the main caveat is that almost all the research is based on criminal records, because they include precise dates of commission of offences; more research on specialization is needed based on self-reported offending, but questions about self-reported offending would need to specify exact dates of commission (which may be difficult to remember for frequent offenders). For the moment, existing research indicates that conclusions about the effects of programmes on offending in general would apply almost equally to violent offending (at least for young offenders).

**References**

Farrington, D. P. (1978) The family backgrounds of aggressive youths. In Hersov, L., Berger, M. and Shaffer, D. (Eds) *Aggression and Antisocial Behaviour in Childhood and Adolescence.* Oxford: Pergamon (pp. 73-93).

Farrington, D. P. (1982) Longitudinal analyses of criminal violence. In Wolfgang, M.E. and Weiner, N.A. (Eds) *Criminal Violence*. Beverly Hills: Sage (pp. 171-200).

Farrington, D. P. (1991) Childhood aggression and adult violence: Early precursors and later life outcomes. In Pepler, D.J. and Rubin, K.H. (Eds.) *The Development and Treatment of Childhood Aggression.* Hillsdale, NJ: Lawrence Erlbaum (pp. 5-29).

Farrington, D.P. (2018) Origins of violent behavior over the life span. In Vazsonyi, A.T., Flannery, D.J., and DeLisi, M. (Eds.) *The Cambridge Handbook of Violent Behavior and Aggression* (2nd ed.). Cambridge: Cambridge University Press (pp. 3-30).

Farrington, D.P. (2019) The development of violence from age 8 to 61. *Aggressive Behavior*, 45, 365-376.

Farrington, D.P. (2020) The Integrated Cognitive Antisocial Potential (ICAP) theory: Past, present and future. *Journal of Developmental and Life-Course Criminology,* 6, 172-187.

Farrington, D. P., Snyder, H.N. and Finnegan, T.A. (1988) Specialization in juvenile court careers*. Criminology*, 26, 461-487.

## How the effect sizes are calculated

**Calculation of the d-statistic**

The review may report the effect size as d, g or an odds ratio. If an odds ratio is reported then it is converted to d using d = 3^0.5 Ln(OR) /π (Borenstein et al, 2021: 44). The g statistic is taken for the d statistic without further adjustment.

**Calculation of the effect size in serious violence**

The assumption that children and young people do not specialize in certain types of crime means that $S=kC$, that is serious violence is a fixed proportion of all crime. From this equation it can be shown that the percentage reduction in serious violence is likely to be the same as the percentage reduction in crime. It can also be shown that the d statistic is likely to be the same for the two outcomes (see Annex 1). Hence, when the review reports offending or reoffending, we take the same effect estimate to apply to serious violence. We acknowledge that this involves an assumption, but this provides the best estimate available based on the current evidence. Where a review reports an impact on violence, this is used rather than an estimate based on overall offending.

**Calculation of the percentage reduction in crime and serious violence**

The percentage reduction in crime and serious violence is calculated by deriving a 2x2 table from the odds ratio. If d or g is reported it is converted to an odds ratio.

The 2x2 table used to calculate the percentage reduction shows the number of offenders and non-offenders in the treatment and control groups. The calculation requires the odds ratio, which is taken from the review, the numbers of children in treatment and control (though the result is not sensitive to either the total number used or the proportion in treatment and control groups), and the prevalence of crime in the comparison group. We assume this prevalence to be 25% for offending and secondary outcomes such as bullying, and 50% for re-offending (see Box 2), with sensitivity of results to that assumption reported in annex 1 of each technical report.

For example, in the case of pre-court diversion, Wilson et al. (2018) report an odds ratio of 0.77 which converts to a relative decrease in reoffending of 13%.

To obtain this percentage reduction we produce the 2x2 table shown in Table 2. Assume a total study population of 200, though the actual number does not matter for the results. We assume half of these are in the treatment group and half in the control. So, there are 100 youth in diversion and 100 in the control group. As the outcome is reoffending, we assume the control group prevalence to be 50%. That is 50 of those in the control group offend.

The odds ratio for the effect of diversion on reoffending is $OR = (Rt/DRt) / (Rc/DRc)$, where R is reoffend and DR is don't reoffend and the subscripts t and c refer to treatment and control respectively. This definition may be re-arranged to give $Rt = T / [ (Rc/DRc) OR + 1 ]$ where T is the number in the treatment group (that is, $T=Rt+DRt$). We know Rc/DRc because we assume prevalence in the control group. As shown in Table 2, application of the formula gives the number who reoffend in the treatment group as 43.5, compared to 50 in the control group. The percentage reduction in children who reoffend is thus (43.5-50)/50 = -13%.

**Table 2 2x2 table for calculation of % reduction in offending from diversion**

| | **Reoffend** | **Don't reoffend** | **Total** |
|---|---|---|---|

|  |  |  |  |
|---|---|---|---|
| **Control** | **50** | **50** | **100** |
| **Treatment** | **43.5** | **56.5** | **100** |
|  | **93.5** | **106.5** | **200** |

**The calculation shown above is reproduced in Annex 1 in each technical report.**

The resulting reduction in crime is dependent on the effect size and the assumed prevalence in the control group. The former is the effect size we have selected as the most appropriate effect size estimate. The latter is assumed to be 25% for the prevalence of offending, and 50% for re-offending; see Box 2 for the justification of these assumptions. In each technical report we include in Annex 1 sensitivity analysis for variations in the control prevalence rate. We examine the effect of varying the control offending prevalence from 10-40% and re-offending rate from 35-65%. Overall, changing the prevalence rates within these bands does not substantially change the results

**Box 2: Justification of the baseline prevalence assumptions**

**Reoffending**

For interventions targeted at offenders we assume a 50% reoffending rate. England and Wales in 2014-15, 38% of juvenile offenders had proven reoffending in only one year after the previous conviction (Ministry of Justice, 2017), so it can be expected to reach 50% within two to three years. The Campbell systematic review by Wilson et al. (2018) on police-initiated diversion for youth makes the same assumption of 50 per cent reoffending by those in the comparison group.

**Offending**

The Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of 411 London boys, 25% were convicted between ages 10 and 17 (Farrington, 2012). Conviction rates have fallen over time – but offending is greater than conviction, and conviction is falling in part precisely because an increasing number of children and young people who offend are diverted before conviction. Moreover, the approaches covered in this Toolkit are generally directed to children and young people who are at risk of offending. The prevalence of offending in this group will be greater than that in the population as a whole.

**References**

Farrington, D. P. (2012). Predictors of violent young offenders. In B. C. Feld & D. M. Bishop (Eds.), The Oxford handbook of Juvenile crime and Juvenile justice (pp. 146–171). Oxford: Oxford University Press.

Ministry of Justice (2017) Youth Justice Statistics 2017/18. London: Youth Justice Board / Ministry of Justice.

Wilson, D.B., Brennan, I. and Olaghere, A. (2018), Police-initiated diversion for youth to prevent future delinquent behavior: a systematic review. Campbell Systematic Reviews, 14: 1-88. https://doi.org/10.4073/csr.2018.5

**Calculation of the effect size with indirect effects**

Not all reviews report criminal behaviour. They may report an intermediate outcome such as bullying perpetration or behavioural difficulties. There are two topics in the Toolkit where this is the case (anti-bullying and parent training). In such cases we combine the effect of the intervention on the intermediate outcome, and the association between that outcome and offending, to calculate the expected effect of the intervention on violence.

There are four steps in this process:

1. Use the odds ratio for the impact of the intervention on the intermediate behaviour to calculate the % reduction in the intermediate outcome (behaviour)
2. Use the odds ratio for the association between the behaviour and offending to calculate the difference in prevalence of offending for those with and without the behaviour
3. Construct tree diagrams with and without the intervention

4. Construct a 2x2 table for offending with and without the intervention, and so calculate the odds ratio from which d can be calculated

The approach is described in detail in Annex 2.

**Discussion of statistical significance**

We report the average effect size and its conversion to percentage reduction in violence regardless of its statistical significance. We believe that effectiveness estimate and evidence rating are a better guide to impact than statistical significance, with some key factors affecting significance captured in our evidence rating. We report statistical significance in the technical reports.

Statistical significance depends on the effect size (smaller effects are less likely to be significant), and the variance in included study effect sizes. This variance depends on the number of included studies (more studies increase the likelihood of statistical significance)[1] and variation in the effect sizes, that is heterogeneity. Both the number of included studies and heterogeneity are factors in the evidence rating.

## Moderator analysis

Reviews report the average effect size across all included studies. That average effect size is used for our effectiveness rating. But effects vary between studies, sometimes considerably. Understanding possible reasons for this variation can provide information about when the approach should be used and how to maximise its impact.

The effect size for offending may vary according to the effect on intermediate outcomes (mediators) and contextual, design and implementation factors (moderators). Information on moderators is important for users in deciding whether an intervention is appropriate for

[1] The sample size in the included studies also affects the variance. However, this figure is not consistently reported in reviews and so not used directly in our evidence rating.

them, and what are important issues in design and implementation. These moderators are discussed on the second level of the Toolkit.

The technical report provides information on all moderators reported in the review. These are discussed in the text and reported in full in Annex 1. Where an estimate of the change in effect size due to the moderator is provided then that is reported. If the review only provides information on statistical significance then we report that.

In some cases we draw on an additional review or reviews, or another paper reporting results from further analysis of the review dataset, to supplement the discussion of moderators.

# Evidence rating

## The issue

The Toolkit is intended to show ‘best bets’. How promising an approach is depends on both the effect size and the evidence rating.

The evidence rating refers to the confidence we have in our assessment of the effectiveness of an approach. The higher the rating then the larger the evidence base which has been well summarized so we do not believe it likely that new evidence will substantially change the effect.

The rating system is as follows:

●●●●● **= Very high confidence in the impact rating**

●●●●○ **= High confidence in the impact rating**

●●●○○ **= Moderate confidence in the impact rating**

●●○○○ **= Low confidence in the impact rating**

●○○○○ **= Very low confidence in the impact rating**

○○○○○ **= There is insufficient evidence to calculate an impact rating**

The evidence rating is based on four criteria. The decision process is summarised in Table 3.

- The number of primary studies used by the meta-analysis to calculate the effect size. The number of studies determines the upper limit for the evidence rating:
  - If there are no studies available, the topic is automatically assigned an evidence rating of 0
  - If there are 1-2 studies available, the highest possible evidence rating is one
  - If there are 2-4 studies available, the highest possible evidence rating is two

  - If there are 5-7 studies available, the highest possible evidence rating is three
  - If there are 8-11 studies available, the highest possible evidence rating is four
  - If there are 12 or more studies available, the highest possible evidence rating is five
- The confidence we have in the methodology of the review from which the impact measure is taken (or based on in the case of an indirect measure). Confidence in the review is assessed by critical appraisal of the review using a modified version of the AMSTAR 2 tool. The AMSTAR rating is used to assign the review a rating of either low, medium, or high confidence.
  - If the AMSTAR rating is low then we drop one rating level.
  - If the AMSTAR rating is medium or high then we do not drop a rating level.
- The consistency of effect sizes from the primary studies used by the meta-analysis to calculate an effect size (i.e. heterogeneity).
  - If the $I^2$ is greater than 60% then we drop one rating level.
  - The $I^2$ is less than or equal to 60% then we do not drop a rating level.
- Whether the impact estimate is based on a direct measure of crime or violence,  or an indirect estimate based on an intermediate outcome such as bullying perpetration.
  - If it is an indirect estimate then we drop one rating.
  - If it is a direct measure we do not drop a rating.

Table 3 captures the decision rule.

| Table 3: Evidence strength decision rule | | | | | | |
|---|---|---|---|---|---|---|
| | Number of included studies | | | | | |
| | 0 (No review or empty review) | 1-2 | 2-4 | 5-7 | 8-11 | 12 or more |

| | | | | | | |
|---|---|---|---|---|---|---|
| AMSTAR rating is low | 0 | * | $I^2$ <= 60%: **<br>$I^2$ >60%: * | $I^2$ <= 60%: ***<br>$I^2$ >60%: ** | $I^2$ <= 60%: ***<br>$I^2$ >60%: ** | $I^2$ <= 60%: ****<br>$I^2$ >60%: *** |
| AMSTAR rating is moderate or high | 0 | * | $I^2$ <= 60%: **<br>$I^2$ >60%: * | $I^2$ <= 60%: ***<br>$I^2$ >60%: ** | $I^2$ <= 60%: ****<br>$I^2$ >60%: *** | $I^2$ <= 60%: *****<br>$I^2$ >60%: **** |

Notes: (1) For indirect effect estimates the evidence rating is dropped 1 level with a floor of *. (2) If Q is reported we calculate I squared (I2 = (Q-df)/Q where df is no. of effect size estimates -1); (3) if tau squared is reported we use 'high heterogeneity' as reported by authors (4) if heterogeneity is unclear drop a * (we may make an exception if the eyeball test shows clearly low-moderate).

## Rating the review

The review is rated using a modified version of AMSTAR 2, a widely used tool for assessing the quality of systematic reviews. The items covered are: (i) Use of all the components of the PICOS; (ii) comprehensive literature search strategy; (iii) study selection and data extraction in duplicate; (iv) description of the included studies; (v) assessing the risk of bias in included studies; (vi) discussion of heterogeneity; and (vii) report potential sources of conflict of interest. Items are score as yes, partially yes or no, which correspond to High, Medium and Low confidence respectively.

The overall rating is given using the 'weakest link in the chain' principle. That is, the overall rating is equal to the lowest rating on any item. If a review receives a 'partial yes' on any item (with no item rated 'no'), the overall rating will be Medium. If a review receives a 'no' on any item, the overall rating will be Low.

The full details are presented in Annex 3.

## Implementation evidence

If there is an available synthesis of implementation and process evaluation on the approach, that is used to present implementation evidence.

If that is not available, then we identify relevant process evaluations of the intervention. Preferably these are evaluations of the intervention implemented in UK or Ireland. If UK and Ireland evaluations are not available we include other available studies. The process evaluations are taken from the YEF Evidence and Gap Map, with supplementary searches using search terms specific to the intervention in England and Wales.

Data are extracted from each process evaluation in a standard template which records information on success factors, challenges, and direct quotes from children and other stakeholders (the template is included as Annex 4 of this technical guide). The completed form is included as an annex in each technical report. The main themes from the process evaluations are summarised for the implementation section in the technical report.

## Cost data

The top level of the Toolkit includes a cost rating. This rating aims to give a simple, overall indication of the likely cost of an approach, relative to other approaches in the Toolkit. The second layer of the Toolkit describes relevant cost information in more detail.

### Collecting cost information

We gathered information about the cost of interventions from several sources:

- The Toolkit evidence review team extracted information on cost from systematic reviews and UK evaluations in the YEF EGM. This information is summarised in the technical reports.
- The YEF team conducted additional desk research. This involved looking for information about the cost of programmes and projects in each approach. We consulted a range of sources, including the EIF Guidebook, EEF Teaching and Learning Toolkit, the College of Policing Crime Reduction Toolkit, and programme developers' websites.
- If the technical reports and desk research did not provide clear information about the likely cost of an approach, the YEF team contacted subject experts to ask for more information.
- If we were unable to find cost estimates but had a good understanding of the components of the approach and the approach is relatively standardised, we used this understanding to place the approach in a rating category. For example, we know that boot camps involve long-term residential care and are therefore likely to be significantly more expensive than most other topics.

### Creating a cost rating

The YEF team used this information to calculate an average cost (or cost range) per participant for each approach. We prioritised cost data from the UK. If estimates from international sources were used, we converted them into GB pounds. The aim is to place each approach in a broad category (high/medium/low); see Table 4.

**Table 4 Cost categories**

| **Band** | **Estimated average cost per participant (£)** |
|---|---|
| **Low (£)** | **£0 – £500** |
| **Medium (££)** | **£500 - £1,500** |
| **High (£££)** | **£1,500 +** |

Costs are calculated from the perspective of commissioners likely to read and use the Toolkit: Violence Reduction Units, Police and Crime Commissioners, and Local Authorities. This means that the cost rating does not consider the counterfactual cost (the cost if the intervention had not been delivered). For example, boot camps might be relatively cheaper than custody but they will be relatively expensive from the perspective of a VRU which does not bear the cost of custody. The second layer of the Toolkit may describe the counterfactual cost, if there is sufficient information available.

The cost rating refers to the cost of starting a new intervention. We included both the costs of setting up and continuing the intervention.

## Updating the Toolkit

The Toolkit is a live resource and will be regularly updated. The YEF will aim to update the Toolkit twice a year. These updates will aim to:

- Add new topics. The first few updates will expand the Toolkit to cover additional topic areas.
- Update existing topics. Later updates will update existing topics to reflect new research.
- Add new functionality. As we learn more about how the Toolkit is used and what our audience needs, we will adapt the Toolkit with new features and functions.

The initial work of adding new topics will focus on areas where systematic reviews already exist. However, the YEF is also planning to commission new systematic reviews to cover topics where they currently do not exist.

# Reference

**Borenstein, M., Hedges, L, Higgins, J.P. and Rothstein, H. (2021) Introduction to Meta-Analysis, 2nd Edition, Chichester: Wiley.**

# Annex 1 - The equivalence of the percentage reduction in serious violence and general crime

The assumption that CYP do not specialize, can be expressed as saying that serious violence offences (S) are a fixed proportion of general offences (O) which can be written as $S = k.O$ where k is a constant. This can be rearranged as $S/P = k\ O/P$, where P is the population of CYP, so S/P is the prevalence of serious violence.

The percentage change in S from an intervention is dS/S. Now:

$dS/S = dS/P\ .\ P/S = dS/P\ .\ 1\ /\ (k\ O/P)$

But $dS/P = k\ dO/P$

Therefore $dS/S = k\ dO/P\ /\ (\ k\ O/P) = dO/O$

That is the percentage change in S and O is the same.

The absolute mean difference, $D_S = dS = k\ dO = k\ D_O$

Using $Var(aX) = a^2Var(X)$ implies $std(S) = k\ std(O)$.

Hence, $SMD_S = D_S/\ std(S) = k\ D_O\ /\ k\ std(O) = D_O\ /\ std(O) = SMD_{O.}$

That is, our assumption that CYP commit serious violence offences (S) at the same rate as general offences means that the effect size for S will be the same as the effect size for O.

## Annex 2 - Estimate of the d-statistic via indirect effects

There are four steps in this process:

1. Use the odds ratio for the impact on the intermediate behaviour to calculate the % reduction in the behaviour
2. Use the odds ratio for the association between the behaviour and offending to calculate the difference in prevalence of offending for those with and without the behaviour
3. Construct tree diagrams with and without the intervention
4. Construct a 2x2 table for offending with and without the intervention, and so calculate the odds ratio from which d can be calculated

The approach is illustrated using the example of bullying. Gaffney et al. (2019) report an odds ratio of 1.32 for the effect of anti-bullying programmes on bullying. We use this odds ratio, and the assumed baseline bullying perpetration rate of 25% to derive the 2x2 table shown in Table A.1. The difference in bullying in treatment and control is used to calculate the relative percentage reduction in bullying (19.4%).

**Table A.1 The reduction in bullying from interventions**

| | **Bully** | **Not bullying** | **Total** |
|---|---|---|---|
| **Control** | **50** | **150** | **200** |
| **Treatment** | **40.3** | **159.7** | **200** |
| **Total** | **90.3** | **309.7** | **400** |
| | | | |
| **% Change** | **-19.4** | | |

In step 2 we use the estimate of the association between bullying and offending from Ttofi et al. 2011 to construct a 2x2 table from which we obtain the prevalence of offending amongst children who haven't bullied others (which is assumed at 25%) and children who have (derived from the odds ratio). These are shown in Table A.2. The difference in prevalence is 0.13 (=0.38-0.25).

**Table A.2 Calculation of offending rates for children who have and haven't bullied others**

| | Offending | Not offending | Total | Prevalence of offending |
|---|---|---|---|---|
| **Haven't bullied** | **50** | **150** | **200** | **0.25** |
| **Have bullied** | **76.3** | **123.7** | **200** | **0.38** |
| **Total** | **126.3** | **273.7** | **400** | **0.32** |

Step 3 is to construct the tree diagram with and without the intervention. Figure A.1 shows the structure of the tree. The two main branches are treatment (with intervention) and control (without intervention). Both groups have children who have and have not bullied others, but there are fewer bullies in the treatment group. The data from Table A.1 are used to calculate the number of children who have bullied others in each group. Next, the data on the prevalence of offending according to whether children have bullied other or not in Table A.2 is used to calculate the number of children who become involved in violence in each group.

**Figure A.1 Tree diagram**

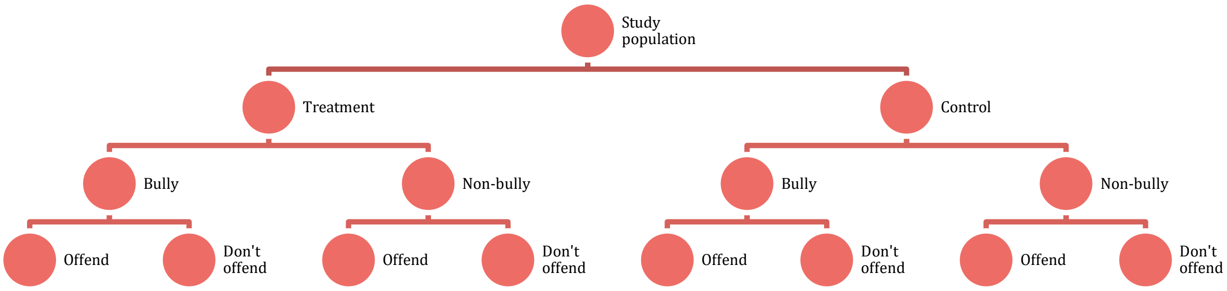


Table A.3 shows the tree diagram with numbers. The population of 800 is equally divided between treatment and control. The top half of the figure shows the number offending in the control group with the assumption of bullying prevalence of 25% and the derived offending rates for children who have bullied and children who have not  in the control (without intervention) group. There are 400 people in this group, so 100 are bullies and 300 not bullies: 38% of the former offend compared to 25% of the latter, giving 113 children who offend in total. Repeating the analysis for the treatment group (with intervention), which has 19% fewer bullies, gives 111 children who offend in total.

**Table A.3 Number of offenders in the control and treatment groups**

| | | | | | | |
|---|---|---|---|---|---|---|
| Population (800 CYP) | Control | Have not bullied | 300 | Offend | 75 | Total:<br>Offend: 113<br>Don’t offend: 287 |
| | | | | Don’t offend | 225 | |
| | | Have bullied | 100 | Offend | 38 | |
| | | | | Don’t offend | 62 | |
| | Treatment | Have not bullies | 319 | Offend | 80 | Total:<br>Offend: 111<br>Don’t offend: 289 |
| | | | | Don’t offend | 240 | |
| | | Have bullied | 81 | Offend | 31 | |
| | | | | Don’t offend | 50 | |

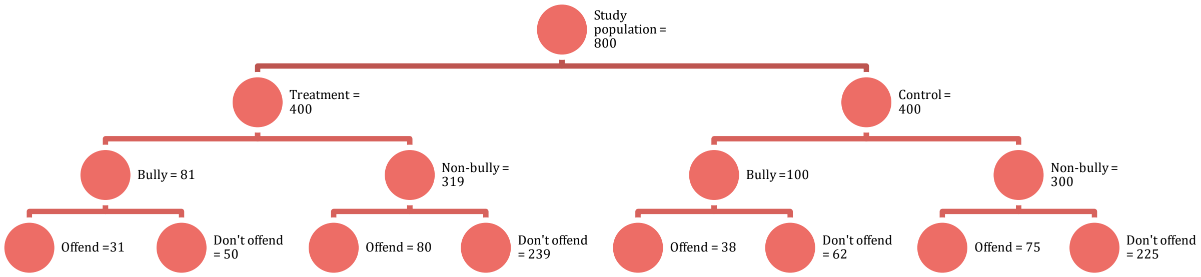
The results in Table A.3 are made to create a new 2x2 table (Table A.4) from which the odds ratio of the impact of the intervention on offending can be derived, and from that calculate d using d=(3^.5/π) l.n(OR). The odds ratio, OR = (111/289)/(113/287) = 0.97, which gives d=-0.02.

**Table A.4 2x2 table with and without the intervention**

| | Offend | Don’t offend | Total |
|---|---|---|---|
| Control (Without) | 113 | 287 | 400 |
| Treatment (With) | 111 | 289 | 400 |

**These numbers are put into the tree diagram in Figure A.2.**

## Figure A.2 Numerical tree diagram



## Annex 3 - Critical appraisal of reviews

We assess the systematic reviews using a modified version of AMSTAR 2.[2] The items are listed below. Each item is rated Yes (=High), Partial Yes (=Medium) or No (=Low). The overall assessment is made using the weakest link in the chain principle. That is, the overall rating is equal to the lowest rating on any item.

| **Modified AMSTAR item** | | **Scoring guide** |
|---|---|---|
| **1** | Did the research questions and inclusion criteria for the review include the components of the PICOS? | To score 'Yes' appraisers should be confident that the 5 elements of PICO are described somewhere in the report |
| **2** | Did the review authors use a comprehensive literature search strategy? | At least two bibliographic databases should be searched (partial yes) plus at least one of website searches or snowballing (yes). |
| **3** | Did the review authors perform study selection in duplicate? | Score yes if double screening or single screening with independent check on at least 5-10% |
| **4** | Did the review authors perform data extraction in duplicate? | Score yes if double coding or single coding with independent check on at least 5-10% |
| **5** | Did the review authors describe the included studies in adequate detail? | Score yes if a tabular or narrative summary of included studies is provided. |
| **6** | Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? | Score yes if there is any discussion of any source of bias such as attrition, and including publication bias. |
| **7** | Did the review authors provide a satisfactory explanation for, and | Yes if the authors report heterogeneity statistic. Partial yes if there is some discussion of heterogeneity. |

[2] https://www.bmj.com/content/358/bmj.j4008

| | | |
|---|---|---|
| | discussion of, any heterogeneity observed in the results of the review? | |
| **8** | Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review? | Yes if authors report funding and mention any conflict of interest |

# Annex 4 – Process evaluation data template

| Author & Title | Intervention | Success | Issues/ Challenges | Young People’s views |
|---|---|---|---|---|
| | | | | |

YOUTH ENDOWMENT FUND

Impetus

Home Office

www.youthendowmentfund.org.uk

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413