Giving Evidence
Enabling giving based on sound evidence

Institutional Responses to Child Abuse: The Set of Rigorous Effectiveness Studies: Summary Report

This document describes the rigorous studies about the effectiveness of institutional responses to child abuse. It discusses what such studies exist, what they say, and suggestions for future research. It builds on and summarises the findings of two products: first, an 'evidence and gap map', which shows what such studies exist, where there are concentrations of them and where there are gaps; and second, a 'Guidebook' which summarises what those studies say, and pulls out the common themes.

Further detail about the project and its products is available at www.giving-evidence.com/csa That includes:

- A more detailed introduction to evidence and gap maps
- A full report about our evidence and gap map about institutional responses to child abuse
- An interactive, searchable version of our evidence and gap map
- Downloadable visual PDF version of our evidence and gap map
- The full 'Guidebook', summarizing the studies on the map
- The results of a (sadly very inconclusive!) attempt to map the amount of activity in child protection globally against the frame used in the evidence and gap map. This aimed to identify areas where there is lots of activity but little evidence, which would be priorities for new research, and converse areas where there is strong evidence that something works but few people running it, which might be priority to encourage new activity.

We are publishing the evidence and gap map with the Campbell Collaboration. At the time of writing, it is going through peer review, nearing the end of that process. The main findings and conclusions presented here are unchanged by that process. The website cited above has the published protocol, and we will publish the final paper once that is published.

Future products will be published on the above website too.

There are some small inconsistencies between the various versions of the map. This is because, for example, the Guidebook was produced while the peer review was underway (it could not wait for it). They do not affect the main findings, conclusions or implications.

November 2020

Corresponding author:
Caroline Fiennes, Director, Giving Evidence
+44 7803 954512, caroline.fiennes@giving-evidence.com

Contents

Executive summary

Recent years have seen significantly increased in interest work on child abuse, particularly within organisations such as churches, schools, youth clubs, residential care. The scale of the problem globally is gigantic compared to the resources available to prevent it and respond to it[i]. Hence it is imperative that no resources be wasted: rather, work and funding in this area should be as effective as possible. That means ensuring that they are informed as much as possible by sound evidence: evidence about where the problem is (in what geographies, and which types of institution), why it happens, who is affected (who are the victims, and perpetrators, who can prevent it and who can respond), and what is effective.

We looked at the effectiveness of interventions – the effect of some intervention(s) on some outcome(s). For example, on the effect on a child's brain development of being raised in Romanian children's homes[ii] vs. in a family. We looked across prevention, disclosure, response and treatment, and all forms of abuse. We asked two questions: first, what evidence of this type already exists that can guide decisions? And second, what does it say? This document summarises our findings on both questions.

Our method began with systematic search for effectiveness studies, including from academia and elsewhere such as practitioner organisations. We were only interested in 'fair tests' of effectiveness, so set that bar fairly high: we included primary studies with a defensible counterfactual, and systematic reviews. We placed the studies that fit our criteria (explained below) on a grid: this shows interventions as rows and outcomes as columns. Thus the placement of studies on the grid (the map) shows where there are concentrations of studies, and where there are gaps. We also coded the completed studies for their reliability. We then read those studies in detail (noting, for example, their sample sizes, the outcomes they measured, the metrics they used, the sizes of effects that they found) and we summarized them.

Our findings, in short, are that these studies are scarce and of limited reliability. We found only 58 completed primary studies, plus three planned new primary studies, and 10 systematic reviews. Some interventions have never been studied in these kinds of primary studies, such as any intervention aiming to encourage disclosure, or the effectiveness of institutional safeguarding policies. Most of the studies look at prevention, and mainly of sexual abuse. There are no studies with clergy or in faith-based institutions, and none of treatment (therapy) interventions begun in the last 20 years. The studies do not match where the world's people are: Western Europe, the US and Canada account for 81% of the completed primary studies. There are no studies from India, only six from China, Africa and Latin America combined.

All the interventions studied have some positive effect(s). Though a few interventions show dramatic results, in general the effects are modest: they reduce the problem but do not eliminate it. Some of the interventions studied make no difference to some of their intended outcomes. On the upside, there was no clear evidence of an intervention producing harm. But most results attenuate over time and most of the studies are low reliability anyway – many are small and short: the picture may not really be as rosy as it seems.

Few studies record actual abuse, or disclosures of abuse: instead, most use intermediate outcomes, and over fairly short time-scales. Few studies explicitly state the theory on which their intervention is based. Very few studies report what the intervention costs.

There is an urgent need for more research, as well as for practitioners and funders to use what already exists.

Scope of our Evidence and Gap Map and Guidebook

We looked for studies relating to child abuse within institutional contexts: in schools, care homes, youth clubs, sports clubs, young offender institutes, churches, etc. We did not look at abuse within families.

We included effectiveness studies of the following types:

- Primary studies (defined in Box 1) which are a 'fair test' of the intervention, i.e., which have a defensible counterfactual (i.e., some way of seeing 'what would have happened otherwise', which isolates and hence shows the effect of the intervention). Those are:
    - Randomised controlled trials (RCTs)
    - Quasi-experimental designs: several designs, all of which have credible counter-factuals.
- Systematic reviews of effectiveness studies (also defined in Box 1).

The logic for the choice of studies to include was that identifying the effect of an intervention means assessing it in a fair test. Ideally that would be a randomised controlled trial, which considered the best design for a single study to assess an intervention's effectiveness, we recognise that there are many institutional interventions where random assignment may not be appropriate or possible. Hence we were open to including studies with quasi-experimental designs (QEDs) which had robust control groups. Non-controlled studies, such as pre-post evaluations were excluded because they lack a credible counterfactual and therefore are not a fair test. We also excluded studies which are not impact evaluations such as qualitative studies, process evaluations, cross-sectional surveys, observational studies, case studies or opinion pieces.

We sought studies published anywhere globally, and we looked in various languages. The studies could have been published at any time: though in practice, the earliest study we found was published in 1985.

We sought studies published:

1. In the academic literature. We searched databases of academic journal articles, and
2. Elsewhere. We solicited suggestions from relevant experts (both academics and practitioners), and searched the websites of some relevant organisations (e.g., statutory inquiries into institutional abuse including child sexual abuse, and some charities).

The scope is detailed in the appendix, and the search strategy is described in detail in the longer report mentioned and the protocol.

The search, screening and most of the coding for the EGM was done with the Centre for Evidence and Implementation and Monash University. The Guidebook was produced with the Campbell Collaboration, who also did some coding. The interactive version of the map was produced with the Campbell Collaboration and the Africa Centre for Evidence. Both the Evidence and Gap Map and the Guidebook were produced with funding from Porticus, a funder. The Guidebook is available at www.giving-evidence.com/csa, as is a long report about the EGM, and a summary report about it. The report about the EGM is being published through the Campbell Collaboration, and will be published when peer-review is complete. The other products vary slightly because some pre-date and some post-date peer-review. The main findings are unaffected.

This document uses 'we' and 'our' to refer to the papers' authors, not to the funder.

The size of the evidence base

We found[1] 58 completed primary studies, three protocols (i.e., plans) for further primary studies which we assume have now started, and 10 systematic reviews[2]. **In total, we found (and the EGM has) 71 studies.** There is one systematic review included which is an update of an earlier one, also included, and two instances where multiple papers have been written about the same underlying experiment. Counting all of those separately, **our map comprises 81 papers.** These are very small numbers given the scale of the problem. Much remains unknown: many more issues in this terrain need to be researched. But:

> *"Where to start isn't the issue, that we start at all is what matters most."*
>
> - Margaret Heffernan's book Willful Blindness

To be clear, the absence of evidence about a type of intervention is no slight on that intervention. It is not normally a comment on that intervention nor on any organisation that runs it: rather, it is a comment about the available research (which fits the scope and criteria of this EGM).

What the studies say

The good news is that **most of the interventions studied have some positive effect(s)**. Very few of them had no effect on any of their intended outcomes. However, to be clear, a positive effect means that the intervention produces *some positive effect*: they reduce the problem but do not eliminate it. Most of the studied interventions have a modest effect. An effective programme may improve knowledge by 20-30 percent and reduce abuse by 10-20%. (The modest-ness of effects is true of most social interventions in any sector.) One of the strongest results was for the *Good Schools Toolkit*, which is studied in multiple papers on the EGM. It is a whole-school approach to reducing violence in schools in Uganda, which the number of students who experienced by school staff[3] in the previous term from 80% (clearly a giant amount) to 'only' 60% of them.

No intervention studied seemed to create harm :-) Remarkably, no study reported any adverse effects - though many studies did look for them: such as whether children's anxiety increased when they learned about 'bad touches'. It is important to be careful here: a few studies appeared find to adverse effects, but none is very clear, and they may all show increased *reporting* rather than increased *incidence*. For instance, Taylor 2010 found that students who received the dating violence programme studied reported (themselves)

[1] To be clear, we had a broad search, so the small number of studies is unlikely to be the result of our search strategy being too narrowly focused, or too niche.

[2] There were three instances of multiple research papers written about the same underlying study:
(1) There are six papers written using the data from a single study (RCT) of Romanian orphanage children: we count that only once to avoid counting those children six times.
(2) There are five research papers written using the data from a single study of the Good Schools Toolkit in Uganda. We count that once.
(3) There is one systematic review from 2007 which was updated in 2015 (with new methods and the new studies), and we count that only once.
In other words, the total map (incl. duplicates) has 81 papers: 57 about primary studies, 3 protocols, 11 systematic reviews.

[3] This does not mean 'just' corporate punishment for purposes of discipline, but rather can be much more severe. For example, Devries 2015 reports a study in Luwero region of Uganda which found that 8% of school pupils had experienced 'severe physical violence' at school, including choking, burning and stabbing.

committing more violence against their dating partners, though this might be because the programme taught them that behaviours they had hitherto considered normal were in fact violent. For example, many students did not previously know that sex between minors is legally considered rape.

Some interventions have no effect, or at least, no effect on some outcomes. For example, a programme run in the Netherlands with at-risk boys living in residential care aimed to reduce sexual harassment by them, but found no effect. This finding is consistent with a finding across social sectors (i.e., outside child protection) that around 80 per cent of interventions in all sectors have small or no effect.

There are some mixed results. For instance, a bystander programme in US high schools found no effects on participants stopping harassment, but did find improvement in denying that rape is possible. Furthermore, some cells have multiple studies, some of which found an intervention to work and others to achieve nothing. This is also unsurprising because, one cell can contain studies with quite diverse, interventions, populations, comparison groups and outcome measures.

What the studies cover

Of those 71 studies: 58 were completed primary studies, 40 of them RCTs[4]; three were protocols for new primary studies, all RCTs; and 10 were systematic reviews[iii]. Most of the included studies came from academic journals (by searching the databases of them); only a handful came from non-academic literature such as practitioner organisations.

Geographically, the studies don't match where the world's population is. Though there are studies from quite a few countries, the studies are markedly concentrated geographically. The US dominates, with 32 of the 58 completed studies; and Western Europe, the US and Canada account for 81% (47) of the completed primary studies between them. Those regions are nothing like 81% of the world's population. Only five of the 61 primary studies (including protocols) are from countries which have Muslim majorities.

By way of demonstrating the mismatch between where studies are vs. where the problem is, we found no studies from South Africa, where child sexual abuse alone is thought to affect[iv] over one child in three. We found no studies from India and only two from China, which obviously have ~2 billion of the world's ~7 billion people between them. We found only three studies from Africa which has over a billion people (two studies from Uganda and one from Tanzania), and one completed primary study from Latin America (Equador).

The major concentration of studies is in programmes educating children about how to prevent violence. They are education-based prevention programmes, delivered in early education and school settings. Fully 50 of the 58 completed primary studies examine such programmes, as do all three protocols for planned studies. Most of the programmes were curriculum-based and aimed to teach children awareness and understanding of sexual abuse and teaching safety skills, e.g., the difference between good touches vs. bad touches, how/when to handle 'secrets' and who to tell (n = 42) These usually involved workshops or lessons, combined with relevant written, audiovisual or other resources (parent materials, activity books), and were

[4] For all cases where this document has a statement such as this, the list of the studies specific referenced is in the longer document. It can also normally be found from the graphical map.

delivered directly to children in small groups via an external agency or existing trained institutional staff. The good news is that these programmes all seem to succeed in increasing children's knowledge and none produces harms: we found this from the studies on our map, as did a systematic review[v] in 2015. However, they have only ever been tested in high-income countries: this is a shame, because, since they appear to work, it would be worth testing them in low-income countries.

Overwhelmingly, studies look at programmes which target children. This is remarkable given that children are not the problem. No primary study looks at legal sanctions or legal response: indeed, only two primary studies look at response at all, and in only one of those was response a major goal of the programme.

Only one completed study assessed an intervention with adults to stop them offending in organisations[5] (either at all, or re-offending). This seems amazingly few. That one study assessed the *Good Schools Toolkit*. There is one completed study of youth-on-youth violence, and one on-going study working with teachers in Jamaica in day-care to reduce violence against children.

We found no causal studies conducted in religious organisations. That is remarkable given the scale and media interest in clerical sexual abuse and the number of countries in which it has been reported[6].

Most of the studies report intermediate outcomes, such as children's acquisition and retention of knowledge, but not actual disclosure of incidence: this may be because actual disclosure can take years, making it expensive and difficult to track.

Most of the studies are about sexual abuse. Sexual abuse was considered by 56 of the primary studies: in 47 primary studies, sexual abuse was the singular focus, and in a further nine studies, sexual abuse was examined alongside other maltreatment types. By contrast, only 12 studies reported on physical abuse, four on neglect, and three on emotional abuse. None had emotional abuse as a main focus. Of the 10 systematic reviews, eight included studies that reported solely on interventions relating to sexual abuse. The remaining reviews included studies that reported on one or more types of child maltreatment: two included studies assessing physical and emotional abuse, as well as neglect, and one included studies reporting on sexual, physical and emotional abuse.

Most of the studies are about prevention. Prevention was examined in 60 papers about completed primary studies (and all three protocols), and 10 systematic reviews. Some studies looked at prevention alongside other issues. By comparison, treatment was studied in only two primary studies and one systematic review; and response was studied in two primary studies and five systematic reviews. On disclosure, we found no primary studies of interventions aiming to facilitate disclosure(!), and only two systematic reviews including

[5] Note that we were looking for interventions with perpetrators of abuse in institutional settings. There are some studies targeting perpetrators of abuse, but where the abuse took place was not specified. Perpetrators of abuse in institutional settings have a different risk profile than those who abuse in other settings, so it is interesting that interventions combine perpetrators into a single group.

[6] The nearest is one RCT in the US (Rheingold) which worked with various types of 'childcare professionals', including teachers, probation officers, coaches, who were variously in schools, day care, healthcare settings, and churches. It does not split out results for the various settings or types of childcare professionals.

interventions around disclosure[7]. However, nine prevention studies included disclosure as an outcome. Some of the school-based prevention interventions were very effective at increasing disclosure as a side-effect[8].

Relatively few studies disaggregated results by sex. Of the 58 completed primary studies, fewer than half (only 24) reported results disaggregated by sex (i.e., analysed and reported any differences between males and females). 43 studies either did not conduct, or did not report, an analysis that would detect whether an intervention's effectiveness differed by gender.

No primary studies about treatment have begun in nearly 20 years. There are only two primary studies of treatment interventions: one non-randomised trial published in 1992, and one study (on which several papers are based) which started in 2000, looking at children raised in Romanian orphanages in the aftermath of the fall of the Ceauşescu regime in 1989.

Only four studies focus on children particularly at-risk. Most (n= 56) of the programmes studied in the primary studies were universal, i.e., were of 'general populations' of children and delivered in schools or early childhood settings (i.e., not focusing on children at risk). These were all prevention focused. Of the four primary studies focused on at-risk populations: one was about special education high school students with cognitive and/or physical disabilities; one was about boys in residential youth care; one was about children in Romanian orphanages; and one was about children sexually abused at a residential school for the deaf.

Only one study looks at educational attainment as an outcome (the *Good Schools Toolkit* study in Uganda). Clearly preventing abuse and dealing with abuse is important in its own right. Given the focus on education in the Sustainable Development Goals and the amount of funding tied to education, this was striking.

Institutional safeguarding practice[9] was studied in seven primary studies. This again is remarkably thin given the amount of activity in organisations on creating and running safeguarding policies. Four of these studies are about operational practice, and three about institutional culture. The outcomes studied in relation to institutional safeguarding practice were: number of cases registered with the authorities, teacher attitudes and their confidence in their ability to manage sensitive issues, and understanding of boundary-violating behaviours. There appear to be no studies of the impact of recruitment practices, structures for reporting disclosures, whistle-blowing procedures, on which we suspect that many organisations rely.

On the positive side, **research interest in this topic has risen dramatically**: before 2014, there were only 15 relevant studies; during and since then, 25 have been published[10], and we found protocols for three more planned primary studies (all RCTs). The other way of looking at that is that many of the studies are quite old:

[7] Sometimes a systematic review (and hence our discussion of it) will reference primary studies that *sound* relevant to the EGM but which are not included in our EGM. This will be because those primary studies did not fit our inclusion criteria e.g., the study might have been about disclosure but did not have a comparison group so did not meet the criteria about study design.

[8] A programme *Red Flag, Green Flag People* had 20 children disclosing vs none in the control group, when studied in both 1987 and 1989; despite small sample sizes, the *Good School Toolkit* generated over 400 additional referrals due to disclosures; and a programme in Spain had eight disclosures vs two in the control group.

[9] We use the term 'safeguarding' to refer specifically to actions taken by organisations that are designed to safeguard children.

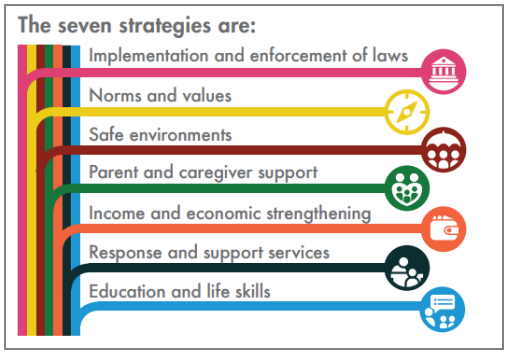
[10] In these numbers, because we are showing the volume of research interest, we do include the multiple research papers on the same raw data (e.g., of Romania, and the updated systematic review), hence these numbers add to more than the total given earlier.

nearly half (46%) the studies on our map published before 2012. Given the delay between a study happening and being published, that probably means that half the studies happened at least ten years ago.

A good range of age groups was studied: early childhood (0- 5 years) got 16 primary studies; middle childhood (6-11 years) saw most studies (38); early adolescence (12-14 years) had 12, and late adolescence 15-17 years had nine. Two studies were across childhood ages (0-18 years).

Very few studies came from practitioners and non-profits. This is not a great surprise, because (sadly) so few 'impact' studies produced by charities, etc., are robust, and hence few made the threshold for the types of study design that our EGM included. Nonetheless, it is striking that so few studies from non-profits met our criteria in terms of being fair tests.

We only found studies relevant to three of **the seven INSPIRE strategies** (pictured), promoted by the World Health Organisation. Those are: norms and values; response and support services; and education and life skills. In other words, we did not find rigorous evidence for four of those strategies, *as they apply to institutions*.



Implications for future research

The stark conclusion is that more research high-quality studies are needed: right across this terrain: across many institutional contexts and maltreatment types. The evidence gaps are particularly evident for low-income countries and countries with large populations. Few studies focused on adults, perpetrators or the organisational environment. We found gaps in the evidence around interventions relating to disclosure, organisational responses and treatment, and few studies that assessed an intervention's impact on perpetrators' maltreatment behaviours, recidivism and desistence. There is also need for more impact evaluations to report on what the programmes cost.

There are some weaknesses in the existing studies which serve as clues as to what future research should focus on and be conducted. We list here in order to inform future research.

Almost all the studies have appreciable risk of bias. We assessed the possibility that the studies might be biased: this uses the material in the study report, e.g., if the report doesn't describe a method of randomization, then we cannot be confident that randomization was done well (i.e., was unbiased), so there is a risk that it was biased. (To be clear, this is about confidence and risk: the study may have been conducted brilliantly, but if the method is not reported clearly, we do not know that. See box.)

We found no RCTs with low risk of bias (i.e., in whose results we can be highly confident). Of the 49 completed RCTs, 18 raised 'some concerns' of risk of bias, and the other had high risk of bias (i.e., we can have only low confidence that the reported results are accurate). The systematic reviews that we found were similar: ten of the systematic reviews are only low reliability, and only one is high reliability. The non-randomised primary studies (i.e., which used quasi-experimental designs) were somewhat better[4]: of the 18 of them, four had serious risk of bias, eight moderate, and six were low risk of bias.

Few studies look at actual incidence of abuse. This is hardly surprising, particularly for sexual abuse (which, as discussed, was the focus of most of the studies) because not all survivors ever report it at all, and some may take decades to do so. Only 10 completed primary studies looked at measures of actual child maltreatment occurrence or reoccurrence. These were usually self-reports from children/young people, which seemed to be used as proxies for incidence. All other studies look at intermediate outcomes, e.g., neural development, or education programmes assessed on whether children acquire and retain knowledge.

The sizes of effects are unclear in some studies. This is sometimes due to unclear reporting, and sometimes to strange use of measurement scales[11] or scales being invented for a study which prevents comparing its results with those of other studies.

Information about what a programme cost is included in very few studies: perhaps three at most. (For one intervention, the *Bucharest Early Intervention Project* (*BEIP*), we found cost information elsewhere.) Consequently, we do not know the cost-effectiveness of the interventions. Clearly for any organisation considering running a programme, cost is a factor so they need this information.

Many studies are worryingly small – so small that the reliability of their results is compromised. For instance, several programmes ran in fewer than seven schools: maybe only three schools got it and three did not. Any number of other factors could have influenced the results of such a small study. Studies need to be large enough to reliably distinguish between the effects of the programme and that of other factors including chance. They should report the 'power calculation' by which their sample size was determined.

The studies do not consistently report on who invented or ran the intervention. Some do, some don't. The papers about *Bucharest Early Intervention Project* say that the foster care programme was created by researchers and eventually supported by the local government; a few other studies say that the programme was run by an NGO (e.g., the *Stewards of Children* programme run by Darkness to Light, an American NGO). But most some studies do not state who ran the intervention/s. This is a shame because it might be possible to gain more information from the implementing agency's public materials.

The trials were generally quite short, which also compromises them. Most measured outcomes only up to about six months after the intervention ended. This is pretty short given that abuse can occur (or be perpetrated) any time in a person's life. We know that knowledge attenuates quite fast (people forget rapidly), so six months - or even 18 months - is unsatisfactory for showing the meaningful effect of a knowledge-gain intervention. The reason that many trials are short is money. We recommend that funders provide enough funding to enable follow-up periods long enough to investigate more enduring effects.

Few studies state the theory on which their intervention is based. This is a great shame because it greatly impedes a practitioner organisation, funder or policy-maker assessing whether it is likely to work in a different context. We recommend that future research includes this in the eventual study reports.

Many programmes are **'branded programmes'** meaning they are available on a commercial basis, often via non-profits working with, or set up by, research teams at US universities. Sometimes branded programmes

[11] For example, Merrill (2018) uses different ranges of Likert scales for different outcomes: one outcome is measured on a scale 0-3, but for another outcome, the scale is 0-12. This prevents us comparing them or identifying the size of the impact.

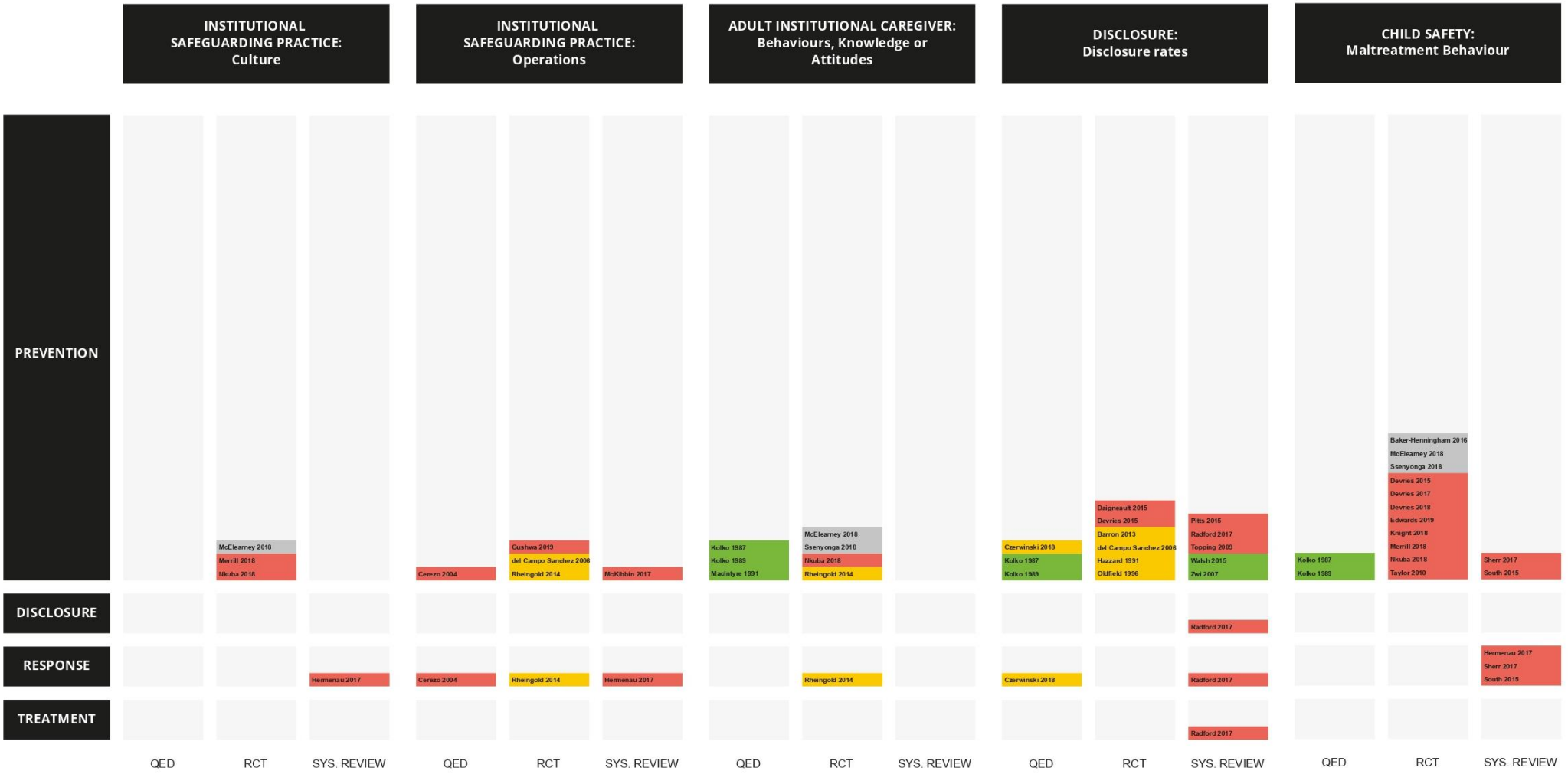
are evaluated by the programme designers, who sell the right to use the programme, which creates a clear conflict of interest: this is precisely what plagues pharmaceutical research, much of which is companies evaluating their own products. Unsurprisingly, in these branded social programmes (and pharmaceuticals) 'own-evaluations' find larger effects than do independent evaluations. There is thus a need for independent evaluations of programmes.

There is also an issue around branded programmes versus usual practice. The use of branded programmes is most pervasive in US education. The US What Works Clearing House[vi] lists 231 programmes to improve literacy. Surely there aren't 231 different ways of teaching children to read. The alternative approach is to identify the elements – or components – which matter in successful programmes and to build those into standard practice. Intensity and duration normally correlate to effectiveness of social programmes (unsurprisingly).

Finally, a note on using the research on the EGM and Guidebook:

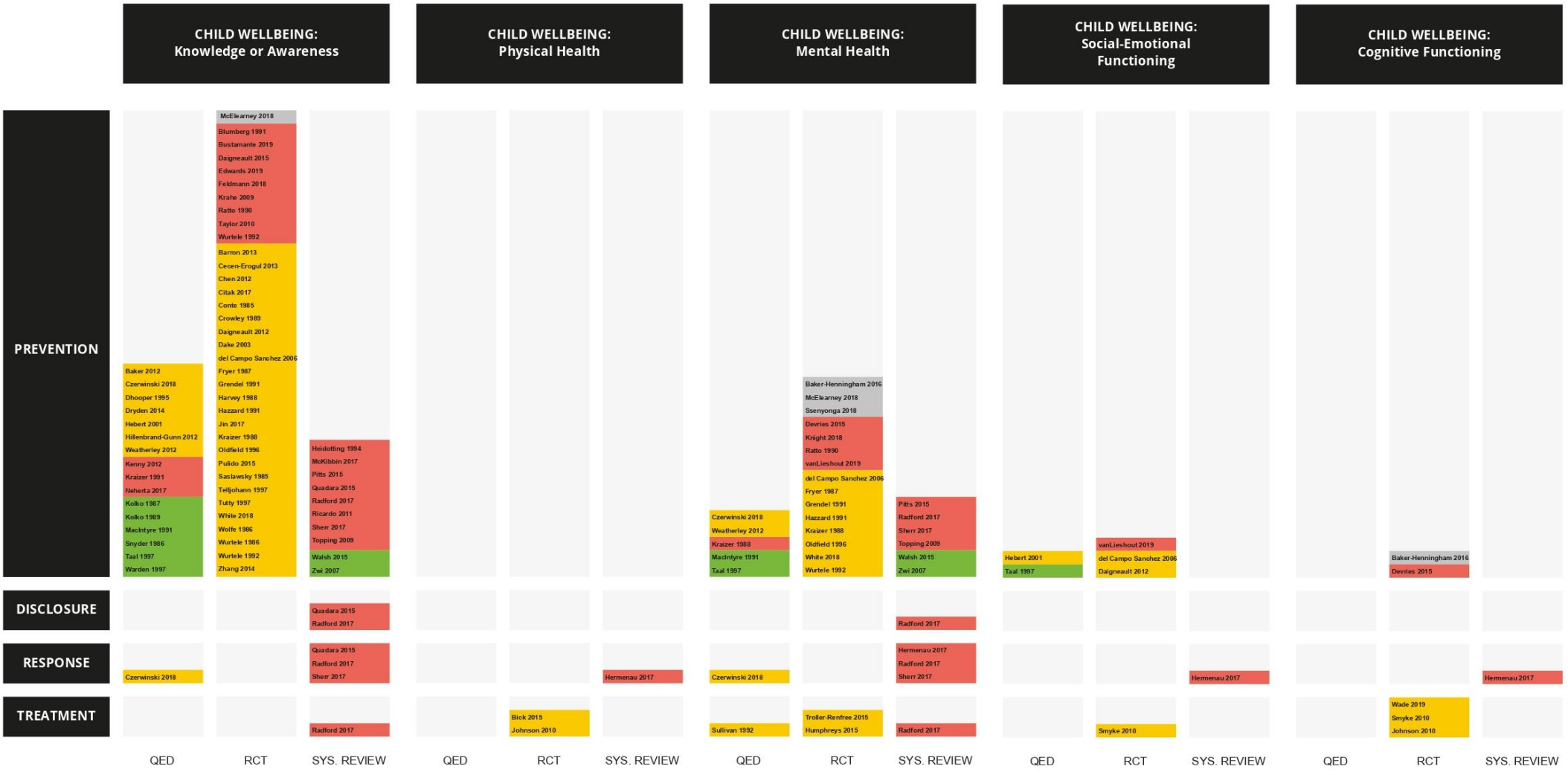
Research the intervention where it will be run. Clearly, if an intervention is moved from one type of setting to another, it may not work as well: for example, if an intervention which has worked in schools is now run in Scout groups, it may not work as well there. Interventions should be evaluated in the contexts where they are proposed to be run.

Evidence gap map of institutional responses to child maltreatment (Part 1 of 3)

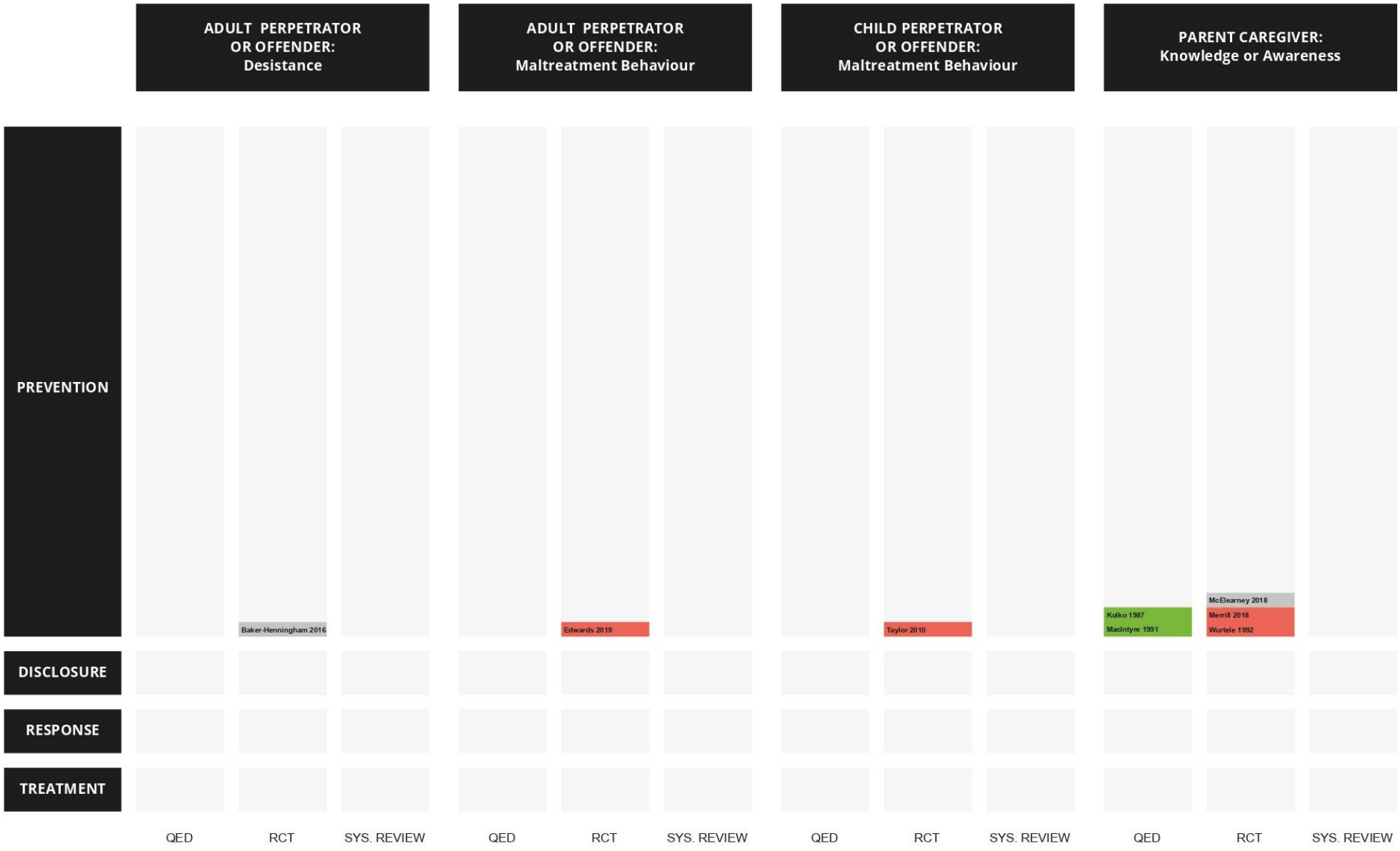


NOTE: We also look for studies which looked for the following outcomes: institutional safeguarding practice (environment); adult perpetrator or offender (recidivism); child perpetrator or offender (desistance); child perpetrator or offender (recidivism); and parent caregiver (knowledge or awareness). We did not find any. Those outcomes would have been shown as columns on this map, and they would have been empty. In the interests of making this map readable, we have removed those columns from this graphic.

(Part 2 of 3)


CHILD WELLBEING: Knowledge or Awareness
CHILD WELLBEING: Physical Health
CHILD WELLBEING: Mental Health
CHILD WELLBEING: Social-Emotional Functioning
CHILD WELLBEING: Cognitive Functioning
PREVENTION
DISCLOSURE
RESPONSE
TREATMENT
Baker 2012
Czerwinski 2018
Dhooper 1995
Dryden 2014
Hebert 2001
Hillenbrand-Gunn 2012
Weatherley 2012
Kenny 2012
Kraizer 1991
Neherta 2017
Kolko 1987
Kolko 1989
MacIntyre 1991
Snyder 1986
Taal 1997
Warden 1997
McElearney 2018
Blumberg 1991
Bustamante 2019
Daigneault 2015
Edwards 2019
Feldmann 2018
Krahe 2009
Ratto 1990
Taylor 2010
Wurtele 1992
Barron 2013
Cecen-Erogul 2013
Chen 2012
Citak 2017
Conte 1985
Crowley 1989
Daigneault 2012
Dake 2003
del Campo Sanchez 2006
Fryer 1987
Grendel 1991
Harvey 1988
Hazzard 1991
Jin 2017
Kraizer 1988
Oldfield 1996
Pulido 2015
Saslawsky 1985
Teljohann 1997
Tutty 1997
White 2018
Wolfe 1986
Wurtele 1986
Wurtele 1992
Zhang 2014
Heidotting 1994
McKibbin 2017
Pitts 2015
Quadara 2015
Radford 2017
Ricardo 2011
Sherr 2017
Topping 2009
Walsh 2015
Zwi 2007
Quadara 2015
Radford 2017
Quadara 2015
Radford 2017
Sherr 2017
Czerwinski 2018
Hermenau 2017
Bick 2015
Johnson 2010
Radford 2017
Czerwinski 2018
Weatherley 2012
Kraizer 1988
MacIntyre 1991
Taal 1997
Sullivan 1992
Baker-Henningham 2016
McElearney 2018
Ssenyonga 2018
Devries 2015
Knight 2018
Ratto 1990
vanLieshout 2019
del Campo Sanchez 2006
Fryer 1987
Grendel 1991
Hazzard 1991
Kraizer 1988
Oldfield 1996
White 2018
Wurtele 1992
Troller-Renfree 2015
Humphreys 2015
Radford 2017
Pitts 2015
Radford 2017
Sherr 2017
Topping 2009
Walsh 2015
Zwi 2007
Radford 2017
Hermenau 2017
Radford 2017
Sherr 2017
Hebert 2001
Taal 1997
Smyke 2010
vanLieshout 2019
del Campo Sanchez 2006
Daigneault 2012
Hermenau 2017
Baker-Henningham 2016
Devries 2015
Wade 2019
Smyke 2010
Johnson 2010
Hermenau 2017
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW

(Part 3 of 3)


ADULT PERPETRATOR OR OFFENDER: Desistance
ADULT PERPETRATOR OR OFFENDER: Maltreatment Behaviour
CHILD PERPETRATOR OR OFFENDER: Maltreatment Behaviour
PARENT CAREGIVER: Knowledge or Awareness
PREVENTION
DISCLOSURE
RESPONSE
TREATMENT
Baker-Henningham 2016
Edwards 2019
Taylor 2010
Kolko 1987
MacIntyre 1991
McElearney 2018
Merrill 2018
Wurtele 1992
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW
QED
RCT
SYS. REVIEW

Box 1: What are primary research and systematic reviews?

Primary research is a study of people. It can involve questionnaires, surveys or interviews, or other measurements about people such as their income, height, or scores in tests.

A systematic review is a study of studies. It is a structured investigation to find, critically appraise and synthesise all the relevant primary research on a specific topic. Systematic reviews are stronger than non-systematic 'literature reviews' in that they: (i) can reconcile differences in the conclusions of different studies by looking across a larger set of participants, (ii) identify gaps to inform further research, (iii) are more transparent and hence can be reproduced by other researchers in future and (iv) are less prone to bias, as science writer, doctor and Oxford academic Ben Goldacre explains:

"Instead of just mooching through the research literature consciously or unconsciously picking out papers that support [our] pre-existing beliefs, [we] take a scientific, systematic approach to the very process of looking for evidence, ensuring that [our] evidence is as complete and representative as possible of all the research that has ever been done."

Thus a systematic review is more likely to be *accurate* and hence useful to practitioners for informing research and programme design than non-systematic literature. It is also more *credible* and hence useful in terms of convincing funders and policy-makers.

Each systematic review defines a ***scope*** (the topics, geography and timescale of interest) and the way that it will search for studies with that ***remit*** (the 'search strategy'). Most set some threshold for the ***quality*** of the primary studies they include in their analysis (the importance of quality of primary studies is discussed in Box 2). This is significant because the systematic review process is not magic: if the primary studies on which a systematic review is based are unreliable, the review's results will be unreliable. As a Yale cardiologist wrote recently on Twitter (Krumholz 2015): '*You can't just combine weak evidence and pretend that when mushed together it is strong. [Rather] it is meta-mush*.'

Box 2: Why we only include RCTs and other studies with robust counterfactual

In short, because different research methods give different answers. The choice matters. We only want the true answer.

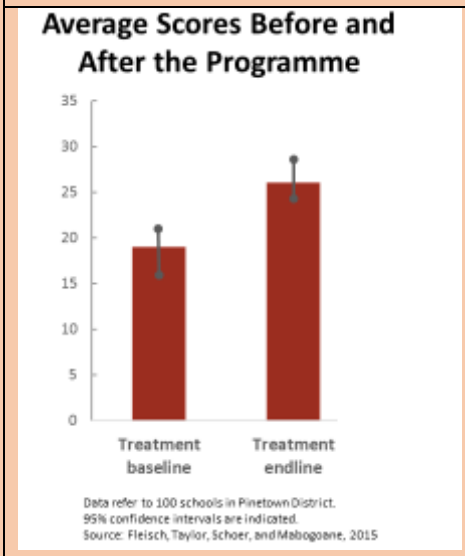
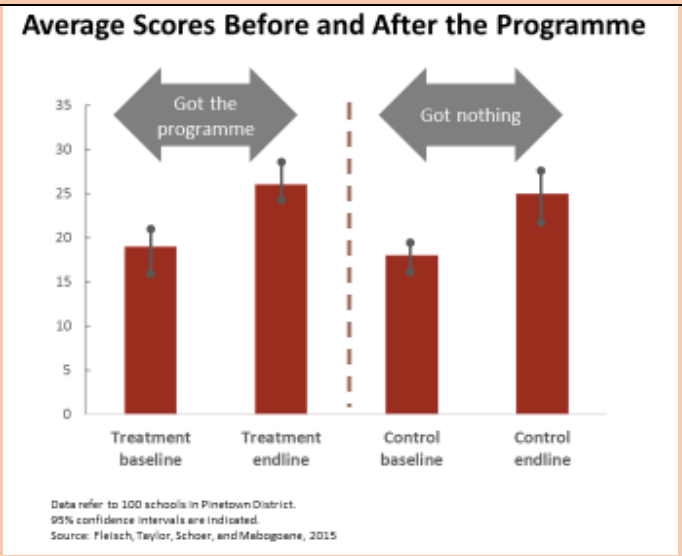
The table below shows the effect of one reading programme in India measured using several research methods. These methods all used the same outcome measures, but the experimental designs were different. The answers vary widely: some show that it works well, others show it to be detrimental. Clearly there can only be one correct answer. All the other answers are incorrect, and may lead donors or practitioners to implement this programme at the expense of another which might be better. The answers vary because research methods vary in how open they are to biases (i.e., systematic errors).

Method	Impact Estimate
(1) Pre-post	26.42*
(2) Simple Difference	-5.05*
(3) Difference-in-Difference	6.82*
(4) Regression	1.92
(5)Randomized Experiment	5.87*

*: Statistically significant at the 5% level

Source: Innovations for Poverty Action

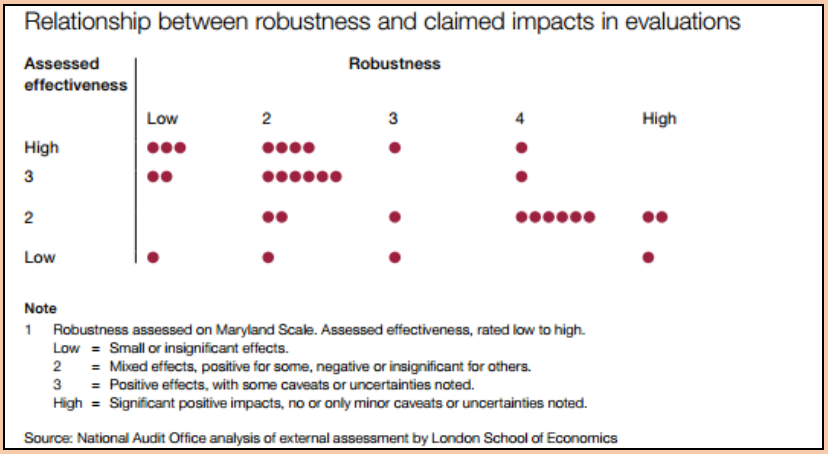
Another example is this remedial education programme in KwaZulu-Natal. Examined just by looking at children's reading ability before the programme and after it, it looks like the programme works. But careful examination with a randomised controlled trial shows that children who do not do the programme also make progress during it: in fact, they make precisely the same amount of progress. The programme achieves nothing: the apparent improvement is simply due to the passage of time (see below).

Simple pre / post assessement: average scores before vs after the programmae	**Assessment using randomised controlled trial: average scores before vs after the programmae**
	

Weaker research allows for more positive findings

The pattern above happens all the time: that weaker research gives totally different answers than does rigorous research.

The National Audit Office examined this in UK government studies. It searched for literally every published evaluation of a government programme. Of those, it chose a sample, and ranked on one hand, the quality of the research method ('robustness' on the x axis), and on the other, the positive-ness ('claimed impact'). It shows that more robust research only allows for modest impact claims whereas weak research allows much stronger claims. Bad research can be persuaded to say almost anything.



Bad research can be persuaded to say almost anything, and won't allow researchers to distinguish the effects of a programme from other factors (e.g., the passage of time, the mindset of participants, other programmes) nor from chance.

Most social interventions have a small effect and a reliable research method will show what that is: bad research is likely to overstate it. The highest estimate for the reading programme above is from the pre-post study which is a weak study design.

This relationship between weak research methods and positive findings has been shown in many fields including in education and in medical research.

Box 3: Risks of bias in trials, including randomised controlled trials

Suppose that you run an RCT of a pharmaceutical drug. Your trial will involve measuring the health (on some metric/s) of participants at the beginning and at the end. Now, during any trial, some people will drop out: perhaps they move away from the area, or become unwilling to take the drug. But suppose that participants die *because the drug kills them*. If you only measure the health of people who are still in the trial at the end, you would miss the – rather important! – fact that the drug is fatal for people. The ostensible result is biased towards people who weren't killed by the drug – perhaps the drug only kills people aged over 70. In that scenario, your results have 'survivor bias', which will make the drug look more effective than it really is.

For this reason, it is important that trials report reasons for 'attrition', i.e., how many people dropped out (i.e., their data is lacking at the end), which types of people they were (so we can see that, in our example, age was a determinant), and why people dropped out insofar as it is known.

Other sources of bias include how the randomisation was done. Many methods can be used to obtain an apparently random allocation. Dr Ben Goldacre explains[vii]:

"*Let's imagine there is a patient who the homeopath believes to be a no-hoper, a heart-sink patient who'll never really get better, no matter what treatment he or she gets, and the next place available on the study is for someone going into the 'homeopathy' arm of the trial. It's not inconceivable that the homeopath might just decide—again, consciously or unconsciously—that this particular patient 'probably wouldn't really be interested' in the trial. But if, on the other hand, this no-hoper patient had come into clinic at a time when the next place on the trial was for the placebo group, the recruiting clinician might feel a lot more optimistic about signing them up.*

The same goes for…tossing a coin. [F]orgive me for worrying that tossing a coin leaves itself just a little bit too open to manipulation. Best of three, and all that. Sorry, I meant best of five. Oh, I didn't really see that one, it fell on the floor."

People have studied the effect of randomisation in large numbers of trials and found that the ones with dodgy methods of randomisation overestimate treatments effects by fully 41%[viii]. The method of randomisation was a cause for concern in 20 of the 33 completed RCTs included in our EGM.

Biases can dramatically alter the answer that trials give. For that reason, The Cochrane Collaboration lists various possible sources of bias[ix] such as this, and we assess reports of RCTs against that list to determine their 'risk of bias', reported as quality.

All of the completed RCTs in our EGM raised concerns about their selection of reported results: that leaves open the possibility that the researchers measured loads of outcomes but only reported on the ones that gave the results that they wanted.

Box 4: The Context of Abuse and Institutional/Organisational Settings

In recent years, child maltreatment in institutional settings has received high public and policy recognition through a range of official inquiries particularly in high-income countries (e.g., the Royal Commission into Institutional Responses to Child Sexual Abuse in Australia, 2017; the Scottish Child Abuse Inquiry). These inquiries have led to a prioritisation of child maltreatment within institutional settings, as both a specific and serious issue among policy-makers, practitioners and service agencies working with children. In addition, the inquiries themselves have produced many key reports examining the impact of institutional child maltreatment, how it can be prevented, victims supported, and appropriate responses implemented.

Institutional settings include schools, out-of-home care, sport clubs, religious institutions and comparable settings in which children live or spend time. In these settings, child maltreatment can be adults abusing children or children abusing other children. Children may be more or less vulnerable for reasons ranging from a lack of proper safeguarding in institutions (e.g., failing to respond to disclosures), to the characteristics of children (e.g., age, developmental or other disabilities). Institutional child maltreatment as a field of empirical research is at an early stage. Whilst recent prevalence studies in residential care facilities suggest that children are at higher risk of sexual abuse compared to the general population, there has been virtually no comparison of other types of maltreatment in other settings.

Box 5: Definitions of the categories of intervention: prevention, disclosure, response, treatment

The EGM divides the interventions into four categories - prevention, disclosure, response, treatment – which are defined as follows.

Prevention interventions were defined as any intervention where the primary aim was to decrease the likelihood or risk of child maltreatment occurring or recurring in the future. This encompassed both interventions for any child / adult ('universal populations'), as well as interventions targeted at specific populations. Examples of types of prevention interventions that could be included were school-based safety programmes, organisational guidelines or practices, or perpetrator targeted interventions to reduce reoffending.

Disclosure interventions were defined as any intervention that aimed to facilitate, support, or promote the disclosure of child maltreatment. This encompassed a range of universal interventions, such as traditional or social media campaigns, or child helplines, as well as therapeutic interventions for children that aimed to promote disclosure (e.g., play therapy). It included tertiary interventions relating to perpetrators, such as mandatory reporting, and also included any intervention that aimed to promote disclosure within an organisational context (e.g., staff training, organisational guidelines).

Response interventions were defined as any intervention that aimed to improve institutional responses to child maltreatment in relation to each of the target populations. Response interventions included enhancing safeguarding practices, legal and policy interventions, supporting the victim and/or family, working with child protection agencies, and providing training and crisis support to staff within organisations.

Treatment interventions were defined as any intervention that aimed to provide a therapeutic response to a target population. This included therapeutic interventions provided to children who experienced child maltreatment in institutions, and interventions targeted at institutional perpetrators of child abuse. The Romania studies are included here, because foster care was provided as treatment for young children who spent their early lives in institutionalised care.

Appendix: About Porticus and Giving Evidence

Porticus is an international organisation managing and developing the philanthropic programmes of charitable entities established by Brenninkmeijer family entrepreneurs. Porticus is involved with and fund a broad range of social service activities, including to both faith-based organisations, and organisations unrelated to religious institutions.

Porticus commissioned this EGM and Guidebook to further support its own and others' ongoing work to enhance organisational safeguarding.

Porticus has supported, and currently supports, efforts among its grantees to improve organisational safeguarding of children. Specifically, Porticus:

- requires from all its grantees to have a safeguarding policy
- works with some grantees to further develop interventions that can make the organisations safer.

These projects are conducted in collaboration with both faith-based organisations and non-faith-based organisations. To ensure that all standards for the production of a Campbell EGM are met, Porticus was not involved in any technical steps taken to produce the EGM or this Guidebook, including information retrieval, data analysis and reporting of findings.

Giving Evidence is a consultancy and campaign, which enables and encourages charitable *giving* based on sound *evidence*.

Through consultancy, Giving Evidence helps donors and charities in many countries to understand their impact and to raise it. Through campaigning, thought-leadership and meta-research, we show what evidence is available and what remains needed, what it says, and where the quality and infrastructure of evidence need improving. We have advised many donors in many countries on many issues.

Giving Evidence was founded by Caroline Fiennes, a former award-winning charity CEO, and now Visiting Fellow at Cambridge University's Centre for Strategic Philanthropy. She wrote the *How To Give It* column in the Financial Times for three years, the first column about philanthropy in any major newspaper globally. She is author of the acclaimed book *It Ain't What You Give, It's The Way That You Give It*, which is a guide for donors. She has also written in Freakonomics, the Daily Mail and spoken at TED, and is one of the few people whose work has appeared in both *OK! Magazine* and the scientific journal *Nature*.

———

The EGM was produced with the Centre for Evidence and Implementation and Monash University. The Guidebook was produced with the Campbell Collaboration, and some input from the Africa Centre for Evidence.

[i] The WHO estimates that 1 billion children aged 2–17 years, have experienced physical, sexual, or emotional violence or neglect in the past year alone, in families or elsewhere: https://www.who.int/news-room/fact-sheets/detail/violence-against-children [accessed 13 November 2019]

[ii] *Effect of early institutionalization and foster care on long-term white matter development: a randomized clinical trial*, Bick et al, 2015, JAMA Pediatr. Doi: 10.1001/jamapediatrics.2014.3212

[iii] *EVIDENCE GAP MAP ON ADOLESCENT WELL-BEING IN LOW- AND MIDDLE-INCOME COUNTRIES: PROTECTION, PARTICIPATION, AND FINANCIAL AND MATERIAL WELL-BEING*

https://www.unicef-irc.org/evidence-gap-map/ and https://www.unicef-irc.org/publications/931-bridging-the-gap-to-understand-effective-interventions-for-adolescent-well-beingan.html [accessed 13 November 2019]

[iv] *The prevalence of child sexual abuse in South Africa: The Optimus Study South Africa*, 2019, S Afr Med J 2018;108(10):791-792. DOI:10.7196/SAMJ.2018.v108i10.13533 http://www.samj.org.za/index.php/samj/article/viewFile/12449/8654%20 [accessed 13 November 2019]

[v] School-based education programmes for the prevention of child sexual abuse, 2015, Walsh et al, DOI: 10.1002/14651858.CD004380.pub3, https://www.ncbi.nlm.nih.gov/pubmed/25876919 [accessed 13 November 2019]

[vi] https://ies.ed.gov/ncee/wwc/

[vii] Goldacre, B., 2008, *Bad Science*, London: 4th Estate

[viii] Ibid, p. 49-50.

[ix] RoB 2: *A revised Cochrane risk-of-bias tool for randomized trials*, https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials [accessed 14 November 2019]

Appendix: Summary of scope of the EGM

The EGM covers studies of the following:

Population	Children aged 0-17 years at the point of baseline measurement, living in and / or engaging in activities **in institutional settings**
Geography (of researcher or the intervention studied)	Anywhere
Language of the study	English, German, French, Spanish, Italian, Portuguese, Dutch, Danish, Swedish, Norwegian
Status	Finalised and on-going studies
Setting	**Institutions** (We list them. Including: Day care, school, sports clubs, churches / religious institutions, camps, residential care including orphanages). It does not include in-family care.
Intervention:	We list them in the protocol and longer report. They include: prevention, disclosure, response, treatment. The 'target' can be child victim, child offender, adult perpetrator, organisational leadership, organisational staff, caregiver / parent. Delivery mode might be individual, group or other.
Outcomes	We list them in the protocol and longer report. They include: institutional practice, disclosure rates, occurrence, child cognitive and educational performance, child health (physical & mental), child social functioning, offender behaviour (adult and child offenders), recidivism
Study design	Primary studies: impact evaluations which have a reasonable counterfactual. We list them, and they include randomised controlled trials, quasi-experimental designs such as propensity score match or regression discontinuity designs, non-randomised trials with at least two intervention sites and two control sites. Secondary studies: systematic reviews and meta-analyses.