

Improving Clarity and Usefulness of Research by Charities:

A checklist of items for charities to include
in published reports of impact research

Caroline Fiennes

“There are four key steps where research can go wrong, which is contributing to an enormous amount of research waste: not asking the right research questions, poor research design, non-publication of research, and poor reporting of research.”

- Paul Glasziou, Professor of Evidence-Based Medicine at Bond University in Australia, and former Director of the Centre for Evidence-based Medicine at Oxford¹

“If I have seen further it is by standing on the shoulders of giants.”

- Isaac Newton

Contents

Introduction and purpose	3
The proposed items to include in reports of impact research	5
Box 1 - Example Structured Abstract	7
Box 2 - Describing an Intervention	8
Why should I report this way?	9
Why should I ask organisations I support to report this way?	
Appendix 1: Fewer than 20% of charities in crime reduction and criminal justice publish adequate evaluation materials	10
Appendix 2: British Medical Journal ‘trial reports’: an example	12
Appendix 3: Bad research flatters more than good research	13

Introduction and purpose

Our purpose is to enable work by charities and others to be based on evidence which is as sound as possible. Charities¹ should not always *produce* research – they often should rather *use* existing research. But when they do produce research, it should enable best practice to spread and to spread fast.

This project concerns reporting of research: the last problem on Paul Glasziou’s list above. That is not because we think that reporting is the sole problem with research – we don’t. Research also needs to address important questions, provide reliable answers, and be published and findable. However, though fixing research reporting isn’t sufficient, it is necessary.

By ‘research’, we mean research by charities which investigates causal links between inputs and outcomes, and which could be useful to external audiences. That includes research into the effectiveness of their interventions, and research to inform policy debates, e.g., about the effectiveness of prison, or public attitudes. Both types of research can be useful to practitioners in their own organisations, and should be used by academics, the police, commissioners and others. We are not so concerned here with research for internal purposes such as process evaluations.

For charities’ research to be useful, it should be high quality, published somewhere findable, clear, and disseminated / available to its intended audiences.

This study began with two observations about much research by charities. First that it could be easier to find: since there is no central repository, material is often published just on charities’ own websites where readers may not find it. Second that it could usefully be clearer – about the intervention, the research question and method, and the findings. We consulted on whether research by charities in criminal justice could/should:

- (a) **Always be published with a few key details. These might include: (i) the intervention used, (ii) the research question/s, (iii) the research method and how it was used (e.g., if 20 people were interviewed, how were those 20 chosen?), and (iv) the results.** This is essentially a checklist of items to be included in the research report. And/or:
- (b) **Sit in a central repository.**

Both proposals met with a great response. Some items were added to the checklist.

This document details the items which we suggest charities include in reports of their impact research. A much fuller report on the consultation responses is at www.giving-evidence.com/info-infrastructure, detailing how evidence is currently produced / shared / used by criminal justice charities, how research reporting works in medicine, how these proposals relate to the Justice Data Lab and government’s What Works Centre on crime reduction, concerns about them, incentives on charities to produce research, how the repository might work, detail about possible solutions to the repository, relevant academic centres.

In the consultation spirit, we welcome your perspective: please send your thoughts to admin@giving-evidence.com.

¹ We use the term ‘charity’ to mean any voluntary or community organisation including social enterprises. We use ‘charity’ and ‘NGO’ interchangeably.

Example responses:

On improving clarity of research reports, and the quality of research:

“Absolutely, absolutely. A thousand times, yes.” Alex Murray, Police Chief Superintendent, Solihull; Founder, Society for Evidence-Based Policing

“Great idea! Would make my life a lot easier. We are, in effect, buying outcomes, so need to be able to see what they’ve done and what the outcomes really were” – funder

“I would love this... It would be wonderful... IF you can get people to do it [because of the incentive problems]...” – funder

“There isn’t anything systematic... it would certainly be useful” – major funder

“A charity I chair just got a grant from [a particular foundation], of which half – £5,000 – was for evaluation. I said to them that that’s ridiculous, and kind of unfair. We obviously can’t do decent research with that” – operating organisation

“When I started in this [the charity sector], I kept talking about evaluation and he [senior person in the charity sector] said to me ‘don’t worry about that. You can just make it up. Everybody else does. At the very least you should exaggerate a lot. You’ll have to, to get funded’ ” – operating organisation

“It [charities’ research] is often ridiculously complementary! ... a flood of positive results” – funder

“Honestly some of what they [charities] say it is just so entertaining. Nothing goes wrong! Magic!” – former funder

“It can be really hard to figure out from the research what they’ve actually done. Well, not hard: impossible, because they just don’t say” – major funder

On describing what the intervention was and how it worked:

“Did everybody get the same thing? Was the delivery consistent? Was the delivery as per the design? Was the dosage and quality as intended? What about staff turnover? It’s not uncommon for different users to get quite different services – just based on which staff member they happened to get. So then you don’t even know what the evaluation is really evaluating” – major funder

“Also, you normally can’t see who was turned away. That’s pretty important [since the cohort so dramatically affects success rates]” – practitioners and funder

“No-one properly describes their cohort. Funders expect a success rate of about 70%, so magically that’s what everybody has, though they patently have quite different client groups. There’s pretty obviously a lot of lying” – former director of large funder²

On making research reports easier to find:

“Yes it would help. The system is such a shambles now that ... well the hurdle you describe is pretty low but clearing any hurdle now is a good thing to do. And this could drive awareness of the shambles and how to improve, so yes ... excellent” – academic / practitioner

“I certainly think that there’s a gap in our sector. It’d be wonderful to have a place that we’d point to for all the research” – Lis Bates, Head of Research and Evaluation, CAADA

“This would be a great contribution” – NGO in criminal justice

“You have to look on every site” - Carol Jackson, Head of Assessment, The Prince’s Trust

The proposed items to include in reports of impact research

Good reporting about research should clearly answer ‘**what did you do and what did you find?**’. We consulted on the notion that any research published by charities should include details of at least the following:

- i) **The intervention:** in enough detail that it could be replicated elsewhere. [See Box on describing an intervention.] For instance, if it is an education programme, how highly trained are the teachers, what is the class size, the session duration, the content of each session, the number and spacing of sessions, and physical classroom environment.
- ii) **The research question.** This may be a monitoring-type question (e.g., what is the demographic breakdown of our beneficiaries?); or it may be an evaluation-type causal question (e.g., what effect does this intervention have on rates of bike theft?).
- iii) **The research method, and how it was used.** Details such as: the sample size (since this determines statistical significance); how they were recruited (since this determines the reliability of the sample: views of 20 randomly-chosen beneficiaries is more reliable than if the programme co-ordinator chooses which 20 beneficiaries to include). If they were interviewed (e.g., for qualitative surveys), how long and where was the interview, and what were they told it would be used for.

Items (ii) and (iii) show the quality of the research and the appropriateness of the research method used. The importance of these are shown in Appendix 3.

- iv) **The results,** e.g., proportion of people employed, breakdown of beneficiaries by socio-demographics.

Nobody wanted any of those items removed. Suggestions of items to add to the checklist included the following. Clearly too many items will make the checklist too long to be useable, so we may need to make some choices. An option is to state some items as ‘essential’ and others as ‘desirable’.

- v) **Detailed description of the cohort of beneficiaries** (research participants). Clearly the success of, say, a back-to-work programme or programme to reduce re-offending vastly depends on the type of people who receive it.
- vi) **Who conducted the research and who paid for it.** The former (e.g., whether it was conducted by an academic, or consultants for hire) allows the reader to estimate quality and motivation, and the latter speaks to incentives. Including details of both is standard in medical research (see examples of British Medical Journal trial reports [in the full report](#)) since, as discussed, studies funded by pharmaceutical companies seem to get different answers than identical studies funded by other sources⁴.
- vii) **Programme cost.** Ideally the total programme cost, the set-up cost, and the unit cost. Clearly this is essential for comparing programmes and deciding between them: all funders and charities are allocating finite resources. (Remarkably, the National Audit Office found that 77% of government evaluations – i.e., analysis of work funded by taxes – included no cost-effectiveness data at all⁵.)

- viii) Theory of change and the evidence for it.** That is, how is the programme *supposed* to create change? The theory of change can help others to gauge whether the programme will produce similar results in their context, and the evidence for the theory of change helps show whether the results were as expected or were quirky outliers.
- ix) Context in which the programme was run, and support for implementation.** The research will primarily be useful for other organisations deciding what to run in their contexts, so detailed description of the situation is necessary. Insights from process evaluations (e.g., were the people served actually the group who the charity set out to serve? Was the intervention delivered actually what the charity set out to deliver?), insights about what is necessary for it to succeed², contexts in which it is likely to work, and materials which can help others to implement it, such as training manuals.
- x) Where there is a control group (randomised or propensity matched or constructed from other methods), **describe what the control group got.** This is essential because otherwise, the reader may assume that the control group got nothing, which will overstate the effect of the treatment being tested if they did get something useful. Describe it fully: as one academic said: “don’t just say ‘probation service as normal’ because in ten years’ time, we won’t remember what that means”. It may be appropriate also to **give the counterfactual cost**, i.e., the costs which arise if the intervention is not done.
- xi) How the researchers **guarded against bias.** For instance, how were people recruited into the research (the potential bias being that only the most optimistic ones offer to be interviewed); who dropped out during the programme (since perhaps people who hate the programme drop out, leading to ‘survivor bias’); how did you ensure that people who were not eligible for, say, a food programme didn’t cadge food from those who were and hence effectively sneak into the programme.
- xii) How we can tell that the results **aren’t simply a product of chance.** Programmes can appear to have an effect which in fact is nothing more than chance. This is particularly likely when they have a small sample size, or there is no robust counterfactual (showing what would have happened to that cohort anyway, in the absence of the programme).

In fact, if the research is well-described, the answers to these latter two questions will be findable. The purpose of asking them explicitly is to (a) make it easier for the reader to find the answers, and (b) alert the researchers to their significance, and hence train them to consider them.

Reporting these items: structured abstracts

Ideally the items on the checklist should all be answered in a structured abstract at the front, since studies have found that structured abstracts are higher quality than traditional descriptive abstracts^{7, 8} and that they allow readers to find information more easily.^{9,10} An example a structured abstract is below (and an example structured British Medical Journal clinical trial report is in Appendix 2).

² Some of this is very practical. The police cite ‘alley-gating’ as an example: sealing alleys behind terraced housing does reduce burglary but is laborious and slow because permission is required from every householder.

Box 1 - Example Structured Abstract

The structured abstract is envisaged as a short front-sheet to a charity's research report. The charity can supplement it with whatever it wants: the structured abstract is simply minimum criteria to cover. The following example is based on research by St Giles Trust¹¹ on its WIRE project, which works 'with female ex-offenders to guide them from release, assist with resettlement and reconnect them with the community'. (For concision, the example may omit some details. The purpose here is to illustrate the structure, not describe WIRE.)

Intervention

The service is intentionally flexible, the workers respond to each individual woman's needs, as a result service provision can be radically different in that not all women require the same intensity of support. What is consistent though is the commitment to provide more than a referral to another agency, the WIRE advocates on behalf of these women, counsels them on the correct decision for them, provides information, will accompany them to appointments and provides firm emotional support. The service involves advocacy and signposting.

Stage 1 involves helping the women with their immediate housing needs, making appointments at doctor's, Homeless Persons Units and Drug Intervention Programmes, etc.; often the women are escorted to one or all of these. This is the most intensive part of the programme (taking hours or days of staff time) and without a client's ID none of these are possible. It is important to ensure women meet their probation and stay within the details of their licence.

Stage 2 relies on the women returning to the service as often staff prioritise new releases. When they do return they are supported with access to solicitors to help reconnect them with children, or to organisations that can help them with any issues around their mental health.

Research question

Understanding the WIRE's impact on re-offending.

Research method

Analysis of 364 cases.

Interviews (25, with staff (5), clients (8) and service delivery partners (12)).

Analysis of client conviction rates.

(Comparison group data from the Police National Computer was unavailable.)

Findings

- WIRE was achieving more outcomes with clients in the early stages of the service rather than further on.
- Seen as an effective service.
- Reconviction rate for the eligible women was 42%, against 51% for the national average for women offenders and 88% for prolific offenders.
- Overall, WIRE was seen as having a substantial impact on reducing re-offending and offending frequency amongst those who are deemed as being at high risk.

Box 2 - Describing an Intervention

Medical researchers ask that the description of an intervention “should allow a [practitioner] to know exactly how to administer the intervention that was evaluated.” The medics have created a 12-point checklist for describing interventions, the Template for Intervention Description and Replication (TIDieR), which is helpful and could easily be adapted for charities in crime reduction (and other sectors):

- The name of the intervention (brief name or phrase)
- The way it works (rationale, theory, or goal of the essential elements)
- What materials and procedures were used (physical or informational)
- What was the procedure (activity / process)
- Who provided the intervention (e.g., nurse, psychologist, and give their expertise and background)
- How was it delivered (e.g., face to face, online, by phone, and whether it was provided individually or in a group)
- Where it took place
- When and how much (the number of sessions, schedule, dosage and duration)
- Tailoring (what if anything could be adapted to the individual, why and by how much)
- Modifications which happened after the study started
- How well was adherence to the plan assessed (i.e., the process for assessing adherence)
- To what extent did implementation adhere to the plan.

This makes for long descriptions, much longer and more detailed than most charities’ descriptions. The full report has an example description which medical researchers view as adequate, and notes how much more detailed it is than the the description of an example intervention analysed by the Justice Data Lab.

Why should I report this way? Why should I ask organisations I support to report this way?

Because it makes the research more useful to other organisations:

- If the intervention seems to work, other organisations can replicate it, so more people can benefit.
- It will prevent people being misled into doing / funding ineffective work. Research quality matters: bad research gives *different* answers to good research, and gives *misleadingly positive* answers: see Appendix 2. When organisations are considering work to do / fund, they need to see the quality of the underlying research, which is only possible if these details are available.
- Sectors may build systems to enable users to search for research by using the categories in the checklist, which can serve as repositories, e.g., a system in which charities upload their research wherever they want, and a central site with a crawler bot would pull the relevant details in response to search.

The editor of the major medical journal which published TIDieR [wrote recently](#) that it was the most important articles that the journal has run. They are ‘akin to turning on a light in a room. It doesn’t clean the room for you [but] tells you what the room looks like.’

We hope that eventually *all* research by charities will be published in this way: not just research conducted in future but also the ‘back catalogue’ of completed high-quality research, since decisions today are based on the whole available evidence base, which includes material from some years ago³.

Towards the goal of improving results by enabling charitable activity and giving to be based on sound evidence, the checklist and research infrastructure discussed here seem important and necessary steps. We found enthusiasm for them in several other parts of the charitable sectors and can imagine trialing this system in other sectors also. We are excited to take this forward.

³ The AllTrials campaign aims to get the results of all clinical trials published, not just those which flatter the researchers. When the drug company Roche announced in 2013 that it would attempt to publish all its research from that point on, the campaign director Tracey Brown responded that: “Does Roche expect applause for announcing that it will continue to keep clinical trial findings hidden? Which bit of All and Trials do they not understand?”

Appendix 1: Fewer than 20% of charities in crime reduction and criminal justice publish adequate evaluation materials

Research conducted for Giving Evidence by Leah Ginnivan.

Background

To begin to assess the prevalence and quality of evaluation research published by charities working in the UK criminal justice sector, we conducted a short review of their published evaluation materials.

Method

To find charities, we used the membership list of Clinks, and randomly selected a 5% sample (by assigning each organisation a random number, and then filtering the results). Clinks has 617 members, and so the 5% sample comprised 31 organisations.

Clinks members without websites were excluded and replaced with the next randomly selected organisation. Organisations that provided a private service (such as individual counselling to people in the criminal justice system) were included. Groups that offer services such as prayer or social support were included, since they could collect some data on their impact.

We spent up to 20 minutes searching for evaluation materials on each website. It's possible that some organisations did have evaluation materials on their website that were not found during this search. However, the 20 minute time limit was a way of searching for material that was intended to be shared.

We collected results for four categories:

- Whether the organisation published any results or evaluation about the impact of its programmes. Operational data (for instance number of staff employed) was not included.
- Whether there was information available about the cohort that was treated (for instance, whether they were helping young men facing court or children with parents convicted of criminal offences).
- Whether there was enough information about the intervention to replicate it (for example, there was a detailed description of what the programme actually involved and how it was run).
- Whether there was some sort of explanation or justification for why this specific intervention had been used.

The results are summarised below.



Discussion

Only six charities in this survey (19%) had evaluation materials that were sufficiently detailed to help a third party replicate some or all of the intervention. These charities had information on the cohort, detailed information on the intervention and data collection, and had justified their approach.

An additional two charities had some evaluation materials available, but these were only summary results - such as a 'success rate' for a programme without data on how this rate was measured.

The remaining 23 organisations (74%) lacked evaluation materials entirely, or had only anecdotal data (such as 'programme X helped me'). The information on these charities' websites could not be used to replicate the interventions.

The six charities which met all four of the criteria above were:

Programme	Method of evaluation	Sample size
1. Intensive support service during adoption process; live-in sentencing alternative for women under court order	Social return on investment calculations (attempt to quantify social and economic benefits of programme against financial cost)	20 families per year in the adoption programme; 205 clients per year in the live-in sentencing alternative
2. Detox facility pilot project	Semi-structured interviews at baseline and follow-up	17
3. 10-week parenting education course with weekly group sessions.	Multiple linear regression to analyse whether days in the programme predicted better outcomes on wellbeing metrics (uncontrolled)	374 parents with 469 children
4. Intensive case management of high-needs chronically homeless people	Semi-structured interviews, review of case files	13
5. Volunteer social support to reduce risk of sex offender recidivism	Review of case files, literature review, interviews with target population of sex offenders and stakeholders	70 interviews (30 with target population)
6. Specialist drug worker / GP collaboration to help patients withdraw from prescription drug addiction	Review of case files	267

Nearly all of these six organisations had very extensive evaluations, including in some cases independent evaluations and publications in peer-reviewed journals. This suggests that the organisations that wished to make public their evaluation materials committed substantial resources to the task and conducted these evaluations to a high standard.

This review did not consider the issue of publication bias in the results published on the websites.

Appendix 2: Example British Medical Journal trial report

RESEARCH

Effects of nutritional supplementation for HIV patients starting antiretroviral treatment: randomised controlled trial in Ethiopia

Mette F Olsen,¹ Alemseged Abdissa,² Pernille Kæstel,¹ Markos Tesfaye,³ Daniel Yilma,⁴ Tsinuel Girma,⁵ Jonathan CK Wells,⁶ Christian Ritz,¹ Christian Mølgaard,¹ Kim F Michaelsen,¹ Dilnesaw Zerfu,⁷ Søren Brage,⁸ Åse B Andersen,⁹ Henrik Friis¹

¹Department of Nutrition, Exercise and Sports, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

²Department of Laboratory Sciences and Pathology, Jimma University, Jimma, Ethiopia

³Department of Psychiatry, College of Public Health and Medical Sciences, Jimma University, Jimma, Ethiopia

⁴Department of Internal Medicine, Jimma University Specialized Hospital, Jimma, Ethiopia

⁵Department of Paediatric and Child Health, Jimma University Specialized Hospital, Jimma, Ethiopia

⁶Childhood Nutrition Research Centre, UCL Institute of Child Health, London, UK

⁷Ethiopian Health and Nutrition Research Institute, Addis Ababa, Ethiopia

⁸MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

⁹Department of Infectious Diseases, Odense University Hospital, Odense, Denmark

Correspondence to: M F Olsen mette.frahm.olsen@gmail.com

Cite this as: *BMJ* 2014;348:g3187
doi: 10.1136/bmj.g3187

This is a summary of a paper that was published on bmj.com as *BMJ* 2014;348:g3187

bmj.com

Read responses to this research at bit.ly/1o6qcVa

STUDY QUESTION

Can three months' intervention with a lipid based nutrient supplement containing either whey or soy protein improve regain of lean body mass, grip strength, physical activity, and immune recovery in patients with HIV starting antiretroviral treatment in a food insecure setting?

SUMMARY ANSWER

Supplementation resulted in greater gains of lean body mass, grip strength, and immune recovery in Ethiopian patients with HIV, compared with patients initiating antiretroviral treatment without a nutrient supplement. No major differences between the two supplements were observed

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Poor nutritional status at initiation of antiretroviral treatment is associated with impaired treatment outcomes among African patients with HIV. This study shows that provision of a lipid based nutrient supplement during the initial phase of antiretroviral treatment has beneficial effects.

Design

In this randomised controlled trial patients received 200 g/day lipid based nutrient supplement containing whey or soy during the first three months of antiretroviral treatment (ART). A control group received the supplement during the subsequent three months. Allocation to intervention groups was based on block randomisation and administered by a person not involved in recruitment or data collection. Supplement type (whey or soy) was masked with codes and blinded to everyone involved, while allocation to early or delayed supplementation was blinded to data assessors and data analysts only. We carried out three main comparisons. Whey and soy containing supplements were each compared with no supplement in participants with BMI >17, as a delayed supplementation groups served as controls. In addition, we compared whey and soy containing supplements with each other among all participants with BMI >16. Secondly, we compared early and delayed supplementation in participants with BMI >17 to investigate potential effects of timing of supplementation.

Effects (95% CI) of whey and soy containing supplements at three months in HIV patients with BMI >17 (n=282). Estimates are adjusted for sex, age, and education. Reference group was 93 patients who received delayed supplement

	Whey (n=94), P value	Soy (n=95), P value
Lean body mass (kg)	0.85 (0.16 to 1.53), 0.018	0.97 (0.29 to 1.64), 0.005
Grip strength (kg)	0.68 (-0.11 to 1.46), 0.090	0.93 (0.16 to 1.70), 0.019
PAEE (kJ/kg/day)	1.06 (0.87 to 1.29), 0.56	1.10 (0.91 to 1.33), 0.31
CD4 (cells/ μ L)	25 (-2 to 53), 0.073	15 (-12 to 42), 0.28
CD3 (cells/ μ L)	150 (24 to 275), 0.020	79 (-44 to 202), 0.21
CD8 (cells/ μ L)	112 (15 to 209), 0.023	60 (-35 to 154), 0.22

PAEE=physical activity energy expenditure.

Participants and setting

Adults with HIV eligible for ART with body mass index (BMI) >16 in Jimma, south west Ethiopia.

Primary outcomes

Lean body mass, grip strength, and physical activity at three months after initiation of antiretroviral treatment.

Main results and the role of chance

The effects of supplementation were considerable, resulting in a more than threefold weight gain, compared with the effects of ART alone, and with substantially more lean body mass gained. The increase of lean mass was accompanied by an effect on grip strength, though no effect on physical activity was observed. Furthermore, the whey containing supplement was associated with increases in CD3 and CD8 counts, and results also suggested an effect on CD4 counts. No such effects on immune recovery were shown for the soy containing supplement, but when the two supplements were compared, there were no significant differences in their effects. Patients receiving delayed supplementation had a greater weight gain but less gain in grip strength and physical activity than those receiving early supplementation.

Harms

No harms of nutritional supplementation were observed.

Bias, confounding, and other reasons for caution

The risk of selection bias was low as patient recruitment was consecutive, group allocation was concealed, and the study had a high follow-up rate (88% at three months). For many patients, however, we had incomplete data on physical activity, and we might not have been able to detect a potential effect of supplementation on this outcome.

Generalisability to other populations

Our findings are relevant for the treatment of all patients with HIV in food insecure settings. We included only patients with BMI >17 in the comparison with an unsupplemented group during the first three months of antiretroviral treatment, but previous observational studies have shown associations between weight gain and improved treatment outcomes across all BMI strata. We therefore conclude that the beneficial effects of supplementation can be generalised to HIV patients with inadequate access to food regardless of initial BMI status.

Study funding

The study was funded by US Dairy Export Council, International Atomic Energy Agency (IAEA), and Ministry of Foreign Affairs of Denmark (DANIDA). Nutriset developed the supplements and partially covered transportation expenses.

Appendix 3: Bad research flatters more than good research

Why we evaluate research methods:

i) Because different research methods give different answers

“Two men say they’re Jesus: One of them must be wrong” (Dire Straits lyric!)

The table below shows the effect of a reading programme in India measured using several research methods¹⁴. These methods all used the same outcome measures, but different experimental designs.

The answers vary widely: some suggest that it works well, others show it to be detrimental. Clearly there is only one correct answer! All the other answers are incorrect: and could mislead donors or practitioners to implement this programme at the expense of another which is better.

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

The answers vary because research methods vary in how open they are to biases (i.e., systematic errors). For instance, if a study only looks at the outcome (here: reading level) before the programme and then afterwards (i.e., is a pre-post study), we don’t know whether any improvement in reading levels was due to the programme or just to the fact that children learn over time anyway.

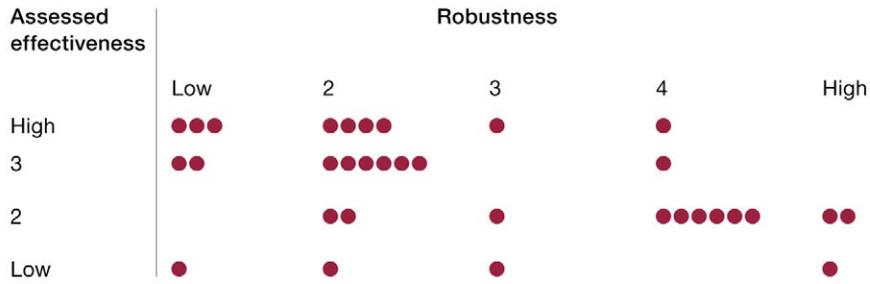
{As an aside, contrary to popular myth, it is not invariably the case that robust research is more expensive than unreliable research, nor that randomised controlled trials (the most reliable design for a single primary study) are invariably terribly expensive: many are cheap or free.}

ii) Because weaker research methods allow for more positive findings

The UK National Audit Office searched for literally every published [analysis of the quality of almost 6,000 government evaluations](#)¹⁵. Of those, it chose a sample, and ranked on one hand, the quality of the research method (‘robustness’ on the x axis, i.e., how insulated the study is from bias), and on the other, the positive-ness of the programme (‘claimed impact’).

The trend line on the resulting graph below would slope diagonally downwards. It shows that more robust research only allows for modest impact claims whereas weak research allows much stronger claims.

Relationship between robustness and claimed impacts in evaluations



Note

- 1 Robustness assessed on Maryland Scale. Assessed effectiveness, rated low to high.
- Low = Small or insignificant effects.
- 2 = Mixed effects, positive for some, negative or insignificant for others.
- 3 = Positive effects, with some caveats or uncertainties noted.
- High = Significant positive impacts, no or only minor caveats or uncertainties noted.

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 also in medical research. We found it in the studies of outdoor learning too.

Endnotes

Many comments were made to us in confidence. Therefore, not all of the statements in this document are referenced fully, though all supported by evidence and our interviews.

- 1 Lane, R. (2014) Paul Glasziou: surfing the wave of evidence-based medicine. *The Lancet*. Vol 383 (January 2014) p.209. [Online] <http://download.thelancet.com/pdfs/journals/lancet/PIIS0140673613623867.pdf> [Accessed: 24.09.14].
- 4 Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomized trials. [Online] p. 22. [http://www.consort-statement.org/Media/Default/Downloads/CONSORT%202010%20Statement/CONSORT%202010%20Explanation%20and%20Elaboration%20\(BMJ\).pdf](http://www.consort-statement.org/Media/Default/Downloads/CONSORT%202010%20Statement/CONSORT%202010%20Explanation%20and%20Elaboration%20(BMJ).pdf) [Accessed: 02.10.14].
- 5 National Audit Office. *Evaluation in Government*. [Online] <http://www.nao.org.uk/report/evaluation-government/> [Accessed: 10.10.14].
- 6 Haynes, R.B., Mulrow, C.D., Huth, E.J., Altman, D.G., Gardner, M.J. (1990) More informative abstracts revisited. *Annals of Internal Medicine*. 113 (1).
- 7 Taddio, A., Pain, T., Fassos, F.F., Boon, H., Ilersich, A.L., Einarson, T.R. (1994) Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *Canadian Medical Association Journal*. 150 (10).
- 8 Wager, E., Middleton, P. (2008) Technical editing of research reports in biomedical journals. *The Cochrane Library*.
- 9 Hartley, J., Sydes, M., Blurton, A. (1996) Obtaining information accurately and quickly: Are structured abstracts more efficient? *Journal of Information Science*. 22 (5).
- 10 Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. [Online] [http://www.consort-statement.org/Media/Default/Downloads/CONSORT%202010%20Statement/CONSORT%202010%20Explanation%20and%20Elaboration%20\(BMJ\).pdf](http://www.consort-statement.org/Media/Default/Downloads/CONSORT%202010%20Statement/CONSORT%202010%20Explanation%20and%20Elaboration%20(BMJ).pdf) [Accessed: 02.10.14].
- 11 St Giles Trust. (2012) Support for female prison leavers reduces re-offending. [Online] <http://www.stgilestrust.org.uk/stats-and-info/p192682-support-for-female-prison-leavers-reduces-re-offending.html> [Accessed: 10.10.14].
- 12 Harris, A.H., Standard, S., Brunning, J.L., Casey, S.L., Goldberg, J.H., Oliver, L., Ito, K., Marshall, J.M. (2002) The Accuracy of Abstracts in Psychology Journals. *The Journal of Psychology*. 136 (2).
- 13 Markets For Good. (2014) *Finding More Value In Charities' Research*. [Online] <http://www.marketsforgood.org/finding-more-value-in-charities-research/> [Accessed: 24.09.14].
- 14 Innovations for Poverty Action
- 15 National Audit Office. *Evaluation in Government*. [Online] <http://www.nao.org.uk/report/evaluation-government/> [Accessed: 10.10.14].