

Charities' research infrastructure:

Improving Clarity and
Usefulness of Research by
Charities in Criminal Justice

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

“Good research reporting in eight words: ‘what did you do; what did you find?’”

– Doug Altman, Professor and Director of The Centre for Statistics in Medicine at Oxford University²

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

– Isaac Newton

“If you want to encourage some activity, make it easy”.

– Richard Thaler, Professor of Economics at the University of Chicago, and co-author of Nudge³

About Giving Evidence

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

Through consultancy, Giving Evidence helps donors and charities to understand their impact and to increase it. Through campaigning and thought-leadership, we show what evidence is available and what remains is needed, where quality and infrastructure need improving: in order that decisions about *giving* can be based on sound *evidence*.

Giving Evidence was founded by **Caroline Fiennes**, a former award-winning charity CEO, and author of *It Ain't What You Give*. Caroline speaks and writes extensively about these issues, e.g., at the Skoll World Forum, the US Center for Effective Philanthropy, in the Stanford Social Innovation Review, Freakonomics, and the Daily Mail. She is on boards of The Cochrane Collaboration, Charity Navigator (the world's largest charity ratings agency) and the US-based Center for Global Development.

Contents

1. Introduction and purpose	4
2. The current evidence system in criminal justice charities	6
2.1 What research is produced and reported by criminal justice charities?	7
2.2 Dissemination	16
2.3 Use: who uses charities' research?	17
3. Beyond the scope of this project	20
4 The proposal and reactions	21
4.1 Making research clearer: The checklist	21
4.2 Making research easier to find: The repository	26
4.3 Concerns, key success factors, scope	29
5. Next steps: the pilot	33
Appendix 1: Research method and people interviewed and consulted	36
Appendix 2: Fewer than 20% of charities in crime / criminal justice publish adequate evaluation materials	38
Appendix 3: British Medical Journal 'trial reports': two examples	40
Appendix 4: CONSORT Checklist and flow chart for reporting medical RCTs	42
Appendix 5: Instructive models for consolidating research	46
Appendix 6: Bad research flatters more than good research	49
Appendix 7: Incentives on charities and researchers in publishing and indexing research	50
Appendix 8: Relevant academic centres	51

With thanks to J Leon & Company Ltd

1 Introduction and purpose

Our purpose is to enable best practice amongst charities working in crime reduction and criminal justice to spread and to spread more quickly. It seems unacceptable that, quite possibly, somebody has discovered an intervention which works well but which doesn't spread, or that organisations are using interventions known to not work but which don't die. It seems equally unacceptable that charities[†] spend money on research which suffers from some or all of the problems which Paul Glasziou identifies above in medical research: money which could more usefully be spent on enabling better research or better delivery.

This project concerns the last problem on Glasziou's list: reporting of research. That is not because we think that it is the sole problem – we don't. We don't claim that fixing research reporting is sufficient, but we do think it necessary. [Ginny Barbour, editor of medical journal PLoS, likens improving research reporting 'to turning on a light in a room. It tells you what the room looks like; it doesn't clean the room for you.'] We also suspect that fixing reporting may be simpler than fixing the other three problems.

By 'research', we mean research by charities which could be useful to external audiences. That is broadly of two types. First, research into the effectiveness of their interventions. And second, research to inform policy debates, e.g., about the effectiveness of prison, or shorter sentences, or public attitudes. We are not so concerned here with research for internal organisational purposes such as process evaluations.

Research of both types is somewhat analogous to medical research, since both aim to identify what works – or, rather, what works best, and/or what is most cost effective. Charities in crime reduction and criminal justice run programmes, (e.g., to reduce offending, to increase employment after release, or to reduce re-offending) and evaluate the effect of the work. That is, they investigate causal links between inputs and outcomes. The findings of both types of research can be useful to many practitioners in their own organisations, precisely like medical research. Hence ideally charities' research should be used by academics, the police, commissioners and others.

For charities' research to be useful, it should be treated like any other serious research: high quality, published somewhere findable, clear, disseminated effectively to its intended audience, and used.

This Giving Evidence study began with two observations about much research by charities. First, that it could be easier to find. There is no central repository and hence material is often published just on charities' own websites so readers may not find it. Second, that reports about research would be more useful if they were clearer – about the intervention used, the research done, and the findings.

The study was simply a consultation 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 standardised checklist of items to be included in the research report. And/or:
- (b) Sit in a central repository.**

[†] We use the term 'charity' to mean any voluntary or community organisation including social enterprises. We use 'charity' and NGO interchangeably.

The conclusion thus far is that both these developments would be useful, at least to many bodies. We also found considerable support for these proposals in other parts of the charitable sector (we were not deliberately consulting there, though we constantly have many conversations with people in many parts of the sector), implying that perhaps the system of a checklist and repository could usefully be replicated beyond the crime reduction and criminal justice arena.

This consultation also identified some other options and suggestions (see Section 4). The probable next step is to pilot both suggestions with a small group of charities (see Section 5).

This document captures: the current situation; how the issues of findability and clarity of research are dealt with elsewhere, e.g., in agriculture and medical research; the response to the two proposals above; and suggestions for the pilot and implementation. In the spirit of our consultation, we welcome your perspectives: please send them to admin@giving-evidence.com.

This project is (the beginnings of) a behaviour change exercise. It aims for charities to produce and publish clearer research reports; and for funders, commissioners, the police, policy-makers and other charities to use them to make informed decisions about which interventions to run and to fund.

One aim is for it to be unacceptable for a charity to say 'please find attached our research': rather the norm should be to say 'please find below a link to our research'.

"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

2 The current evidence system in criminal justice charities

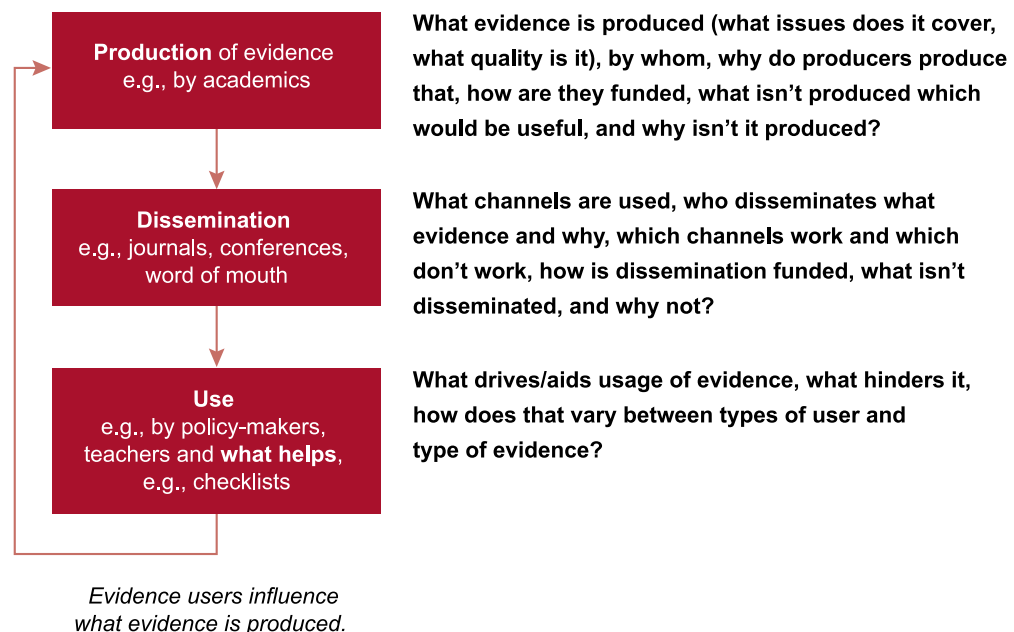
In encouraging practitioners to use more evidence, the concept of an 'evidence system' is helpful. An evidence system comprises:

- what evidence[†] is **produced** - and how, by whom, why, and how it's funded. Also what evidence isn't produced and why not;
- how it is **disseminated** and stored; and
- how it is **used**, and what aids and hinders usage.⁴

This project focuses on part of production and part of dissemination, though other aspects of the evidence system in criminal justice probably need strengthening too.

The objective of an evidence system is to enable good decisions. It should enable practitioners, policy-makers, funders and others to (a) find any existing research, such as about the effectiveness of interventions for ameliorating particular situations, (b) compare those interventions, in terms of effectiveness, cost and cost-effectiveness: that in turn relies on being able to assess the quality of the research, (c) ascertain or estimate whether the intervention is likely to produce those same outcomes in the decision-maker's own context, and (d) give the reader enough detail that they could replicate the intervention. A good evidence system also reduces waste (e.g., research into questions to which the answers are already known) simply by making the existing research findable and clear.

Figure 1 - An evidence system



[†] This document uses the terms 'research' and 'evidence' interchangeably.

2.1 What research is produced and reported by criminal justice charities?

Many charities publish nothing about their effectiveness. Giving Evidence analysed a small sample of UK charities working in criminal justice, and found that 74% publish no research at all about their effectiveness. (Details in Appendix 2.) That doesn't necessarily mean that no research is happening: perhaps research is conducted but withheld. A study of charities in Canada, as yet unpublished, found that perhaps as much as 90% of research by charities is unpublished.

The importance of clear reporting

Maximum reduction in re-offending 'if you do everything right' according to academics⁵: about 20%.

Overstatement of an effect which can arise from using bad method of randomisation: over 40%.⁶

For research which is published by charities, there are (as far as we could find) **no standards or templates for what is reported**. Charities appear to publish as much or as little information as they see fit, in whatever sequence they want, and publish it wherever they want or can. We heard this from charities and others interviewed, and have observed it in many other parts of the voluntary sector over many years. Therefore often much information is lacking from charities' research reports.

"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

For example, the British Medical Journal (BMJ) publishes single-page 'trial reports' in every issue which have a standard template covering basic information (two example trial reports are in Appendix 3). To see whether charities in criminal justice publish this basic information, Giving Evidence found the best evaluations published by criminal justice charities (method in Appendix 2) and, of those, selected two at random, and tried to complete the BMJ template for them. [Somewhat atypically, these two were both published in academic journals.] As you see in Figure 2 below, from even the best, some items were missing.

Figure 2 - Comparison of adoption support charity evaluation vs BMJ guidelines

BMJ trial report item	Present in study 1? (Adoption service)	Present in study 2? (Social support for sex offenders after jail)
Authors and their organisations	✓	✓
Study question	✗	✓
Summary answer	✓	✓
What is known and what this paper adds	✗	✓
Participants and setting	✓	✓
Design	✓	✓
Primary outcomes	✓	✓
Main results & role of chance / bias, confounding and other reasons for caution	✓	✓
Funding	✓	✓
Potential conflicts of interest	✗	✗

Medical research probably leads the way in terms of the sophistication with which practitioners are trained about research, the reporting of research and understanding of how to minimise waste in research. Medics have developed various standards for research reporting⁷. The standard for medical randomised controlled trials is CONSORT⁸ (Consolidated Standards of Reporting Trials), endorsed by over 600 influential journals. (See Box 1.) It contains two sections. First, a checklist, of 25 items (37 when sub-categories are included), which include those recommended here (e.g., describing the intervention) and others (e.g., how sample size was decided, method of randomisation, and how a trial changed once underway). Second, a flow-chart showing how many beneficiaries dropped out at various stages and the reasons for that. Both could usefully be used for reporting research in criminal justice, and both are shown in Appendix 4. Medicine has other checklists for reporting other types of research, over 200 in total, such as STROBE for observational studies (STrengthening the Reporting of OBservational studies in Epidemiology)⁹ and PRISMA for systematic reviews and meta-analyses (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).¹⁰ A version of CONSORT is being developed at Oxford University to cover RCTs in social and psychological interventions¹¹ (which will include crime reduction, education, public health, social work etc.)

Box 1 - CONSORT (Consolidated Standards of Reporting Trials)

CONSORT¹² aims to improve the reporting of medical clinical trials. Its main goal is the *reporting* of trials: to make clear the method used in the trial, in order that the journals and their readers can assess the quality and reliability of the answer. However a secondary goal is improving the *design* of trials: if bad design becomes easier to spot, it becomes harder to publish badly-designed trials, which reduces the incentive to do them. Growing evidence suggests that reporting guidelines do have this effect.¹³

Clarity about a trial's method is important because it can significantly affect the trial's results. As mentioned, the method for generating random assignment can influence the answer by 40%; trials that were not blinded (e.g., participants knew whether they were getting the 'treatment' or a placebo) typically overstate the treatment's effect by 25%; and trials with poor methods for enforcing the randomisation (which give the trial administrator some influence on who goes into which group, so the groups may not actually be random) typically overstate the treatment's effect by 31%. Those overstatements obviously can lead practitioners to bad decisions – which lead to bad outcomes for patients and wasted money.

The first CONSORT 'statement' was published in 1996, in response to growing evidence that method affected results. It was revised in 2001, and again in 2010. The checklist includes only those items deemed 'absolutely fundamental to reporting a randomised controlled trial (RCT)', and for each checklist item, CONSORT provides evidence from studies of trials (i.e., meta-research, that is research about research) that can influence the result.

It suggests what a trial report should contain: it does not include any judgments on how a particular study should have been run. For example, CONSORT suggests that authors state how trial participants were randomised. Methods of randomisation vary in how good they are though CONSORT itself doesn't give the reliability of any particular method. A knowledgeable or resourceful reader of a CONSORT-compliant research report can make that judgement.

CONSORT has the explicit endorsement of over 600 journals in many countries and languages, the International Committee of Medical Journal Editors, the Council of Science Editors, and the World Association of Medical Editors¹⁴.

Quality of research

In medicine and some other sciences, there are often assessments of the quality of published research, and those assessments are used to identify problems such as resources being wasted on research too unreliable to be used or not geared towards adding value to the current body of knowledge, and to rectify them. By contrast, the quality of research by charities in crime reduction and criminal justice appears not to be routinely assessed: indeed we could not find any such assessments.

However, research quality does seem to be a problem. For example, the Arts Alliance is a coalition of arts organisations working in the UK criminal justice system, managed by Clinks, the umbrella body of voluntary sector working with offenders in England and Wales. Its 'evidence library' (described further in Appendix 5) contains evaluations of arts-based practice. In 2013, it had 86 evaluations, only four of which met the quality criteria for inclusion in a 'rapid evidence assessment' commissioned by the National Offender Management Service (NOMS)¹⁵. The poor quality of research by charities

in general is indicated by a report from The Paul Hamlyn Foundation¹⁶, which analysed the quality of research received by the foundation from grantees. Using a rather generous scale of quality, the foundation found that only a third of research it had received was 'good', and found "some, though relatively few, instances of outcomes being reported *with little or no evidence to back this up*" (italics ours).

"My first thought is that the reports which satisfy funders shouldn't satisfy them." – academic expert in crime

The problem has several causes, most notably that charities have scant incentive to produce high-quality research nor do many have the skills (discussed in detail later). Funding is certainly a factor: many funders ask that a proportion of every grant is used for evaluation. Those amounts are individually too small for reliable research. It's a tragedy because it produces many low-grade studies, though the total budget might have enabled a few good studies.

"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. That set-up forces people to do bitty research just to prove their point." – operating charity

The most reliable form of research is systematic reviews, which synthesise all findable research on a particular topic above a quality threshold. In medical research, that threshold normally means that only RCTs are included[†]. Ideally, charities' research should be strong enough to be included in systematic reviews, and hence drive forward the frontier of knowledge. The UK's new What Works Centre on Crime Reduction (see Box 2) has identified 337 systematic reviews in crime reduction. Though the Centre hasn't rigorously coded those systematic reviews for the source of the studies included, it said that the amount of charity-generated research included in them was 'very little'¹⁷.

This poor showing for charities' research may be because charities' research is too weak to include, or simply because it couldn't be found. Systematic reviewers – normally academics – vary in their diligence in searching for material to include, normally responding to constraints of time or budget¹⁸. It's easy for them to find material published in journals which is hence almost invariably considered, whereas material which is published but not anywhere obvious ('grey literature') may not be included. Charities' research may be being under-used in systematic reviews, which is a wasted opportunity.

[†] Reviews by The Cochrane Collaboration, perhaps the most respected producer of medical systematic reviews, *only* consider randomised controlled trials.

Box 2 - The What Works Centre on Crime Reduction

In 2013, the UK government launched several 'What Works Centres', each tasked with identifying good practice in a particular area and improving policy and outcomes there. Based on the National Institute of Health and Care Excellence (NICE), which determines what medicines and procedures should be available on the NHS, the what works centres arose from wanting, as Jeremy Heywood said in his first speech as Cabinet Secretary, 'a NICE for social policy'¹⁹.

The What Works Centre on Crime Reduction is led by the College of Policing, supported for three years by a consortium of universities[†]. Its initial tasks include listing all systematic reviews relevant to crime reduction, rating and ranking them so as to identify effective interventions, producing guidance for practitioners (the police and others), and training practitioners²⁰. The Centre may conduct or commission primary research, though clearly only where there are gaps in the existing evidence. It is more likely to produce tools and guidance, inspired by the Education Endowment Foundation's [Toolkit](#), which compares the effectiveness and cost of various interventions by synthesising the available evidence about them (see below.)

Whereas systematic reviews often largely look at *whether* something works, the What Works Centre on Crime Reduction is also focused on understanding *why* it worked (or didn't), what enables and hinders it, the contexts in which it might succeed, and what implementers need to know in order to implement it.

The College of Policing is well-aware that crime reduction covers much more than policing.

The What Works Centre should be a fantastic – and free – resource for charities and funders interested in reducing crime. It should also use charities' research: that is, charities' research should be of adequate quality and relevance to be included in its studies and thereby shared with a much larger audience.

[†] University College London, the Institute of Education at the University of London, the London School of Hygiene and Tropical Medicine, Birkbeck College, and Cardiff, Dundee, Surrey and Southampton universities.

Extract from the Education Endowment Foundation's toolkit:

The screenshot shows the Education Endowment Foundation's Toolkit website. The header includes navigation links like 'About Us', 'Print', 'Share', 'Document Library', 'Contact Us', and 'Log In'. A search bar contains the text 'e.g. Teaching and Learning Toolkit'. Below the header, there are tabs for 'Toolkit info', '+ Pupil Premium Calculator', 'Latest Updates', and 'Downloads'. The main content area displays a list of interventions with columns for cost (represented by pound signs), duration (in months), and a 'Downloads' button. A 'Toolkit Filter' sidebar on the left allows users to sort by 'A-Z', adjust 'Average Impact' with a slider, and filter by 'Cost' (represented by pound signs).

Intervention	Cost	Duration	Downloads
After school programmes	£££££	+2 months	Download
Arts participation	£££££	+2 months	Download
Aspiration interventions	£££££	0 months	Download
Behaviour interventions	£££££	+4 months	Download
Block scheduling	£££££	0 months	Download
Collaborative learning	£££££	+5 months	Download
Digital technology	£££££	+4 months	Download
Early years intervention	£££££	+6 months	Download

Quality control

There is no quality control process for most research by charities. There is nothing analogous to peer review, used in academic research to check that the research has been carried out correctly and that the results have been interpreted appropriately²¹: it also raises questions about results which are unlikely (though there are well-documented problems with peer-review^{22, 23}).

“When I first started in this, 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’s worth understanding how peer review works. There are two ways. In the traditional model, it serves an editorial function *before* publication. The journal, a commercial publisher, wants a reputation for only publishing material of a particular quality. To assess the quality of each piece of submitted research, it recruits reviewers, normally academics, who are experts in the topic and hence peers of the submitting researcher. [Importantly, the incentives of the reviewing academics are not clear and sometimes misalign with those of the reader. For example, it’s said that reviewers waive through articles by their pals and block articles by their rivals. This may result from specialisation in academia, which means that there are often few people qualified to comment on a particular topic, so despite the reviewers being officially anonymous, it’s obvious to the producers who they are, so blocking poor research can be embarrassing. Surprisingly, reviewers are seldom paid by the journal, so arguably they have no incentive to do a good job.] In the newer model, review largely happens *after* publication. For example, PLoS (the Public Library of Science), a non-profit journal, has a quality standard for everything it publishes, which is lower than for traditional journals and hence PLoS publishes a higher proportion of the material it receives. However, anybody can review and comment on its articles. Any problems with the research become clear from the comments, and the research’s usefulness becomes clear from the amount that it’s cited. That is, the quality control beyond a minimum standard is crowd-sourced.

Charities’ research has neither safeguard. It is published by the producer, without having to meet anybody else’s editorial or quality standards. And neither is there an audience which actively critiques the research once it is published, which discerns its quality.

Perhaps consequently, charities could, if they so wished, fabricate not only the research but also the underlying data. The check on this in academia is that people replicate each other's studies, but that requires the intervention and study design to be described in detail, which isn't the case in most charities' research.

Reporting about intervention details

We heard complaints that charities often inadequately describe their interventions[†] or, for policy research, their research methods. This prevents their research being included in systematic reviews, and precludes anybody else from replicating their programme. This latter may be in part intentional – or, at best, a response to a badly-designed incentive: when charities compete for contracts from commissioners, their interventions are their products, their IP. This provides a strong financial incentive to protect their 'secret sauce'. This is a grave problem on two counts. First, many successful programmes serve a need larger than their inventor-organisations can serve and hence replication by other organisations is necessary for reducing crime and creating justice as fast as possible. And second, replication is essential for testing whether the research was accurate^{††}.

One academic / practitioner who we interviewed pin-pointed a problem which would be solved by greater detail and transparency about research, including pre-registering research to avoid publication bias: *"It's all rather suspicious, frankly. NGOs rarely seem to publish evidence that doesn't support their existing position. The more I've researched this, the more outrageous it seems. Get this: The economists' blind trials show that prison works... but the left-wing charities consistently say that it doesn't. Those on the right produce evidence that's on the right; those on the left, they produce evidence that's left. I bet they're just choosing what suits them."*

We also heard complaints that reports about charities' work inadequately describe the quality of implementation:

"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. If the organisation has a process review [which would normally detail this], that's really useful to see alongside the 'scientific' evaluation" – 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

[†] Interestingly, one person told us that simply being a charity is an essential part of the intervention, e.g., that women prisoners' answers on whether they have children are quite different depending on whether they are asked by somebody employed by the state or somebody perceived as independent, because they may fear the state removing their children.

^{††} A famous example of this correcting-mechanism comes from economics, where two leading American academics were caught out by a student replicating their analysis, which had to be corrected, rather embarrassingly. Harvard University professors Carmen Reinhart and former International Monetary Fund chief economist Ken Rogoff wrote a research paper called *Growth in a Time of Debt* showing that economic growth slows dramatically when a country's debt rises above 90% of GDP. A student elsewhere couldn't replicate their results and requested their calculations. He discovered that the academics had accidentally only included 15 of the 20 countries under analysis in their key calculation, and used a rather debateable way of averaging figures from the others.

Box 3 - The Justice Data Lab

Charities working with offenders often struggle to gather data from which to assess whether and which parts of their services are working because they cannot track their beneficiaries over time (sometimes for legal reasons) and/or they cannot see data on whether and how their beneficiaries re-offend. The state alone has those data. The Justice Data Lab, currently being piloted by the Ministry of Justice (MoJ), aims to solve this. Charities submit data about offenders with whom they have worked and the services they have provided, and the lab returns to them the one-year re-offending rate for that group and that of a matched control group of similar offenders. (It currently cannot provide data other than the one-year re-offending rate.)

In terms of the evidence system, the Justice Data Lab is a producer of evidence. In terms of the quality of the research, the method is published by the Ministry of Justice²⁴, though clearly sample sizes vary and hence the statistical significance of results varies. On dissemination, the MoJ publishes all results²⁵ and requires that all results be published on the charities' own websites within four months²⁶.

In terms of being *clear* (which the *checklist* aims to improve), the MoJ's report summarises the intervention too briefly for anybody to replicate it or make reliable comparisons[†].

In terms of being *easy to find*, the results from the Justice Data Lab are relatively easy to find if you know where to look. However, the MoJ publishes them in a table in a PDF document, and issues a new such document every month with that month's results. The results may therefore not be picked up by search, and it is rather laborious to find all results from interventions of a particular time in order to make comparisons. That problem will obviously worsen over time.

The Justice Data Lab appears a tremendous idea and its first year has been well-received. It's also cheap: apparently costing about £170,000 per year, and producing about 60 reports per year²⁷: at less than £3000 per report, it is considerably less expensive than most evaluations. The research it produces – like much other research in the sector – will be clearer and easier to find if the approach explored in this consultation become enacted.

Who produces the research published by charities?

Research published by charities is produced by three main types of researcher, outlined below. A detailed breakdown of their prevalence is beyond what was necessary for this study (we did not find a quantified breakdown of these categories – though we didn't look terribly hard for one) but it's worth outlining the main issues with each.

† For example, the summary of an intervention by The Footprints Project (chosen at random for this illustration) is simply: "a mentoring service to individuals leaving custody or serving a community sentence ... a "through-the-gate" mentoring service. Trained volunteers from the community act as mentors by guiding and supporting individuals with various needs, often signposting them to where they can further access particular support that they need including accommodation, finance, health services, substance misuse agencies and access to training/voluntary work." Of the cohort of people, the summary says only that: "Many of the individuals ... have mental, physical, social, and educational/employment issues, alongside difficulties with relationships, substance misuse and housing." Clearly those 'issues' could be of many grades of severity.

1. The charity's own staff. We suspect that this is the vast majority. There are two major problems with this category²⁸. First, **most charities are operators, they're not research-houses**. They aren't organised around the research function – which perhaps explains why the charity sector has no peer review process or boards which grant ethical approval for experiments on human subjects as academic researchers do. Nor should most charities be researchers: their primary role is supporting victims of domestic violence or helping ex-offenders to find jobs. This work is obviously important, and equally obvious is that *doing* things differs fundamentally from *studying* things. Hence **charities rarely have strong research skills**. Less than a third of charities in criminal justice have anybody for whom 'impact measurement' is part of their role²⁹. Furthermore, of the estimated 1475 charities in the UK working with offenders, ex-offenders and their families, about 90% are small community groups, with budgets of less than £150k.³⁰,³¹. It would be bizarre and wasteful for so many organisations, and such small organisations, to each have sophisticated research skills. Focused on delivery, rather than research, **charities don't have incentives to conduct quality research and publish** in prestigious channels as researchers and research institutions do.

"We don't really look at research by charities. If we want quality research, we go to the research producers. Universities and so on." – senior manager in large funder

The exception is where charities are conducting research for policy purposes. Here we found considerable skepticism about the validity and integrity of their research. This might be improved by greater transparency about their research methods: it's easy to support a case by cherry-picking the data to publish, and/or withholding research whose message is unwelcome.

"It's remarkable that charities with a particular position: their research always confirms that position, even if independent, academic research consistently shows something quite different. It's not remarkable, it's incredible... literally, not credible. It doesn't help anybody to believe their case" - senior academic and practitioner

Second, **charities' incentives around research are horrible**. Often, their research into their effectiveness is motivated, or at least, influenced by the need to win funding. (See Appendix 5.) Therefore, without besmirching anybody, it's clear that they have an incentive to produce research which flatters them: to use the methods most favourable to them and to publish only the subset which helps them[†]. They're not being evil: they're simply responding to badly-designed incentives. [Pharmaceutical companies face similar incentives – they want their drugs to get licensed – and hence their research is demonstrably horribly skewed, about four times as likely to show a favourable result than independent research into the same question³², ³³. There are safeguards, such as requiring that all research is registered before it starts which makes it easier to spot research which mysteriously doesn't get published, not that they yet work very well³⁴.]

"It's often ridiculously complementary! ... a flood of positive results" – major funder

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

"Academics will know that if you do everything right with [that group], you can reduce re-offending by maybe 20%. They've known that for years. Then along comes [charity] claiming 40%. Well, tough act if it's true" – analyst

[†] Giving Evidence is starting a study to find if, where, and how much the literature published by charities is biased. To our knowledge, this has never previously been studied.

2. 'Independent' researchers, consultants and academics, funded by the charity. This is probably the next most frequent category. As an aside, it's worth noting that medical clinical trial reports contain details of how the research was funded (see examples in Appendix 3): charities rarely report this.
3. Academics or other researchers funded independently produce a little.

Presumably, some of the research by academics in categories (2) and (3) here appears in journals. That is probably peer reviewed, though the extent of this is not clear.

2.2 Dissemination

There is no standard place for charities to publish their material – unlike, say, medicine or academia which uses journals and databases of journal abstracts. Neither are there standard ways for charities to 'tag' their research online (i.e., meta-data) to ensure that it is findable through search. That is, the information infrastructure (also called information architecture) is weak: good information architecture allows the user to find all and only what they want relatively swiftly.

"To be honest, I don't even know where our own research goes. Where would I go to find it? Er, Google! [laughs]" – researcher in charity in criminal justice

If I were looking for research ...*"I'd have to do a big trawl... government sites, NGOs, academics. Would probably take ages"* – researcher in charity in criminal justice

"There certainly is no clear go-to place"- Harvey Koh

"When we set up the [organisation], we really needed good evidence to make our case but it was really hard to find" (You mean that it didn't exist, or that it existed but it was hard to find?) "Hard to find. It was there... hiding. But you couldn't use much: they all measure different things, apples and pears. Some don't even tell you what they're measuring, so maybe it's not even apples and pears. Maybe bananas!" – operating organisation

There are some databases of materials by charities (e.g., the Third Sector Knowledge Portal built by the Third Sector Research Centre at the University of Birmingham) though they are far from comprehensive. There are various journals about crime and criminology (e.g., the Journal of Experimental Criminology) though these generally take articles about research more rigorous than charities produce. Furthermore, unlike academics, charities have scant incentive to publish in journals since their key audiences (e.g., funders, commissioners) rarely read them and it's not important for professional advancement.

For some other parts of the charitable landscape, research is accessible via rather unsystematic lists, e.g., the European Venture Philanthropy Association has a 'knowledge centre' with links to reports which its members (funders using venture philanthropy models) may find useful³⁵.

"There currently is essentially no system. This would be a great improvement." – senior funder

"You have to look on every site" - Carol Jackson, Head of Outreach, Assessment and Outcomes, The Prince's Trust

Charities said that they mainly disseminate material themselves through websites, and newsletters, and/or through conferences and umbrella organisations.

Box 4 - Why do charities produce research?

When asked why they measure the impact of their work, seven times as many UK charities cited funders' requirements as cited 'wanting to improve our services' (52% against 7%)³⁶. This suggests that charities' evaluations are often essentially compliance or designed to help fundraising, rather than as a serious attempt to learn.

This is supported by a rigorous study by two American universities³⁷: in a randomised controlled trial with 1,419 micro-finance institutions (MFIs), they found that the MFIs were 'significantly' more likely to want to study their own effectiveness if they had previously been shown evidence which flatters MFIs than if they had been shown evidence critical to MFIs.

Funders and commissioners (arguably) amplify this incentive by requiring charities to include assessments of their effectiveness in applications and tenders for commissioned services.

Several commentators reported that charities' research is largely 'post-match analysis': "it's all done too late... after the action, not before. It's not designed to inform operational decisions", said one.

A study of charities in Ireland³⁸ found that: "There is a deep-rooted fear of finding out (or 'being found out') that one has not had the impact that was intended... Organisations are incredibly reluctant to admit that programmes have not gone to plan... Some simply do not tell funders the truth; others are very opaque when reporting back to funders; yet others cherry-pick clients to ensure low success rates are minimised... Lessons of 'failure' are rarely shared... When funders become aware that the desired results have not been achieved for whatever reason, they are seemingly equally reluctant to take constructive action, for fear of damaging the organisations' (and possibly their own) reputations."

That is, the system encourages charities to research their own effect essentially as an advert. This has a predictable effect on research quality. Rigorous research is typically less flattering than sloppy research (as illustrated by a study by the UK's National Audit Office³⁹ of all UK government evaluations, in Appendix 6). Therefore charities have a conflict of interest between honesty and the need to raise money. It's therefore hardly surprising that much research is withheld. ('Publication bias' which results from withholding of unflattering results is rife in pharmaceutical research⁴⁰, where companies have commercial incentives.) To our knowledge, withholding and publication bias in charities' research have never been studied.) We heard that the NOMS has lost interest in qualitative research, since it is often simply case studies with little or no discussion of how that case study came to be chosen, and hence wide open to bias.

2.3 Use: who uses charities' research?

Again, we did not need to explore this in detail, but did encounter some curious data. The short answer is that charities' research seems not to be used very much. When most people want reliable answers, they turn to the academic literature. This chimes with our experience in many other parts of the charity sector.

We did not find that charities themselves routinely look for nor use research by each other. The public sector entities we interviewed had a few examples of using charities' research though were

far from convinced that it is a valuable resource. As discussed, relatively little is eligible for inclusion in systematic reviews.

One interviewee talked about practitioner colleagues using charity research “blindly. Like they use everything else [meaning, other research], because they don’t know how to critique it. So it lands on their desk and they just believe it. Could be a systematic review, could be just one random case study.” Hence some ‘facts’ which ‘everybody knows’ turn out not to be true at all. This speaks to the need for all users of research to understand enough about research methods to discern the value.

Box 5 - Reducing crime versus policing

Work to reduce crime normally focuses on *criminals*: identifying who they are, or who is likely to become a criminal – based on the assumption that criminals are a distinct group, different to everybody else[†] – and intervening to prevent them committing crime.

Nick Ross, former presenter of BBC’s *Crimewatch*, attacks this approach in his book *Crime: How To Solve It, and Why So Much of What We’re Told Is Wrong*⁴¹. He cites many examples which show how crime doesn’t involve some group of ‘criminals’ who are somehow different from ‘non-criminals’, but rather is the product of opportunity and the lack of defence. “Crime has been with us since Adam and Eve, and, surprisingly, God didn’t spot the solution. It might have been better had he put the forbidden fruit higher up the tree.”⁴²

For instance, shoplifting soared when shops moved from having staff serving and all the products behind the counter, to the open-format we know today in which customers can serve themselves. That creates opportunity for people to *help* themselves. Did the criminal under-class suddenly grow? No. People caught at it included even a High Court judge.

Ross cites an experiment showing that people with no ill intent will engage with material they know to be illegal if given the opportunity, *even in preference to what they were trying to do*. Two British researchers set up online adverts for ‘freeware’ or ‘free software’. Those adverts actually directed surfers to a site offering three options: hard-core porn, soft-porn and free software. Of the 800 people who saw that page, only 26 had been looking for pornography: over 96% hadn’t, and yet fully 500 opted to see hard-core porn. The software – which people had sought in the first place – was the least popular.

Going back to shoplifting, the rate there fell again once scanners were installed at shop exits, because they make theft more difficult and so deter the opportunists. The pattern repeats elsewhere. Household burglary, for example, rose dramatically in the 1960s and 70s, by when barely anybody still had staff at home to guard their possessions, yet growing prosperity, cheap credit and increasing affordability of consumer goods meant that many houses were newly stocked with stuff worth nicking. Once home-owners started installing better window-locks and alarms – i.e., making crime difficult – the rates fell again. “Opportunity makes the thief. To a vast extent, it also creates the football hooligan, drink-driver, fraudster and murderer too”⁴³.

The implication is that **we can decrease crime simply by making crime harder**, not (just)

[†] As an aside, it’s striking that pictures on websites and reports of charities and public services working on crime are dominated by pictures of young men, often black, in hoodies and on council estates etc. We’ve yet to encounter a single picture depicting, say, corporate fraud or misselling of financial services, despite the recent huge fines and public disquiet.

by better policing and increasing sentences. It's the flip-side of Richard Thaler's comment which opens this document. Yet crime prevention concerns just one of the 190 training modules for police recruits. Nick Ross concludes that "It is hard to exaggerate the extent to which the criminal justice process is dissociated from the business of protecting victims and cutting crime."⁴⁴

Nick Ross now chairs the Jill Dando Institute of Crime Science at University College, London, (UCL) which describes its work:

"Crime science is a radical departure from the usual ways of thinking about and responding to the problem of crime and security. The distinct nature of crime science is captured in the name. First, crime science is about **crime**. Traditional criminological approaches are concerned largely with *criminality*, focusing on distant causes such as poverty, social disadvantage, parenting practices, and school performance. In contrast, crime scientists are concerned with near causes of *crime* – why, where, when, by whom, and how a particular offence is committed. They examine ways in which the immediate situation provides opportunities and provocations that account for the highly patterned distribution of crime events. Second, crime science is about **science**. Crime science is an evidence-based, problem-solving approach that embraces empirical research. Adopting the scientific method, crime scientists collect data on crime, generate hypotheses about crime patterns and trends, and build testable models to explain observed findings."

For instance, the Institute's Urban Lights project worked with UCL's experts in the built environment to establish that sodium (yellow) street lights makes it harder to recognise faces than white light. This is a key factor in reducing crime and fear of crime, and the project led to a change in UK standards for pedestrian lighting.

Similarly, the Design Against Crime Centre at leading design school Central St Martins⁴⁵ works to reduce temptation to opportunistic crime. For example, it has designed better bike locks, and better bike stands which are easier to lock a bike to, redesigned moneybelts as sexy lingerie, created an accessory in which a mobile phone can be carried by hand but not stolen, and raised awareness amongst bag designers of how to prevent theft e.g., by using Velcro seals which are impossible to open silently. These may all reduce crime but are far from standard activities for the police or criminal justice system.

Now that we've looked at each of the three parts of the current 'evidence system' around research by charities in crime reduction and criminal justice, we should note that the system really isn't working. Its basic purpose is to enable practitioners, policy-makers and funders to make evidence-informed decisions about which programmes to run, and it doesn't do that. The material is too often unpublished, too hard to find, and too unclear. We therefore turn now to our proposal for improving the system, pausing briefly to note some other factors which also need fixing: they, like those addressed by our proposal, are each necessary but insufficient.

3 Beyond the scope of this project

This project focuses on making charities' research more findable and clear. We found many others who share our view that improvements on these two fronts could make charities' work more consequential and improve outcomes.

However, it does not solve everything, and further work would remain around:

- **Data quality.** We heard several times about systems where information (e.g., about participants) is recorded on paper and then entered electronically. Clearly this creates ample potential for errors, and imperils the integrity of research using those data.
- **Research quality.** Research methods vary in their appropriateness to various circumstances and questions, and in their reliability. Poor methods are generally cheaper, quicker and more flattering. Even good research methods can yield poor quality research if they are used badly: for example RCTs are normally considered high quality though it is quite possible to conduct one so badly that the results are meaningless.

As discussed, research by charities is often poor quality. We did not explicitly explore how to fix this, though two people suggested a system for brokering academics and PhD students into charities to run their evaluations, and/or to advise on research design before any research starts. Those academics might need to be vetted for the relevance of their skills, but such a system might be a relatively cheap way to improve quality.

- Normally **the reader must discern the quality of the research**, which requires them to understand research methods. [Occasionally an intermediary has marked the research quality, e.g., the Education Endowment Foundation in its Toolkit⁴⁶ and medical Royal Colleges do it in the clinical practice guidelines they produce for doctors.] The information infrastructure system advocated by this document should give the reader enough information about the research to assess its quality: it won't do that assessment for them.
- **Ensuring demand.** Simply making material available does not guarantee that funders, commissioners, charities, or other target consumers will look for it nor use it. That requires incentives and motivations: for instance, when funders require applicants to cite evidence for their theory of change, this encourages applicants to seek out such evidence; and payment by results (pay-for-success) contracts encourage providers to research the most effective interventions. Equally, there is concern that target users will not understand the research nor be able to interpret and use it intelligently. This requires readers to know the merits and short-comings of various types of research and discern its appropriateness for their contexts. Nonetheless, organising the *supply* of evidence is clearly also necessary, and is the focus of this project.
- **The orchestra issue.** The purpose of research about effectiveness of interventions is to guide funders and practitioners to the best interventions. One funder used the compelling analogy that it helps them choose the best of several oboeists ... but does not normally help them see what other instruments are needed. (In other words, perhaps it's possible to have a set of great interventions without that being a great set of interventions.) For example in campaigning, it is sometimes important to have a multiplicity of organisations involved – to demonstrate the breadth of support and build a movement – even if some are sub-optimally effective.

4 The proposal and reactions

The proposal is to create (a) a short checklist of items to be clearly covered in charities' research (The checklist would define the minimum content: the organisation could publish anything else alongside), and (b) a repository for charities' research. Our first step was a consultation on this.

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

"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 (Co-ordinated Action Against Domestic Abuse)

"This would be a great contribution" – NGO in criminal justice

"There isn't anything systematic... it would certainly be useful" – major funder

"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..." – funder

The result, in summary, was that almost everybody agreed that we had identified an important problem and they supported both ideas to resolve it. Everybody agreed that solving these two problems would leave other problems outstanding, as discussed. On the repository, a few alternatives emerged and are discussed below. On the checklist, some additional items were suggested.

4.1 Making research clearer: The checklist

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 detail at least:

- i) **The intervention:** in enough detail that it could be replicated elsewhere. [See Box 7 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.

From (ii) and (iii), one can tell the quality of the research and the appropriateness of the research method used.

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

Nobody wanted any of those items removed. They are, as it happens, almost precisely the IMRAD structure which has become standard for scientific research (see Box 6), and the sections required by the Annals of Internal Medicine for the discussion section of research reports.[†]

Box 6 - IMRAD

Introduction, **M**ethods, **R**esults, **A**nd **D**iscussion. This structure is required by the International Committee of Medical Journal Editors, 'not [as] an arbitrary publication format but rather a direct reflection of the process of scientific discovery'.⁴⁷

- Introduction - Why was the study undertaken? What was the research question, the tested hypothesis or the purpose of the research?
- Methods - When, where, and how was the study done? What materials were used or who was included in the study groups (patients, etc.)?
- Results - What answer was found to the research question; what did the study find? Was the tested hypothesis true?
- Discussion - What might the answer imply and why does it matter? How does it fit in with what other researchers have found? What are the perspectives for future research?

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** (trial 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.

"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 funder^{††}

Several people saw value in having a standard 'vocabulary' for describing cohorts of beneficiaries. The descriptions should cover risk factors, not just socio-demographic data and crime history. Equally, perhaps definitions of interventions or outcomes could usefully be (more) standardised. This problem has been addressed in agriculture, which has developed a standard vocabulary (see Box 8). However, others felt that this is unnecessary, and that current descriptions are adequate. We can perhaps experiment with this.

- vi) Outcome and measurement tool.** This might be 'number of pregnancies' or 'literacy measured by Initial Assessment Tools for Literacy and Numeracy'. This is essential for enabling readers to

[†] "Annals of Internal Medicine recommends that authors structure the discussion section by presenting (1) a brief synopsis of the key findings, (2) consideration of possible mechanisms and explanations, (3) comparison with relevant findings from other published studies, (4) limitations of the present study (and methods used to minimise and compensate for those limitations), and (5) the clinical and research implications of the work, as appropriate."

^{††} This view is exacerbated by the (perception, accurate or not) that the 'triage' is inadequate. People in one area, or prison, or school can receive the services run by the organisations which happen to be there, rather than those necessarily best suited to their needs, which may 'belong' to organisations which happen to be elsewhere. Bad triage would clearly tend to decrease success rates.

compare interventions. One problem with research by charities is that they often use bespoke measurement tools which don't allow for comparisons. It's a problem in medicine too: a study of 2000 studies of schizophrenia found 640 different instruments, of which 369 had been used only once.⁴⁸

- vii) 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 BMJ trial reports in Appendix 3) since, as discussed, studies funded by pharmaceutical companies seem to get different answers than identical studies funded by other sources⁴⁹.
- viii) 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 all⁵⁰.) However, there were concerns about whether charities would shy from publishing cost data, for fear of seeming too cheap, or sometimes too expensive. Furthermore, cost data may prove too commercially sensitive since charities compete for funding. Thirdly, interpreting unit cost data requires care, since work to monitor need, outcomes, consistency of delivery and so on increase cost. And fourth, many research producers are unused to gathering cost data: even academics often don't because, remarkably, some economics journals not only don't require it but will in fact remove it from submitted research⁵¹.
- ix) 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.
- x) 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 succeed[†], contexts in which it is likely to work, and materials which can help others to implement it, such as training manuals.
- xi) Where there is a control group (randomised or propensity matched or constructed from other method), 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.

[†] 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.

- xii)** 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.
- xiii)** 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.

There is a choice as to whether the items on the checklist should be addressed 'just somewhere' in the research report, or whether they should all be answered in a structured abstract at the front (rather like a BMJ trial report). On the basis of empirical research, CONSORT "strongly recommend[s] the use of structured abstracts for reporting randomised trials⁵². Some studies have found that structured abstracts are of higher quality than the more traditional descriptive abstracts^{53, 54} and that they allow readers to find information more easily^{55, 56}. **Therefore we recommend a structured abstract**, organised around the items on the checklist. This makes it harder for charities to inadvertently omit items, and also makes the technology easier (see below). We refer to this for the remainder of this document as the 'structured abstract'. An example is in Figure 3.

Figure 3 - 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 Trust⁵⁷ 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, we worked with the team to map out the two stage process.

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

[†] This is the reason for the 'follow up' section of the CONSORT reporting flow-chart, in Appendix 4. Perhaps people receiving the treatment are literally more likely to die than are people in the control group, which is obviously important to know.

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 interviews 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 7 - Describing an Intervention

CONSORT asks that "The description should allow a [practitioner] to know exactly how to administer the intervention that was evaluated in the trial." 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:

- 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 (each procedure, activity, and/or 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 more so than most charities' descriptions. The example which CONSORT uses of an adequate description is as follows. Compare to the description of the mentoring programme analysed by the Justice Data Lab on page 14.

“In POISE [a study], patients received the first dose of the study drug (i.e., oral extended-release metoprolol 100 mg or matching placebo) 2-4 h before surgery. Study drug administration required a heart rate of 50 bpm or more and a systolic blood pressure of 100 mm Hg or greater; these haemodynamics were checked before each administration. If, at any time during the first 6 h after surgery, heart rate was 80 bpm or more and systolic blood pressure was 100 mm Hg or higher, patients received their first postoperative dose (extended-release metoprolol 100 mg or matched placebo) orally. If the study drug was not given during the first 6 h, patients received their first postoperative dose at 6 h after surgery. 12 h after the first postoperative dose, patients started taking oral extended-release metoprolol 200 mg or placebo every day for 30 days. If a patient's heart rate was consistently below 45 bpm or their systolic blood pressure dropped below 100 mm Hg, study drug was withheld until their heart rate or systolic blood pressure recovered; the study drug was then restarted at 100 mg once daily. Patients whose heart rate was consistently 45-49 bpm and systolic blood pressure exceeded 100 mm Hg delayed taking the study drug for 12 h.

Patients were randomly assigned to receive a custom-made neoprene splint to be worn at night or to usual care. The splint was a rigid rest orthosis recommended for use only at night. It covered the base of the thumb and the thenar eminence but not the wrist. Splints were made by 3 trained occupational therapists, who adjusted the splint for each patient so that the first web could be opened and the thumb placed in opposition with the first long finger. Patients were encouraged to contact the occupational therapist if they felt that the splint needed adjustment, pain increased while wearing the splint, or they had adverse effects (such as skin erosion). Because no treatment can be considered the gold standard in this situation, patients in the control and intervention groups received usual care at the discretion of their physician (general practitioner or rheumatologist). We decided not to use a placebo because, to our knowledge, no placebo for splinting has achieved successful blinding of patients, as recommended⁵⁸.

4.2 Making research easier to find: The repository

Three options emerged for making research easier to find, as follows.

1. A database

The advantages include that, being purpose-built, it can have whatever functionality the sector would like. By allowing any charity (as a research-producer) to submit material, the risk of small charities' material being overlooked is reduced. Material can be indexed suitably and consistently. Some databases (e.g., the ELDIS database of international development material, run by the Institute of Development Studies at the University of Sussex) scours the web for material as well as taking submissions, determines suitability of 'candidate' material, and writes or checks the abstracts for accuracy. (Abstracts are essentially an advert, luring people to read the full research, and surprisingly often are actually a poor match for what the research report contains^{59, 60}.)

Disadvantages include cost, in that it needs to be newly-built, and clearly there are cost implications of an editorial process, having writers to create or check abstracts, and having people to upload material. Such databases are rarely viable commercially, so require funding, which may run out, leaving the database stranded and unable to update. (ELDIS needs about three FTE staff, and gets around 600,000 unique visitors a year. It has a freemium model, and is subsidised by donors. Medical journals are commercial because of the incentives on academic researchers to get published. UNESCO has been creating a portal of evidence about education in less developed countries, including finding and indexing information, which has cost at least \$1.2m⁶¹.)

“People underestimate the work in doing them” – major funder

Options here include:

- a. Holding just the structured abstract vs. holding also the full research report. The latter is clearly more expensive, but the former is problematic because it allows for links to get broken and for contentious research to magically disappear from the publisher's site.
- b. Building onto an existing database, such as the Third Sector Knowledge Portal or the Arts Alliance Evidence Library (see Appendix 5).
- c. Taking only submitted material (as journals mainly do, though even they will sometimes commission research) or proactively looking for material to include (as ELDIS does).

2. Google search

This completely decentralised model is the opposite of a centralised database. Here, charities would publish their research on their own (or any) websites but tag it with particular meta-data in order that it is findable by search. (Scholar Google and various other engines are better than just basic engines.)

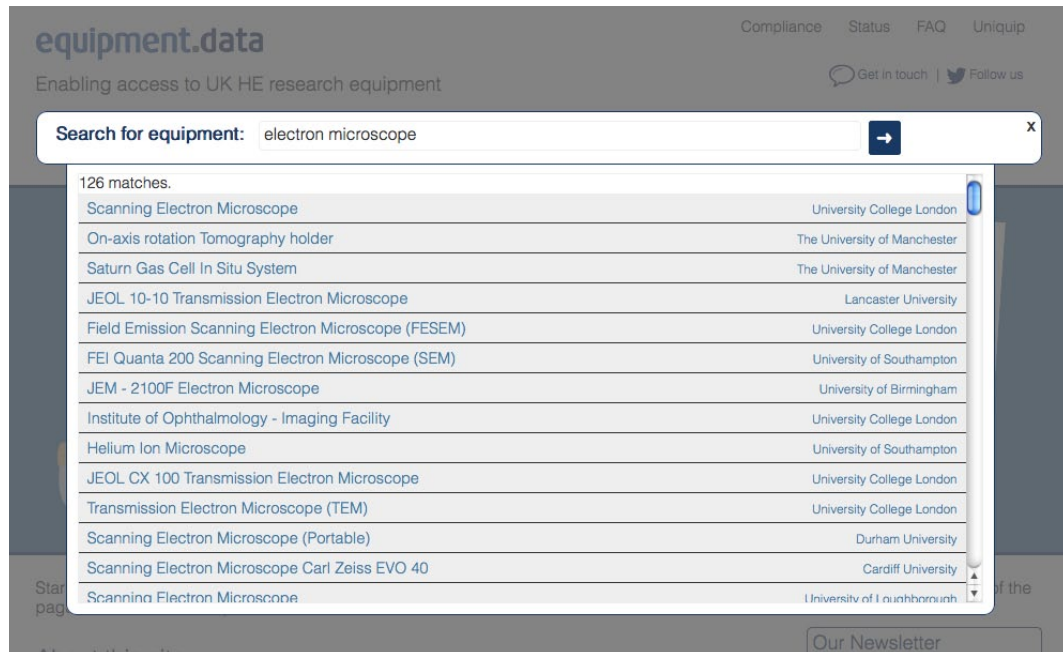
The advantages and disadvantages are the converse of those for a bespoke database. It's cheap, but this is because nobody is doing any sort of indexing or quality control, and there is no standardisation. Material can be removed without trace, abstracts may be misleading or wrong, or the tagging may be done wrong such that a search finds irrelevant material and misses target material. The burden of discerning relevance is shifted entirely to the reader.

3. Meta-data with a bespoke search-bot

This is an intermediary solution, and is our preferred option. The research and structured abstract are published on any website (i.e., a decentralised model), tagged with systematised meta-data. Rather than relying on Google or other generic search engines to find it, we create a simple central website with a bespoke search 'bot' which looks for (only) research published with that meta-data. Such 'bots' reports their results more usefully than normal search (see screenshot below) and can cache the results in case they become unfindable in future.

Examples of this system include Equipment.Data for finding equipment held by UK universities e.g., electron microscopes. The university simply uploads a spreadsheet onto its .ac.uk website with a column headed "Description" and each row describing an item of equipment. The university tells Equipment.Data the URL of that spreadsheet, the university's details, physical location (to enable search by proximity) and a contact for all items. The search function on Equipment.Data is much more user-friendly than, say, Google, picking up all and only the relevant information.

Screenshot of the Equipment.Data search 'bot'



The search results created by the 'bot' for research by charities in crime reduction and criminal justice would pick up the entries in the structured abstract for each item in the checklist, returning a table. So if you search for, say, reducing bike theft, it would return a table showing each piece of research[†] relevant to reducing bike theft, and for each (probably in a table, somewhat like that shown below), the full text which the charity has entered: intervention details, research question, research method, and results (and for any other checklist items which get added to the structured abstract). If a charity has left a field in the structured abstract blank, it would return a blank for that cell in the table.

[The system we build here may need to be more complicated than the Equipment.Data system because Equipment.Data is picking up simple and unambiguous data (equipment type, phone number etc.), whereas research about charities and crime reduction and criminal justice may require free-text fields which are harder to search.]

The meta-data can be open data even if the research report isn't, e.g., because it is behind a pay-wall. An advantage of the meta-data being open is that it can allow third parties to comment: this enables crowd-sourced quality-control, e.g., if the abstract radically differs from the actual research report, or if there are inaccuracies (e.g., 'this claims to be a randomised controlled trial but is in fact a case control study').

The advantages and disadvantages sit between those of the two previous options. It is cheaper and 'lighter' than a database, but requires less work from the user and less scope for error than does a completely decentralised model.

[†] An issue to be resolved is how the items are sequenced. Most likely, at the outset this might be something simple such as alphabetically by name of the charity publishing it. Over time it could plausibly become geographic or with some quality rating.

Figure 4 – Possible output of the search-bot

Research report from:	Intervention description	Research question	Research method	Findings
AAA Charity	Reducing burglary by doing aaa	Does doing X thing reduce incidence of burglary	RCT, 200 participants	20 percentage point reduction
BBB Charity	Reducing burglary by doing bbb	Does doing Y thing coincide with fewer burglaries	Cohort study, 3 years, 200 people	15 percentage point reduction
CCC Charity	Reducing burglary by doing ccc	How safe victims felt	In-depth interviews, 20 people	Strong support for the programme

Box 8 - Agrovoc

Natural resources
of Living resources
of Resource evaluation
of Resource exploration
Agricultural resources
Animal resources
Energy resources
Feed resources
Fishery resources
Food resources
Forest resources
NT Production forests
NT Protected forests
NT Protective forests
NT Genetic resources
NT Animal genetic resources
NT Gene banks
NT Gene pools
NT Germplasm
NT Plant genetic resources
NT Marine resources
NT Nonrenewable resources
NT Land resources
NT Abandoned land
NT Arable land
NT Common lands
NT Farmland
NT Cultivated land
NT Fields

AGROVOC⁶² is a 'controlled vocabulary' for indexing publications in agricultural science and technology. It aids researchers, librarians and information managers in indexing, retrieving and organising data in agricultural information systems and web pages. It covers food, nutrition, agriculture, fisheries, forestry, environment etc., and all areas of interest of the Food and Agriculture Organization (FAO) of the United Nations, which publishes it. It consists of over 32,000 concepts, and is available in 20 languages. A sample is to the left.

4.3 Concerns, key success factors, scope

There were two main concerns. First, about the **feasibility, complexity and cost** to charities of complying with the checklist and publishing in the database / with the meta-data. Second, about whether charities would do it: that is, what **incentive** they have to do it.

Our suggested next step is to run a pilot test with a small group of charities, in which they will complete the checklist, and will tag them with the meta-data needed for the intermediate option for the 'repository'.

Key success factors include:

1. Organisation to co-ordinate the system

There was wide agreement that the checklist and repository (whichever form it takes) needs to be run and co-ordinated by an organisation (a) widely respected by the criminal justice charity sector; (b) outside government, since there needs to be no impairment to charities publishing research critical of government; and (c) ideally outside academia, since charities' research isn't (usually) intended to be academic standard and is subject to different motivations and incentives. The obvious solution is **Clinks**, whose experience running the Arts Alliance Evidence Library is clearly valuable.

Key 'customers' for the improved information infrastructure system include the UK government's What Works Centre for Crime Reduction (see Box 2), academia, and the police, as well as other NGOs. The co-ordinating body should have or build good relationships with these audiences.

2. Ability to cope with research of different types, standards, purposes

Audiences differ in what they consider to be 'good' research, since they differ in their purposes. Therefore the system needs to not discriminate against some types of research: for example, it needs to be able to take research which is experimental (e.g. randomised controlled trials (RCTs)), qualitative (e.g., interviews with victims), statistical (e.g., re-offending rates), observational (e.g., cohort studies) and which investigates historical data and natural experiments (e.g., changes in sentences before and after the London 7/7 bombings). This need to take a range of research types was used by some interviewees to argue against the system being managed by academics, in that much 'operational' material may be useful to practitioners but might be screened out by academics to whom it is of no value.

Ideally the system would be able to cope with publishing data-sets, as well as research.

3. Charities' ability to complete the checklist

Accuracy: There may be a problem with charities being unable to accurately complete the structured checklist and therefore providing information which is wrong. For example, the checklist will ask for the research method used (such as those listed above) and charities may not know the technical term for the method they are using. This arises even in medicine: for example, in a study of studies in obstetrics and gynecology, 5% of 206 supposed RCTs were in fact not randomised⁶³.

We will address this in two ways. First by having a drop-down list of research types, which hopefully will minimise research getting lost for trivial reasons such as misspelling 'quantitative'. And second, by having a set of notes alongside the checklist to explain how to answer each item (a little like the notes from HMRC on completing each box in a tax return). We will use the pilot to fine-tune the guidance. It may transpire that (some) charities need training about research methods (on-line or in-person) before they can answer these questions reliably.

Project Oracle, which works to improve the evidence produced by charities involved with children and young people in London, reportedly has considerable difficulty getting charities to describe their interventions, research and findings. So the cause and degree of difficulty of getting charities to report accurately is currently unknown, and we hope that the pilot will illuminate it.

Quality control: It is of course possible that self-reporting of research will fail and the system will need some central mechanism for quality control. This is the role of editorial staff in journals and curated databases like ELDIS. It increases accuracy but also cost. The pilot should indicate whether this is necessary.

Cost of completing the structured abstracts: In medical and scientific research this is (reportedly) a non-issue: we heard of no complaints about the costs of using a checklist in reporting, possibly because it is genuinely easy and/or researchers are used to doing it.

Rights and ownership: There may sometimes be concern about who owns the research and therefore who has the right or obligation to decide whether to publish it and to write the structured abstract. This may arise where a researcher has been commissioned and paid by the charity, or where a researcher (e.g., an academic) has studied the charity's work (with their permission) but funded independently.

4. Free-text searching

Most of the items in the checklist will be answered in free-text (such as describing the intervention). Therefore the 'search-bot' will need to be able to search free-text.

Clearly, over time the system may develop to be less 'free' and more precise. Take research method as an example. At the outset, the checklist may simply ask for a description of this, and a charity may write 'interviews with 20 service users'. Eventually, there may be a drop-down list for research method (e.g., 'interviews') at which point the structured abstract may ask for the duration of those interviews, the number of them (in order that the number isn't omitted) and how those 20 people were chosen.

5. Copyright and caching

We are yet to explore whether copyright will be problematic for this system. The 'search-bot' might plausibly be viewed as re-publishing material, so could infringe the publisher's copyright. However, the publisher needs to tell the 'bot' the web address of material to be included, so perhaps the terms and conditions could deal with this. Copyright is unlikely to be a problem for anything published with a Creative Commons Licence, though might be a problem for research which goes into commercial journals (notably by academics).

As mentioned, an option is for the central website to create a cached copy of the research. Again, this may create copyright issues.

6. Incentives and enforcement

Of course, creating a checklist is a technical exercise, and quite different from the behaviour change work of getting people to use it.

Clearly charities will only complete the structured abstracts, publish research or publish using the meta-data if there is incentive to do so. This is most likely to involve funders and commissioners, for obvious reasons. We hope for a world in which funders and commissioners see sufficient value in research being published with the abstracts and meta-data that they will only accept research which is published and which has them. That is, that it becomes unacceptable to send research as an attachment and instead it must be sent with a link to a public document.

Several funders said to us that they might consider making publishing in this way a requirement of funding, or even eligibility to bid. Several other prestigious and important institutions – the Ministry of Justice, some academic centres – said that they would consider making a formal request to charities to publish their research in this way.

On enforcement, we do not envisage any person or organisation having a role of enforcing adherence to the checklist or meta-data. However, the effect of the search-bot is that it will be obvious if a

charity has published a research report whose structured abstract omits some fields: the 'bot' will return a blank for any empty items in the structured abstract. (This is in fact a stronger enforcement mechanism than CONSORT has.) The incentives on charities and scientists publishing research are analysed in Appendix 7.

7. Scope

The most natural research for the system described here is evaluations of interventions' effectiveness (just as CONSORT only covers medical clinical randomised controlled trials). This is where we will focus in the first instance. It is possible that it could be extended eventually to other research, such as policy research, how-to type policy research (e.g., how to design a social impact bond).

8. Is this 'build it and they will come'?

Related to the problem of charities' low incentives to produce and publicise research, is the question of whether the system suggested here (of the checklist and meta-data) assumes that once we build it, people will use it. The answer is no: hence we already started talking with funders and commissioners about whether they might ask charities to do this, or indeed require it. Our approach is to pilot the checklist and meta-data, to see if it is feasible and produces something of value, and then to create incentives and/or requirements to use them only if they turn out to be valuable. It is of course possible that no external incentive is needed, if making their research clearer and more findable is immediately beneficial to the charities which produce it.

5 Next steps: the pilot

As mentioned, medicine has many checklists for reporting research: so many that the CONSORT group published guidance on how to create them⁶⁴. This is very helpful: our process to date (unwittingly) virtually followed the steps it recommends and our outline process from here on in draws on it. CONSORT's experience is that making the guidelines credible and accepted 'is as much politics as science'⁶⁵.

To move from the current system to the system described, we envisage the following stages:

1. Pilot: in which we (a) have a few charities complete the checklist for reports, and (b) create the meta-data and have those few charities use it. If the system seems workable, we then:
2. Run a Delphi Exercise (a structured written consultation), inviting 100-200 people in relevant charities, academics, users of the research and others to comment on the checklist, research reports created and meta-data. Work with the steering group to adapt the system in light of the findings (for example, they may sort suggested additions to the checklist into 'omit', 'desirable' and 'essential'.)
3. Build the website and search 'bot'. This is probably the expensive part.
4. Finalise the materials to enable roll-out. These include:
 - a. A set of notes (and possibly worked examples) to help charities complete the checklist. These may be rather like the notes which accompany a tax-return ('how to complete box 10').
 - b. A more detailed note explaining the logic and provenance of the checklist and meta-data, its goals, endorsers and the logic for each item.
 - c. A simple website to hold the materials, worked examples and notes (This can be part of the search website.)
5. Roll out to the sector, possibly including having academics and commissioners formally call for charities to report in this way. There may be a need for training, either in-person and/or online.

Thereafter there is of course work to monitor the system to see where it needs tweaking or stronger incentives or enforcement in order to work better. Questions here include whether the structured abstracts are accurate; whether charities people are using it; what unexpected problems have emerged; and where would additional support be useful.

Discussed below is the detail of the first stage, the pilot. The purposes of the pilot are to assess in detail the feasibility of the proposed system (both of the checklist and tagging research with meta-data) and the costs and benefits, and to find the unexpected problems. Its specific goals include assessing:

- Feasibility, cost and time: Is it possible to create meta-data which charities can complete, and how long does it take them to complete it?
- Support: What proportion of charities need support completing the checklist and/or the meta-data, and what sort of support do they need? What is the nature and cost of that?

- Benefit: Do charities and their funders and commissioners find value in the structured abstracts?
- Accuracy of the structured abstracts: What proportion of charity-generated structured abstracts are accurate and hence is any central checking or support needed? Perhaps we experiment with having a central person complete some to see what that costs, its benefits and how charities feel about it.
- Cost: Do charities see enough value in the system to do it, or do they need to be paid or otherwise incentivized in order to do it?

The stages are roughly as follows:

1. **Secure funding** for the pilot.
2. **Set up a 'steering group'**. The purpose here is to bring insight and credibility to the process and the eventual checklist and system, which is designed to influence a whole sector. They will help to refine the checklist criteria and explanatory notes which go with it. The steering group may eventually endorse – and even enforce – the system. Several people have offered (unsolicited) to serve on it.
3. **Recruit** a group of charities for the pilot. We will invite a random sample of [Clinks' members](#) (of which there are over 600). We may do a first round of this with very few organisations, say about six; and then a second round with about 50 if it works. This may save on cost and reputational damage if it all turns out to be a nightmare.

Of course, the pilot may wobble at this point if too few charities are interested in being involved. Perhaps we will need to pay them to be in the pilot. It will be interesting to look at who does and doesn't respond: perhaps only the more progressive, or large, or state-funded charities will want to be involved. A 'sell' to the charities is that by being involved in the pilot they get more sway over the eventual system. It may therefore help to have MoJ or commissioners endorse or be involved in the pilot.

4. **Checklist**: We would ask the pilot group to complete the checklist. We might start by providing no guidance at all, providing it gradually on an as-needed basis in order that we can assess what is really needed. That support might be 'notes' on each checklist item, and/or worked examples. We will check the resulting structured abstracts to assess their accuracy and whether a central quality control mechanism is necessary. For some, the 'host' will write the checklist for them (as ELDIS does) to test the feasibility, accuracy, cost and popularity of that model.
5. **Design the meta-data**: Here we would work with the Open Data Institute who are experts at this, and it would be a workshop-based process, collaborative with the charities in the pilot.
6. **Pilot the meta-data**: We ask the charities in the pilot to tag their research using the meta-data, looking again for the time it takes them to do this, the accuracy, the support they need, and need for quality-control.
7. **Gather feedback from the charities, and also from some funders, police and commissioners**: Clearly we will have been in discussion with the charities all along, but will gather their views systematically at this point. We will take the structured abstracts to some funders, police and commissioners to gauge how useful they find them and their appetite for supporting a roll-out.

Clearly findings at this point determine what happens next. With luck, the process turns out to be relatively painless and produces valuable products. In that case, we proceed. Of course if the process is horrible and the product adds no value, we may stop at this point. If we proceed, we will publicise the findings (e.g., through the sector press), and look to roll the system to charities across the crime reduction and criminal justice sector. We hope that eventually *all* research by charities is 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 ago[†] ⁶⁶.

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. As mentioned, 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.

[†] 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: Research method and people interviewed and consulted

This project was essentially a non-formal consultation on the concepts outlined, of a standardised checklist and a repository. The concept arose from Giving Evidence's long-standing work on improving the quality and use of evidence by NGOs and funders, and our recent study of the evidence system in medicine and its relevance to education in less developed countries (published at www.giving-evidence.com/education). The project had two main strands. First, reviewing relevant literature. Second, interviewing relevant people. On the latter, we contacted various relevant organisations, and publicised the project (e.g., through social media, Giving Evidence's newsletter and website, a post on Markets4Good's blog⁶⁷, interview with 80,000 Hours, etc.), which yielded further interviewees and ideas. We also conducted some original meta-research (research about research) e.g., that in Appendix 2. The study ran June-October 2014.

We're grateful to the following people (and any others who we have inadvertently omitted) for sharing their time and insights:

Dr Mark Abrams, Senior Research Officer, College of Policing
Professor Doug Altman, Centre for Medical Statistics, University of Oxford; member, CONSORT Executive Group
Owen Barder, Director, Center for Global Development Europe
Lis Bates, Head of Research and Evaluation, CAADA (Co-ordinated Action Against Domestic Abuse)
Ken Berger, Executive Director, Charity Navigator
David Bonbright, Director, Keystone Accountability
Liz Carolan, International Development Manager, the Open Data Institute
Professor Sir Iain Chalmers, founder, The Cochrane Collaboration
Professor Mike Clarke, Director of the All Ireland Hub for Trials Methodology Research; former chair, The Cochrane Collaboration
Jon Cracknell, co-ordinator, Environmental Funders Network
Rebecca Endeans, Chief Scientific Adviser and Director of Analytical Services, Ministry of Justice
Lucy Gampell, Trustee, Clinks; former CEO, Action for Prisoners' Families
Dr Ben Goldacre, Bad Science and London School of Hygiene and Tropical Medicine
Shaks Gosh, former CEO Crisis and the Private Equity Foundation
Professor Jonathan Grant, Director, The Policy Institute, King's College London
Suzanne Grant Lewis, Director, Plan for 4 Learning Portal, UNESCO
Roma Hooper, founding Chair, Prison Radio Association; founding Director, Make Justice Work campaign
Carol Jackson, Head of Outreach, Assessment and Outcomes, The Prince's Trust
Harvey Koh, Monitor Inclusive Markets, formerly Chief Operating Officer, the Private Equity Foundation
Sara Llewellyn, Chief Executive, the Barrow Cadbury Trust
Sarah Lucas and Kristen Stelljes, Program Officers at The William and Flora Hewlett Foundation
Marc Maxson, Global Giving
Dr Nicky Miller, Research Director, College of Policing
John Mohan, Director, Third Sector Research Centre and Professor of Social Policy, University of Birmingham
Paul Montgomery, Professor of Psycho-Social Intervention and Course Director for Evidence-Based Social Intervention, University of Oxford

Jess Mullen, Project Co-ordinator, Clinks
Alex Murray, Chief Superintendent, Solihull; Founder, Society for Evidence-Based Policing
Professor Ken Pease, Professor of Crime Science at the UCL Department of Security and Crime Science
Jess Plant, Arts Alliance Manager, Clinks
David Pritchard, Head of Measurement and Evaluation, New Philanthropy Capital
Ruth Puttick, Principal Researcher for public and social innovation, NESTA
Charlotte Ravenscroft, Head of Policy and Research, NCVO
Nick Ross, Chair, Jill Dando Institute of Crime Science, University College London
Paul Simpson, Deputy Director, PLoS (Public Library of Science)
Barbara Storch, Portfolio Director, Impetus-Private Equity Foundation
Dr Jeni Tennison, Technical Director, Open Data Institute
Victoria Vrana, Senior Program Officer, Gates Foundation
Dr Nicola Wagner-Rundell, Head of Impact Management, Impetus-Private Equity Foundation
Karl Wilding, Director of Public Policy, NCVO
Jo Wilkinson, Practice Manager, College of Policing
Two major anonymous family foundations

Appendix 2: 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 3: British Medical Journal 'trial reports': two examples

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.

RESEARCH

Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials

Alexander C Ford,^{1,2} David Forman,³ Richard H Hunt,⁴ Yuhong Yuan,⁴ Paul Moayyedi⁴

¹Leeds Gastroenterology Institute, St James's University Hospital, Leeds LS9 7TF, UK

²Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, Leeds, UK

³Cancer Information Section, International Agency for Research on Cancer, Lyon, France

⁴Farncombe Family Digestive Health Research Institute, Gastroenterology Division, McMaster University, Health Sciences Center, Hamilton, Ontario, Canada

Correspondence to: A C Ford alexf12399@yahoo.com

Cite this as: *BMJ* 2014;348:g3174 doi:10.1136/bmj.g3174

This is a summary of a paper that was published on *bmj.com* as *BMJ* 2014;348:g3174

bmj.com

● Clinical review: The diagnosis and management of gastric cancer (*BMJ* 2013;347:f6367)

● Letter: Should *H pylori* always be eradicated? (*BMJ* 2012;344:e2145)

● Clinical review: Treatment of *Helicobacter pylori* infection (*BMJ* 2008;337:a1454)

● Editorial: Who benefits from *Helicobacter pylori* eradication? (*BMJ* 2006;332:187)

STUDY QUESTION

Does searching for *Helicobacter pylori* and treating with eradication therapy among healthy asymptomatic infected individuals reduce the subsequent incidence of gastric cancer?

SUMMARY ANSWER

Yes, only 51 (1.6%) gastric cancers occurred among 3294 individuals who received eradication therapy compared with 76 (2.4%) in 3203 control subjects (relative risk 0.66, 95% confidence interval 0.46 to 0.95).

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Searching for and eradicating *H pylori* could, theoretically, reduce the incidence of gastric cancer, but evidence is conflicting. In this meta-analysis of randomised controlled trials gastric cancer incidence was significantly lower in healthy asymptomatic infected individuals who received eradication therapy compared with those who received placebo or no treatment, with a number needed to treat of 124 overall. If the benefit of eradication therapy was assumed to persist lifelong the number needed to treat was as low as 15 for Chinese men, and as high as 245 for US women.

Selection criteria for studies

We searched Medline, Embase, and the Cochrane central register of controlled trials through to December 2013, conference proceedings between 2001 and 2013, and bibliographies of relevant studies. Eligible studies were randomised controlled trials examining the effect of at least seven days of eradication therapy on subsequent occurrence of gastric cancer in adults infected with *Helicobacter pylori* who were

otherwise healthy and asymptomatic. The control arm had to receive placebo or no treatment. Subjects had to be followed for at least two years, and studies had to report at least two cases of gastric cancer to be included.

Primary outcome(s)

The primary outcome, defined a priori, was the effect of eradication therapy on the subsequent occurrence of gastric cancer, expressed as a relative risk of gastric cancer with 95% confidence intervals.

Main results and role of chance

The search strategy identified 1560 citations, of which six individual randomised controlled trials were eligible. Fifty one (1.6%) gastric cancers occurred among the 3294 individuals who received eradication therapy, compared with 76 (2.4%) in the 3203 control subjects (relative risk 0.66, 95% confidence interval 0.46 to 0.95), with no heterogeneity between studies ($I^2=0\%$, $P=0.60$) (figure). If the benefit of eradication therapy was assumed to persist lifelong the number needed to treat was as low as 15 for Chinese men (who have a high lifetime risk of gastric cancer) and as high as 245 for US women (who have a low lifetime risk).

Overall, there were 24 deaths (1.1%) from gastric cancer among 2242 individuals randomised to eradication therapy, compared with 36 (1.6%) deaths in 2233 participants allocated to placebo (relative risk of death from gastric cancer 0.67, 0.40 to 1.11), with no heterogeneity ($I^2=0\%$, $P=0.90$). In total, 192 (7.3%) of 2639 subjects who received eradication therapy were dead from any cause at the last point of follow-up, compared with 175 (6.7%) of 2614 individuals who received placebo or no treatment (relative risk of death from any cause 1.09, 0.86 to 1.38), with no heterogeneity ($I^2=6\%$, $P=0.36$).

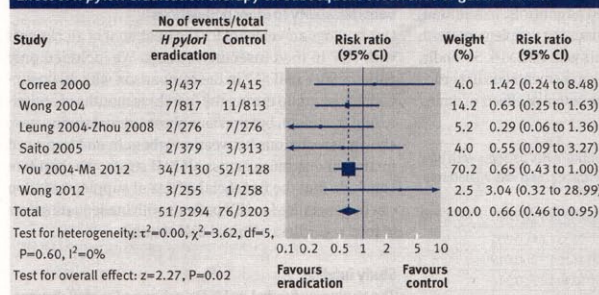
Bias, confounding, and other reasons for caution

Only three of the six randomised controlled trials we identified were at low risk of bias. All but one study was conducted in East Asia, so it is not possible to assess the effect of searching for and eradicating *H pylori* in Western populations. Individual adverse events data were not reported by many of the trials we identified, so we were not able to assess the balance of benefits and harms if searching for and eradicating *H pylori* infection were to be adopted in the general population.

Study funding/potential competing interests

This study was unfunded. The authors declare no competing interests.

Effect of *H pylori* eradication therapy on subsequent occurrence of gastric cancer



Appendix 4: CONSORT Checklist and flow chart for reporting medical RCTs

CONSORT 2010 checklist of information to include when reporting a randomised trial^{68*}



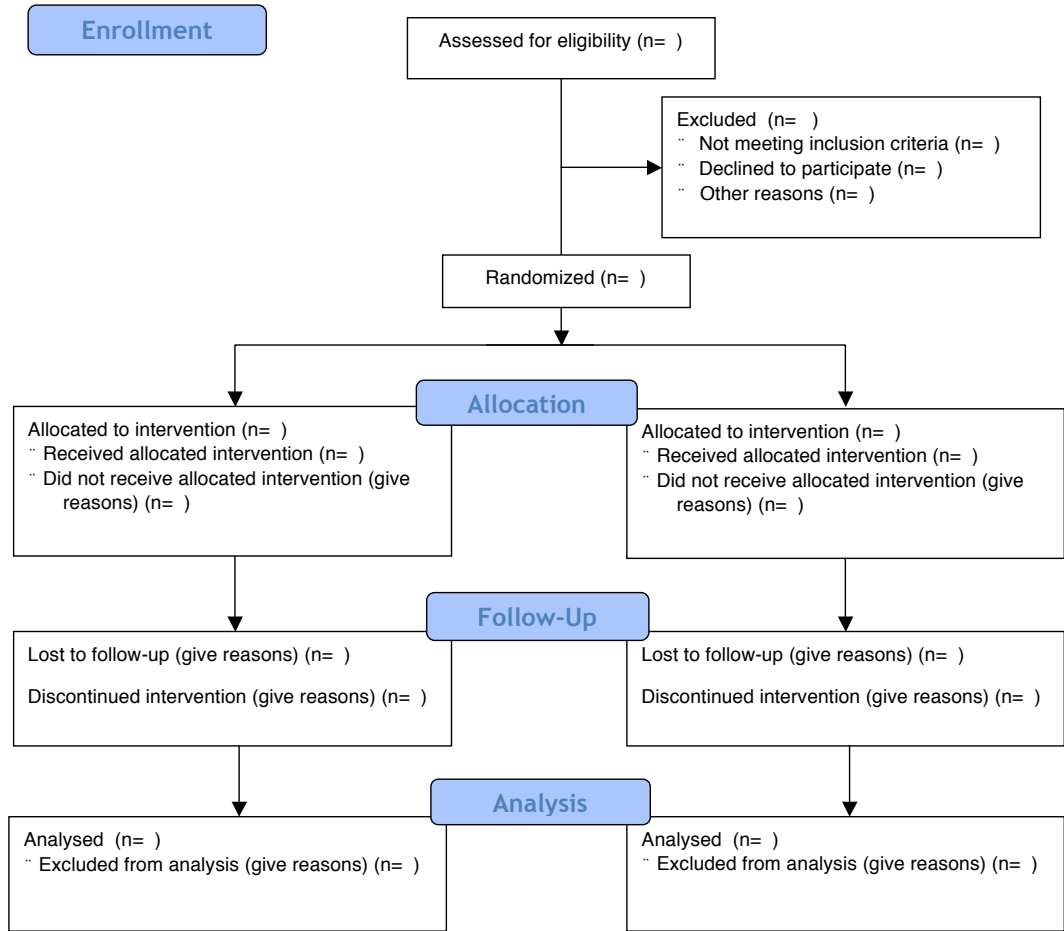
Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	

Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	

Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

*The CONSORT statement says: We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 Flow Diagram



Appendix 5: Instructive models for consolidating research

These collections of evidence offer lessons for the 'repository' and, since we should avoid replicating them, influence its scope.

The Arts Alliance Evidence Library

The Arts Alliance is a coalition of arts organisations working in the UK criminal justice system, managed by Clinks. Its evidence library grew from a literature review, and now contains evaluations of arts-based practice across the criminal justice system conducted independently by academics or research organisations. It currently holds 87 evaluations.

The library is managed by Clinks which does the work of assessing whether submitted material should be included. Clinks also sources external material to include.

There are no quality criteria. As mentioned, only four of the evaluations in the library (of which there were then 86) met the quality criteria for inclusion in a 'rapid evidence assessment' (a speedy systematic review) commissioned by the National Offender Management Service in 2013⁶⁹.

Clinks reports that NGOs use material in the library to support funding applications, by finding evidence which supports their theory of change.

Homeless Pages

Homeless Pages is "the most comprehensive listing of UK publications, research and good practice on homelessness and related issues". It contains 1,800 publications from 650 organisations, including 'about 120 voluntary organisations, 30 commercial publishers, 30 statutory bodies and over 400 local authorities'. It is managed by Homeless Link, the membership charity for organisations working with homeless people in England, and anybody can suggest content to be included.

Youth Justice Board's Effective Practice Library

The Youth Justice Board (part of the Ministry of Justice) collates in this online library details of 'practice which produces the intended results'⁷⁰. Anybody can submit an intervention, which a panel of academics and Youth Justice Board research staff sort into: research-proven; promising evidence; emerging evidence; treat with caution; and ineffective. It judges quality by using the following fairly standard hierarchy for quantitative evidence⁷¹ and set of considerations for qualitative evidence⁷².

Hierarchy of quantitative evidence used by the Youth Justice Board

<p>Level 1 A relationship between the intervention and intended outcome. Compares the difference between outcome measure before and after the intervention (<i>intervention group with no comparison group</i>)</p>
<p>Level 2 Expected outcome compared to actual outcome for intervention group (<i>e.g. risk predictor with no comparison group</i>)</p>
<p>Level 3 Comparison group present without demonstrated comparability to intervention group (<i>unmatched comparison group</i>)</p>
<p>Level 4 Comparison group matched to intervention group on theoretically relevant factors e.g. risk of reconviction (<i>well-matched comparison group</i>)</p>
<p>Level 5 Random assignment of offenders to the intervention and control conditions (<i>Randomised Controlled Trial</i>)</p>

Considerations for assessing quantitative evidence used by the Youth Justice Board

<p>FINDINGS</p> <ul style="list-style-type: none"> • How credible are the findings? • How well does the evaluation address its original aims and purpose? • Is there scope for generalisation?
<p>DESIGN</p> <ul style="list-style-type: none"> • How defensible is the research design?
<p>SAMPLE</p> <ul style="list-style-type: none"> • Sample selection / composition – how well is the eventual coverage described?
<p>DATA COLLECTION</p> <ul style="list-style-type: none"> • How well is the data collection carried out?
<p>ANALYSIS</p> <ul style="list-style-type: none"> • How well has the detail, depth and complexity of the data been conveyed? • How clear are the links between data, interpretation and conclusions – i.e. how well can the route to the conclusion be seen?
<p>REPORTING</p> <ul style="list-style-type: none"> • How clear and coherent is the reporting?

The incentives for researchers or practitioners to submit their research is not clear. However, the library contains details of around 200 interventions. Most have been run by public agencies, but some by charities⁷³. For most items in the library, there is a description not unlike the abstract suggested here (not all descriptions are complete):

- Target cohort (description, age, gender). Note that these descriptions, though sometimes detailed, are sometimes not detailed and do not appear to be standardised
- Cost (though for some programmes, this simply says 'yes'⁷⁴)
- Intended outcomes
- Intervention description, in reasonable detail
- Some points on 'What makes it work'
- Some points on how to implement
- Contact details for the individual and organisation which ran it
- 'Supporting documents', which in some cases include the detailed submission to the library (around 25 pages), or materials describing the theory of change.

POLKA (the UK Police OnLine Knowledge Area) and Research Map

POLKA is 'a secure online collaboration tool for the policing community to network, ask questions, share insights, discuss ideas' and only available to people in the police and parts of government. POLKA and the Research Map are managed by the College of Policing within the What Works Centre.

The Research Map on the College of Policing website⁷⁵ is designed to avoid duplication. It takes details of any research about policing or crime reduction which is in-progress and intended for eventual peer-review, Masters projects, PhD, research by universities and practitioners. When research is completed, it is removed from the map and the College of Policing asks that it be added to the National Police Library. The College would like for more charities' research to be included.

US Council of State Governors Justice

The What Works in Reentry Clearinghouse⁷⁶ has details of various schemes for offenders re-entering normal life after detention. Each is graded thus:

Ratings Key	
High Rigor	Basic Rigor
	Strong evidence of a beneficial effect
	Modest evidence of a beneficial effect
	No statistically significant findings
	Modest evidence of a harmful effect
	Strong evidence of a harmful effect

For each intervention, the database has fields for the following details (again, not all fields are completed in all cases):

- Overview of the intervention
- Summary of findings
- Recommendations for practice. (This is sometimes answered simply Yes/No, meaning whether the study contains any recommendations.)
- Beneficiaries' age, gender, and US state where it was run
- Research method (e.g., sample size, how a control group was constructed)
- Underlying research papers from which the summary is drawn. Where several evaluations have studied the same or similar interventions, the results are shown together (which is obviously easier for the reader trying to find them). Where those studies show different results, the grade for each finding is shown:

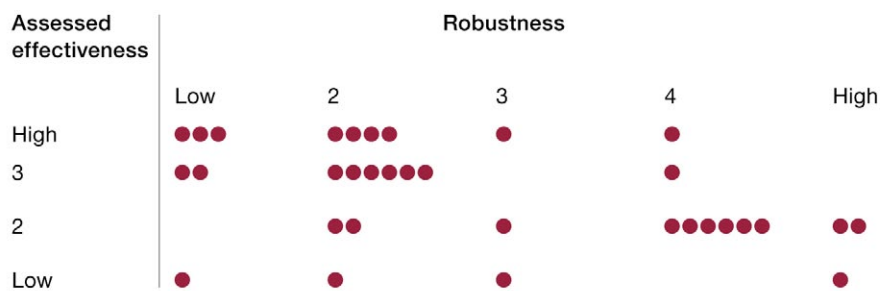
Outcome	Evaluated Findings
Recidivism	
Employment	

Appendix 6: Bad research flatters more than good research

The UK's National Audit Office (NAO) published analysis of the quality of almost 6,000 government evaluations, which contains a salutary nugget. It found that **the strongest claims about effectiveness are based on the weakest research**. This (probably) isn't because the researchers are wicked, but rather because you can infer almost anything from a survey of two people: most social interventions have quite small effects, and robust research won't let you show anything bigger.

Figure 11

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.

Source: National Audit Office analysis of external assessment by London School of Economics

Appendix 7: Incentives on charities and researchers in publishing and indexing research

It may be useful to analyse the incentives in the current system for publishing scientific and medical research, and thereby seeing what the incentives might be on charities in the system we are advocating:

Activity	Who does it, and their incentive	
	Medical / scientific research (currently)	Charities in criminal justice (in the system advocated)
Producing research	Researchers in academia or pharma companies. (In medicine, academic research is co-mingled with practice: hospital doctors often conduct research.) Partly in pursuit of knowledge, partly because of professional incentives: for their research to publish in prestigious journals, and to be cited (hence PLoS is able to charge researchers to publish). Pharma companies have commercial incentives.	Published by the charity itself. Incentives currently aren't clear, but charities are under constant existential pressure to raise money. Funders would need to require material to be published this way.
Publishing research	Published by journals. Commercial journals want material which will generate sales (of the magazine, subscriptions, and/or reprints of articles ⁷⁷), hence run their own editorial and quality control processes.	
Writing abstracts	The researchers. Abstracts are essentially adverts for the research: well-written ones increase readership and hence citations.	<i>Not normally done currently.</i> Funders would need to require that material be published with the structured abstract.
Indexing research	The journals, in order to make their articles (i.e., their product) easier to find, which drives citations, and increases their journal's attractiveness to researchers, both to read (i.e., to buy) and in which to publish.	<i>Not normally done currently.</i> The meta-data would have this effect. Funders would need to require that material be published with the meta-data.

Appendix 8: Relevant academic centres

In no particular order:

<p>The Centre for Research on Families and Relationships, based at the University of Edinburgh⁷⁸</p>	<p>A consortium research centre based at The University of Edinburgh, with partners at the Universities of Aberdeen, Glasgow, Glasgow Caledonian, Highlands & Islands, and Stirling.</p> <p>Funded by the Scottish Government, the EU, the public sector (e.g., through ESRC and NHS) and foundations (e.g., Big Lottery Fund, Joseph Rowntree Foundation).</p> <p>Research areas include: gender-based violence, effect of family environment on environmental attitudes and behaviours, sources and solutions for inequality. Also research interest around the influence of research on public policy.</p>
<p>Jill Dando Institute of Security and Crime Science, UCL</p>	<p>Focuses on ways to cut crime and increase security, which don't necessarily relate to criminals as such (see Box 5). Interdisciplinary centre, drawing on architecture, economics, engineering, geography, medicine, psychology, statistics and town planning. Chaired and co-founded by Nick Ross, who co-presented Crimewatch on the BBC for years.</p>
<p>The POP Center (Center for Problem-Oriented Policing), University at Albany, New York</p>	<p>The POP Center produces guidance on dealing with particular types of crime, e.g. drug-dealing at open markets, robbery in taxis. Each guide is informed by a thorough review of the literature and police practice, and anonymously peer-reviewed by police and a researcher. They cover risk factors and prevention, e.g., the guide on violence in pubs talks about crowding and bad lighting, and pointers for analysing causes and determining tactics for reducing incidence⁷⁹.</p>
<p>Universities' Police Science Institute, Cardiff University</p>	<p>Established in 2007 in partnership with South Wales Police to develop evidence-based policing approaches to tackling crime and anti-social behaviour. Evolved from work by a Professor Jonathan Shepherd (from whom we learnt about evidence systems), a maxillofacial and dental surgeon who noticed that he was treating facial injuries from many more drunken fights than the crime statistics said had happened⁸⁰.</p>
<p>Scottish Institute for Policing Research</p>	<p>A collaboration between the Police Service of Scotland and 12 of Scotland's universities*. Funded by Scottish Funding Council and the Association of Chief Police Officers in Scotland. Focuses (as the name implies) on policing interventions.</p> <p>* Universities of Abertay Dundee, Dundee, Edinburgh, Edinburgh Napier, Glasgow, Glasgow Caledonian, Heriot-Watt, Robert Gordon, St Andrews, Stirling, Strathclyde, and The West of Scotland.</p>

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].
- 2** Altman, D. (2014) Professor and Director of The Centre for Statistics in Medicine at Oxford University. Interview with Fiennes, C. October 2014.
- 3** Thaler, R.H. (2012) Watching Behavior Before Writing the Rules. *The New York Times*. [Online] 7 July. http://www.nytimes.com/2012/07/08/business/behavioral-science-can-help-guide-policy-economic-view.html?_r=2and [Accessed: 24.09.14].
- 4** Shepherd, J. (2007) The production and management of evidence for public service reform. *Evidence and Policy*. 3(2), pp. 231-251.
- 5** Interviewee. (2014) Conversation with Fiennes, C., September 2014.
- 6** Goldacre, B. (2010) *Bad Science*. Fourth Estate. pp. 49-50.
- 7** U.S. National Library of Medicine. *Research Reporting Guidelines and Initiatives: By Organization*. [Online] http://www.nlm.nih.gov/services/research_report_guide.html [Accessed: 24.09.14].
- 8** Schulz, K.F., Altman, D.G., Moher, D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.
- 9** Strobe. STROBE Statement. [Online] <http://www.strobe-statement.org/> [Accessed: 02.10.14].
- 10** Prisma. [Online] <http://www.prisma-statement.org/> [Accessed: 02.10.14].
- 11** University of Oxford. *CONSORT-SPI: A CONSORT Extension for Social and Psychological Interventions*. [Online] <http://www.spi.ox.ac.uk/research/details/consort-spi-a-consort-extension-for-soc.html> [Assessed: 10.10.14].
- 12** Consort. *The Consort Statement*. [Online] <http://www.consort-statement.org/> [Accessed: 24.09.14].
- 13** Moher, D., Schulz, K.F., Simera, I., Altman, D.G. (2010) Guidance for Developers of Health Research Reporting Guidelines. *PLoS Medicine*. 7(2), p. 8.
- 14** Consort. Impact of CONSORT. [Online] <http://www.consort-statement.org/about-consort/impact-of-consort> [Accessed: 02.10.14].
- 15** National Offender Management Service. *Intermediate outcomes of arts projects: a rapid evidence assessment* [Online] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254450/Intermediate-outcomes-of-arts-projects.pdf [Accessed: 24.09.14].
- 16** Paul Hamlyn Foundation. *Assessing Impact*. [Online] http://issuu.com/paulhamlynfoundation/docs/assessing_impact?e=1112503/2404563%20page%209 [Accessed: 24.09.14].
- 17** Wilkinson, J. (2014) Practice Manager, College of Policing. Private correspondence with Fiennes, C., September 2014.
- 18** Gough, D., Oliver, S., Thomas, J. (2012) *An Introduction to Systematic Reviews*. SAGE Publications.
- 19** Nesta. *Why We Need to Create a 'NICE for social policy'*. [Online] http://www.nesta.org.uk/sites/default/files/why_we_need_to_create_a_nice_for_social_policy.pdf [Accessed: 24.09.14].
- 20** College of Policing. *Commissioned partnership programme update*. [Online] <http://www.college.police.uk/en/20825.htm> [Accessed: 24.09.14].

- 21 Hopewell, S., Collins, G.S., Boutron, I., Yu, L-M., Cook, J., Shanyinde, M., Wharton, R., Shamseer, L., Altman, D.G. (2014) Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *The British Medical Journal*. 349. [Online] <http://www.bmj.com/content/349/bmj.g4145> [Accessed: 24.09.14].
- 22 Bohannon, J. (2013) Who's Afraid of Peer Review? *Science*. 342(6154), pp. 60-65.
- 23 *Laboratory News*. (2013) The Problem with Peer Review. [Online] <http://www.labnews.co.uk/features/peer-review/> [Accessed: 24.09.14].
- 24 Ministry of Justice. (2013) *Justice Data Lab, Methodology Paper*. [Online] <https://www.justice.gov.uk/downloads/justice-data-lab/justice-data-lab-methodology.pdf> [Accessed: 24.09.14].
- 25 Ministry of Justice. (2014) *Justice Data Lab pilot statistics*. [Online] <https://www.gov.uk/government/collections/justice-data-lab-pilot-statistics> [Accessed: 24.09.14].
- 26 Clinks. Justice Data Lab – frequently asked questions. [Online] Available from: <http://www.clinks.org/sites/default/files/JusticeDataLabFAQs.pdf> [Accessed: 24.09.14].
- 27 Pritchard, D. (2014) Head of Measurement and Evaluation, New Philanthropy Capital. Interview with Fiennes, C., September 2014.
- 28 Fiennes, C. (2013) Most Charities Shouldn't Evaluate Their Work. *Stanford Social Innovation Review*. [Online] http://www.ssireview.org/blog/entry/most_charities_shouldnt_evaluate_their_work [Accessed: 24.09.14].
- 29 Project Oracle. (2014) *How Can We Make Evidence Easier?* [Online] <http://project-oracle.com/news-and-events/how-can-we-make-evidence-easier/> [Accessed: 24.09.14].
- 30 The Centre for Social Justice (2013). *The New Probation Landscape*. [Online] <http://www.centreforsocialjustice.org.uk/UserStorage/pdf/Pdf%20reports/landscape.pdf>
- 31 NPC. (2005) *Inside and Out. People in prison and life after release: A guide for donors and funders*. [Online] <http://www.thinknpc.org/wp-content/uploads/2012/09/Inside-and-Out.pdf> [Accessed: 24.09.14].
- 32 Lexchin, J., Bero, L.A., Djulbegovic, B., Clark, O. (2003) Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *The British Medical Journal*. 326.
- 33 Bohannon, J. (2013) Who's Afraid of Peer Review? *Science*. 342(6154), pp. 60-65.
- 34 Goldacre, B. (2013) *Bad Pharma*. Fourth Estate.
- 35 European Venture Philanthropy Association. *Knowledge Centre*. [Online] <http://evpa.eu.com/research-and-policy/knowledge-centre/> [Accessed: 02.10.14].
- 36 NPC. (2012) *Making an Impact: Impact Measurement Across Charities and Social Enterprises in the UK*. [Online] <http://www.thinknpc.org/publications/making-an-impact/making-an-impact/> [Accessed: 24.09.14].
- 37 Findlay, M. *Aversion to Learning in Development? A Global Field Experiment on Microfinance Institutions*. [Online] http://www.michael-findley.com/uploads/2/0/4/5/20455799/mfi_learning.22mar13.pdf [Accessed on: 24.09.14].
- 38 The Wheel. (2011) *Demonstrating Impact: Current Practice Amongst Social Purpose Organisations in the Republic of Ireland*. [Online] http://www.wheel.ie/sites/default/files/Demonstrating_Impact_Report_Final-1.pdf [Accessed: 24.09.14].
- 39 National Audit Office. *Evaluation in Government*. [Online] <http://www.nao.org.uk/report/evaluation-government/> [Accessed: 24.09.14].
- 40 Goldacre, B. (2012) TED talks. *What doctors don't know about the drugs they prescribe*. [Online] https://www.ted.com/talks/ben_goldacre_what_doctors_don_t_know_about_the_drugs_they_prescribe [Accessed: 24.09.14].
- 41 Ross, N. (2013) *Crime: How To Solve It, and Why So Much of What We're Told Is Wrong*. Biteback Publishing.
- 42 *Ibid.*, Chapter 1.

- 43 *Ibid.*, Kindle location 155.
- 44 *Ibid.*, Kindle location 3261.
- 45 Design Against Crime. [Online] <http://www.designagainstcrime.com/> [Accessed: 24.09.14].
- 46 Education Endowment Foundation. *Toolkit* [Online] <http://educationendowmentfoundation.org.uk/index.php/toolkit> [Accessed: 24.09.14].
- 47 International Committee of Medical Journal Editors. *Recommendations*. [Online] <http://www.icmje.org/recommendations/> [Accessed: 02.10.14].
- 48 Schulz, K.F., Altman, D.G., Moher, D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.
- 49 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].
- 50 National Audit Office. *Evaluation in Government*. [Online] <http://www.nao.org.uk/report/evaluation-government/> [Accessed: 10.10.14].
- 51 Dhaliwal, I., Tulloch, C. From Research to Policy: Using Evidence from Impact Evaluations to Inform Development Policy. [Online] p. 7. <http://www.povertyactionlab.org/publication/research-policy> [Accessed: 24.09.14].
- 52 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).
- 53 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).
- 54 Wager, E., Middleton, P. (2008) Technical editing of research reports in biomedical journals. *The Cochrane Library*.
- 55 Hartley, J., Sydes, M., Blurton, A. (1996) Obtaining information accurately and quickly: Are structured abstracts more efficient? *Journal of Information Science*. 22 (5).
- 56 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].
- 57 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].
- 58 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).
- 59 Pitkin, R.M., Branagan, M.A., Burmeister, L.F. (1999) Accuracy of Data in Abstracts of Published Research Articles. *The Journal of the American Medical Association*. 281 (12).
- 60 Ward, L.G., Kendrach, M.G., Price, S.O. (2004) Accuracy of abstracts for original research articles in pharmacy journals. *The Annals of Pharmacotherapy*. 38 (7-8).
- 61 Grant Lewis, S. (2014) Director, Plan for 4 Learning Portal, UNESCO. Interview with Fiennes, C., September 2014.

- 62 Agricultural Information Management Standards. (2014) *Latest AGROVOC release: July 2014*. [Online] <http://aims.fao.org/standards/agrovoc/about> [Accessed: 24.09.14].
- 63 Schulz, K.F., Chalmers, I., Grimes, D.A., Altman, D.G. (1994) Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *Journal of the American Medical Association*. 272 (2).
- 64 Moher, D., Schulz, K.F., Simera, I., Altman, D.G. (2010) Guidance for Developers of Health Research Reporting Guidelines. *PLoS Medicine*. 7(2).
- 65 Altman, D. (2014) Professor and Director of The Centre for Statistics in Medicine at Oxford University. Interview with Fiennes, C., October 2014.
- 66 BBC News. (2013) *Drug firm Roche pledges greater access to trials data*. [Online] <http://www.bbc.co.uk/news/health-21595895> [Accessed: 24.09.14].
- 67 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].
- 68 CONSORT. *CONSORT 2010 checklist of information to include when reporting a randomised trial*. [Online] <http://www.consort-statement.org/download/Media/Default/Downloads/CONSORT%202010%20Checklist.doc> [Accessed 24.09.14].
- 69 National Offender Management Service. (2013) Analytical Summary: Intermediate outcomes of arts projects: a rapid evidence assessment. [Online] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254450/Intermediate-outcomes-of-arts-projects.pdf [Accessed: 24.09.14].
- 70 Great Britain. Youth Justice Board. (2013) *Effective Practice Identification and Dissemination*. [Online] <http://www.justice.gov.uk/downloads/about/yjb/effective-practice-identification.pdf> [Accessed: 24.09.14].
- 71 *Ibid.*, p. 14.
- 72 *Ibid.*, p. 15.
- 73 Home Office/Women's Aid. *Expect Respect: A Toolkit for addressing Teenage Relationship Abuse*. [Online] http://www.justice.gov.uk/downloads/youth-justice/effective-practice-library/Home_Office_teen-abuse-toolkit1.pdf [Accessed: 24.09.14].
- 74 Reading Youth Offending Service (2014) Rapid English Project: Effective Practice Submission 2014. [Online] <http://www.justice.gov.uk/downloads/youth-justice/effective-practice-library/reading-yos.pdf> [Accessed: 24.09.14].
- 75 College Of Policing. *Policing and Crime Reduction Research Map*. [Online] <http://www.college.police.uk/en/researchmap.htm> [Accessed: 24.09.14].
- 76 Justice Center. *What Works in Reentry Clearinghouse*. [Online] <http://whatworks.csgjusticecenter.org/> [Accessed: 24.09.14].
- 77 Handel, A.E., Patel, S.V., Pakpoor, J., Ebers, G.C., Goldacre, B., Ramagopalan, S.V. (2012) High reprint orders in medical journals and pharmaceutical industry funding: case-control study *British Medical Journal*. 344 (6).
- 78 Centre For Research On Families And Relationships [Online] <http://www.crfr.ac.uk/> [Accessed: 24.09.14].
- 79 Center for Problem-Oriented Policing. *Assaults in and Around Bars 2nd Edition*. [Online] <http://www.popcenter.org/problems/assaultsinbars/> [Accessed: 02.10.14].
- 80 Cardiff University. *My Cardiff – Professor Jonathan Shepherd* [Online] <http://www.cardiff.ac.uk/125/getinvolved/mycardiff/100408.html> [Accessed: 24.09.14].