Commentary: Randomized trials of controversial social interventions: slow progress in 50 years

Ben Goldacre

Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK. E-mail: ben.goldacre@lshtm.ac.uk

The attached paper—a study of caning schoolboys—is remarkable in many respects. In particular, it illustrates battles that are ongoing even half a century later around the ethics, feasibility, desirability and political acceptability of randomized trials (RCTs) outside medicine.

For many working in medical research, it can seem peculiar that randomized trials are resisted in other fields. There are frequently situations in education, policing, criminal justice, employment and other areas of policy where RCTs could be practical and informative, especially: where there is arbitrary variation in the use of a particular intervention; where there are broadly reliable and valid outcomes to be measured; and where strong claims for effectiveness are being made. Trials outside medicine have, however, been slow to come. The prospect of an RCT on beating children provides a good—if extreme, and rhetorically dangerous—opportunity to explore the broader objections to RCTs outside medicine.

First, it is important to note that the attached paper was not a randomized controlled trial of caning. This was not for want of trying. In fact the work by Palmer on caning is mentioned twice, in passing, by Archie Cochrane in his 1972 monograph *Effectiveness and Efficiency*,1 one of the founding documents of evidence-based medicine (it is also discussed at some length in Cochrane’s autobiography).2 Initially he uses the paper to illustrate the weaknesses of data from research into the effectiveness of interventions where participants are not randomized:

Observational evidence is never very satisfactory. An example comes from Palmer’s work in my own unit...He took very detailed smoking histories from all the boys in a secondary modern school on two occasions, one year apart. He then obtained a list of all the boys who had been caned for smoking during that period. He was thus able to compare ‘change in smoking habit’ in caned and uncaned boys. The results appear at first very striking. Those caned increased their cigarette consumption significantly more than those who were not caned, but when one thinks about it the results do not tell us anything at all. They are equally compatible with caning increasing, decreasing, or having no effect on cigarette consumption. Observational evidence is clearly better than opinion, but it is thoroughly unsatisfactory.

He goes on to point out that all effectiveness data in medicine were in a similarly poor state until the 1950s, except for those treatments whose effects were so dramatic that no trials were needed. The revolution in the quality of medical evidence was driven largely by Austin Bradford Hill and his randomized trial of streptomycin. But it was clear even in 1972 that Cochrane regards the adoption of RCTs in medicine as only the first step: ‘His ideas [Bradford Hill’s] have only penetrated a small way into medicine, and they still have to revolutionise sociology, education, and penology. Each generation will, I hope respect him more’.

Progress has been slow, and worse in some countries than in others. A recent Cabinet Office paper3 (of which I am a co-author) sets out the basics of randomized trials for a lay audience of interested politicians and civil servants, and works through the benefits and disadvantages, before giving responses to some common objections and misunderstandings. The cultural challenges around conducting randomized trials are not limited to medicine, however, and examples of shortcomings and barriers throughout this paper are taken from both medicine and policy.

A key barrier to conducting randomized trials, for example, is excessive certainty among practitioners. In the world of policy, this can be especially pressing, as politicians may not wish to have good quality evidence on whether their preferred intervention really does achieve its stated outcome. In my experience, politicians can be surprisingly open to the suggestion; but nonetheless, the movement for more randomized trials in government has so far focused on uncontroversial areas, such as randomizing the content of reminder letters from Her Majesty’s Revenue and Customs (HMRC), allowing the methods to be normalized and capacity to grow.
Spurious certainty, however, is not a problem unique to trials on policy issues. The Corticosteroid Randomization After Significant Head Injury (CRASH) trial—of steroids for head injury—was notoriously resisted by clinicians and ethics committees, on the grounds that steroids were already proven to be effective, and also that unconscious patients could not give consent to participate in research. This last objection left open the alarming prospect of very slow progress for much of emergency medicine. When the trial was finally conducted, it reported that steroids—a widely used intervention—actively increased a patient’s chances of dying.6 Similarly, the Scared Straight programme was widely used in schools, with good intentions: children at high risk of subsequent offending were taken into prisons to be shown the consequences of a life of crime, and observational data showed that participants were much less likely to offend in future. Unfortunately, these data were hopelessly confounded, and when a randomized trial was finally conducted, it was found that participation in Scared Straight actively increased children’s chances of subsequent offending.5

So how might one think through the ethical and practical issues around a modern randomized trial on caning, or any other form of punishment? Medical researchers may feel uneasy about any research that delivers punishment—or any deliberately unpleasant intervention—to participants, especially without their consent. But that is a separate issue to whether such interventions are already being delivered anyway. Corporal punishment of children continues to remain in widespread use around the world. The Global Initiative To End All Corporal Punishment Of Children maintains a regular survey, and reports that only 43 out of 198 UN member nations have protected children by law from all corporal punishment (including at home, whereas 122 have prohibited corporal punishment in schools).6 It is permitted in the home, with certain restrictions, in most African and Asian nations, the US, Canada and the UK.

Despite corporal punishment being so widespread, and therefore implicitly regarded as an ethically acceptable intervention by many, there is little good quality evidence to inform decisions about its benefits and harms. Observational research suggests that corporal punishment for children is actively harmful, even on the outcomes where benefit is anticipated by those using it; all 27 studies in a recent systematic review, for example, reported increased aggression among children receiving physical punishment.7 However, such evidence is clearly vulnerable to both confounding and reverse causality, as children who are prone to aggression may also be more likely to be punished.

A 2002 paper in the Canadian Medical Association Journal attempts to summarize the ethical and evidential issues around corporal punishment for paediatricians—who are likely to be asked for parenting advice—and includes one telling comment: ‘As much as possible today, we are all looking for hard evidence on which to base any opinion or statement. But there isn’t the kind of evidence we would like on this issue because it doesn’t lend itself to randomized trials.’8

It is worth exploring whether this last claim is true, on ethical or practical grounds. Corporal punishment is widespread. Society has not regarded it as unacceptable enough to ban. It seeks to have a positive impact, and a randomized trial comparing caning against some other punishment intervention would almost certainly be feasible. Many of the key positive outcomes that caning seeks to elicit are quantifiable (in the case of smoking cessation for the Palmer paper, they certainly were). The intervention is delivered to individuals, and could be randomized at individual level or, if delivered in schools, then in clusters, with whole schools randomized to cane or not cane, since resentment on grounds of fairness is likely (unless the control intervention is widely perceived as equivalently unpleasant). Blinding would be impossible, but this is the case with many randomized trials of social, psychological, educational and public health interventions. Longer-term negative outcomes could be followed up as with observational research on corporal punishment, assuming that the period over which it was randomly assigned was long enough to plausibly have a long-term impact. Inclusion criteria could stipulate that only those children already exposed to the practice of caning were admitted into such a trial.

There are naturally concerns about any trial where the intervention is either unpleasant or dangerous, especially where there is no consent, and there is already a grim history of such activity in medical research. The Red Wing experiments in the 1940s9 used a strangulation cuff around the neck to investigate why pilots in high-speed aircraft were losing consciousness, and whether prodromal features could be identified, in order to help the war effort: participants were prisoners and patients with schizophrenia. A 1958 study10 involved shutting 2- and 5-year-old children in refrigerators to investigate their escape strategies, leading to changes in fridge design that have saved many lives,11 but the descriptions of the children’s reactions in the original research are disturbing (‘about half were upset but could be comforted easily, and a small group (11%) required some help to become calm’; at follow-up 8 months later ‘a number of children still talked about the tests, some with pleasure, a few with resentment’).

In both these cases, however, participants were randomized to unpleasant or dangerous experiences that they themselves would not normally have encountered. Trials of punishment are somewhat different, since participants are already exposed to the unpleasant intervention.
Furthermore, without good quality research into the impact of punishments, there is a risk that progress in rehabilitation and prevention of re-offending, for example, might suffer; and also that recipients might experience an unpleasant intervention needlessly, without clear evidence of benefit either for themselves or society.

There is already a small literature of randomized trials comparing sanctions for those convicted of crimes. The challenges encountered by trialists in this field mirror those in medical research, such as excessive certainty among practitioners and practitioner preferences that over-ride randomization. In an RCT from 1974, judges were directed to randomly allocate fines and probation to convicted drunk drivers: in more than half the cases the judges subverted the randomization process, mostly in the direction of more lenient sentences.

In the field of criminal justice, there is already a literature on how to work around practitioners breaking randomization on account of their own preferences. For example, Weisburd has suggested that militaristic hierarchical agencies, such as the police, are likely to be less vulnerable to such sabotage, since rigid organizational structures make protocol compliance more likely. He has also suggested that, for ‘sanctioning experiments’, trials are more likely to run smoothly when the punishments being newly introduced are perceived as more lenient than prevailing norms, rather than equivalent but different, or more harsh.

In effect Weisburd suggests that trialists, where they are perceived to be introducing a new intervention, should try to ensure that the new intervention is more palatable to practitioners than the old one, in order to make the process of random assignment itself more palatable. This is paralleled in a common ethical (and, perhaps, rhetorical) strategy for those conducting trials that seek to explore the impact of practices regarded as unethical, harmful or unpleasant: we randomize an intervention to reduce the intervention that is perceived as negative, and measure the impact of this. (Such approach is arguably implicit in trials that seek to assess the impact of poor diet, smoking and poor medication compliance, by using interventions to improve diet, reduce smoking and improve compliance).

In this vein, a randomized trial on the California prison population compared longer and shorter sentences, by randomizing some prisoners to early release; through this, researchers helped avoid triggering any anxiety among practitioners that prisoners had received a longer sentence simply by virtue of participating in an experiment (although thousands of offenders did, nonetheless, still receive longer prison sentences than they might otherwise have had, on account of random allocation).

Similarly, a 2013 meta-analysis of research on corporal punishment for children states that ‘Although randomized control trials can be used to study the effect of reducing physical punishment... they cannot be used to study the effect of imposing such punishment because it would be unethical to assign children to a group receiving painful treatment when research suggests that such pain poses harm not outweighed by potential benefit’. And so, despite their objection to the arbitrary deployment of physical punishment to children on random allocation, they cite—with approval—one RCT of an intervention to reduce parents’ physical punishment of their children: even though the children in this trial were assigned to receive either violence, or less violence, by simple random allocation.

Perhaps the arbitrary nature of randomization is so harsh that we sometimes need to draw a discreet veil across it. But random allocation of an intervention is an irreducible reality of randomized trials, and it should not be shied away from. Participants are randomly assigned to one intervention or another for good reason, on the grounds that both interventions have been deemed acceptable in terms of risk and ethics, and that we do not know which is best able to achieve the interventions’ stated objectives.

In medicine, this bitter pill has been swallowed, albeit after some significant battles. Sadly, simple opportunities for randomized trials on policy, criminal justice, education and more have been left undone, whether through a lack of will, misplaced ethical concerns or a lack of insight into the need for better evidence. Drug Testing and Treatment Orders were introduced, as an alternative to custodial sentences, with the intention that they should reduce subsequent offending and drug use; but no randomized trial was ever conducted, only a pilot study focusing largely on the practicalities of delivering the new service. The Blueprint drugs education programme was introduced, with the intention that it should reduce drug use among children; but the official evaluation—at a cost of several million pounds—simply consisted of following up 23 participating schools and 6 ‘control’ schools, selected after the evaluation period was completed.

It is frustrating to realise how slow progress has been. In his autobiography, Cochrane describes his own efforts—many of them provocative, humorous and confrontational—to stimulate more RCTs in medicine during the 1960s and 1970s. Less well known are his efforts to promote randomized trials on punishment during the same period. At one point, he describes his experience of after seeing arbitrary variation at his local magistrates court.

What I found was rather like general practice, with a wide variety of ‘treatments’ being given for what seemed to me identical ‘diagnoses’. For instance, a boy of 10 caught breaking and entering could be given
either a warning or probation, or his parents fined; although just occasionally he might be referred to his headmaster for summary punishment. There seemed little in the way of rhyme or reason in the way such punishment was awarded, and I discussed the problem privately with a number of magistrates I knew. I suggested to them the value of controlled trials, but was horrified when they reacted like elderly physicians and headmasters. They too suffered from the god complex.

Nothing in this commentary should be taken as a defence of corporal punishment; for a randomized trial of corporal punishment, the issues are more complex. I would personally find it impossible to work on a trial that involved hitting children, but only because I believe that hitting children is foul. For those who are happy to hit children—and there are, sadly, a great many—their reasons for avoiding randomization are the same as those used by doctors, politicians and others to dodge good quality evidence on the true impact of their interventions, for decades past, and doubtless for many more decades to come.

Towards the end of Effectiveness and Efficiency, Cochrane discusses his own department’s efforts to move on from Palmer’s study and conduct a robust randomized trial of caning. His experience will read as fresh to anyone on from Palmer’s study and conduct a robust randomized Cochrane discusses his own department’s efforts to move on from Palmer’s study and conduct a robust randomized.

References


