



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Stewart-Brown, S., et al

Article Title: Should randomised controlled trials be the "gold standard" for research on preventive interventions for children?

Year of publication: 2011

Link to published article : <http://dx.doi.org/10.1108/17466661111190929>

Publisher statement: Stewart-Brown, S. L., et al. (2011). Should randomised controlled trials be the "gold standard" for research on preventive interventions for children?. *Journal of Children's Services*, 6(4), pp. 228-235.

Should randomised controlled trials be the ‘gold standard’ for research on preventive interventions for children

Sarah Stewart-Brown*

Professor of Public Health, University of Warwick,

Rebecca Anthony

Research fellow, Swansea University

Lynsey Wilson,

Research Fellow, Swansea University

Sarah Winstanley

Research Fellow, Swansea University

Douglas Simkiss

Associate Professor Warwick University

Nigel Stallard

Professor of Statistics Warwick University

Helen Snooks

Professor, Deputy Director Clinical Trials Unit, Swansea University

*Lead author contact details:

Warwick Medical School

University of Warwick

Coventry CV4 7AL

024 765 74510

sarah.stewart-brown@warwick.ac.uk

Should randomised controlled trials be the ‘gold standard’ for research on preventive interventions for children.

Abstract (152 words)

Randomised Controlled Trials (RCTs) have been offered a privileged position in terms of the evidence base for preventive interventions for children, but there are practical and theoretical issues that challenge this research methodology. Well recognised issues include the impossibility of blinding participants, the problem of identifying a pre-eminent outcome measure for complex interventions, and problems with limiting access to equivalent interventions in real world settings. A further theoretical problem is the exclusion from RCTs of families who are most ready to change, resulting in an apparently reduced level of intervention effectiveness. Qualitative evidence from a recent RCT suggests that this problem could be operating in some prevention trials. Increasing sample sizes can overcome some of these problems, but the cost of the necessarily huge trials becomes disproportionate to the low cost of preventive interview. Given the limitations on RCTs in preventive settings, their privileged position in terms of research evidence maybe undeserved.

Summary of policy and practice implications (135)

Policy makers and commissioners value certainty and RCTs are considered most likely to provide that certainty. In times of scarce resources, interventions without an RCT evidence base are an easy target for cuts and/or decisions not to invest. In the context of preventive services for children this view is debatable. Because social, educational and public health interventions are complex and are offered in the context of a complex world, it is simplistic to believe that one research approach can provide all the necessary information on what works. A more sophisticated understanding of RCTs and complex interventions suggest that we should rebalance the hierarchy of evidence and give more credence to good studies of other designs. This might make policy makers and commissioners lives more complex and less certain, but it might improve outcomes for children.

Category 2 Methodological Issues and Developments

Word Count 3684

Keywords Methodology, randomised controlled trials, preventive interventions, children

Background

The randomised controlled trial occupies a privileged position in health care research, at the top of the hierarchy of evidence, but within the fields of public health, social care and education there is debate about whether this approach deserves such privilege. Views range from those which acknowledge that RCTs are just not possible in some situations, but should always be done when they are (Bonell et al., 2011) to those who claim they are rarely helpful. Amongst the latter are academics working in health promotion who have declared that in their field RCTs are, in most cases, ‘inappropriate, misleading and unnecessarily expensive’ (WHO, 1998).

There are many good reasons for the privileged position of RCTs in health care research. Evaluation studies are prone to bias because practitioners and professionals are usually keen to believe that what they do is useful and clients and patients are often keen to please and will say they are better when they are not. Observational studies are criticised because groups are rarely balanced in terms of symptom severity, age, or other important participant characteristics, leading to difficulties in interpretation. These biases are well documented, have occurred often in evaluation studies and have misled practitioners, commissioners and policy makers. They are effectively dealt with by double blind randomisation and on this the advocates of RCTs rest their case. This reasoning lay behind the decision of the team which undertook the research for Graham Allen’s review *Early Intervention: The Next Steps* (Allen, 2011) that good quality meant at least **one** RCT or **two** quasi experimental studies. It also certainly informed the National Academy of Parenting’s criteria for kite marking parenting programmes (NAPR, 2011) which also gives more quality points to programmes with an RCT evidence base. These two initiatives are proving influential in cash strapped local and health authorities, who are now cutting funding for, or refusing to invest in programmes that do not appear on kite marked or recommended lists.

So why do people in some fields feel so strongly that RCTs are inappropriate? There are several issues, some of which have been previously rehearsed and some of which are only now beginning to enter the debate.

The impossibility of blinding

The first issue is widely recognised but remains important. It relates to the type of intervention. RCTs are at their best for studies of pills which can be made to look indistinguishable from an inactive

placebo control. It is relatively easy in this situation to undertake a genuine double blind study in which neither clinician or patient knows whether they are taking the active ingredient or not. For social, educational or public health interventions, as well as some health care interventions like surgery, this is, of course, not possible. It is possible to offer the control group an alternative intervention, but such trials cannot show the intervention in question to be better than nothing or current practice. Participants still know which intervention they are getting and can pre-judge which is likely to be better. It is also possible, in theory, to keep the research team blind to the intervention. In practice however, during data collection participants often spontaneously reveal their group even when asked not to. When they have just had an interesting experience, participants are understandably keen to tell researchers about it. So, in this way, it is very difficult for social or behavioural intervention trials to meet the full rigour available for RCTs of pharmaceutical interventions.

Selection of outcome measures.

Statistical significance is based on the likelihood of an outcome having occurred by chance. Conventionally, a chance of less than one in twenty has been accepted as unlikely. So to show an intervention has worked the difference between control and intervention groups needs to be big enough to be likely to have occurred by chance less than one in twenty times. This means that the trial result needs to depend on change in one outcome, because if a high number of outcomes are measured, one is likely to reach statistical significance by chance. If two or three are analysed the conventional level of significance needs to be raised and the sample size increased. So trials need to define, a priori, a primary outcome, and it is on the characteristics of this outcome that the necessary sample size is calculated.

Complex interventions, particularly those involving an element of personal development like parenting programmes, may impact on families in many ways. The outcome of most interest to policy makers from parenting programmes has, to date, been a reduction in levels of problem behaviours in children, but parenting programmes can produce many other beneficial outcomes. For example parents understanding of their children may change, they may feel more confident in themselves, they may start to see the negative effect of hostile or derogatory comments on their child's self esteem and stop doing this, their relationship with their child may start to feel less fraught, or their relationship with their spouse may improve. All of these could be beneficial to children's social and emotional development and future mental health in the longer term without necessarily impacting on the child's

behaviour. This is particularly an issue in universal level programmes where only a proportion of the children will have problem behaviour at the outset. In this situation some change could occur in almost all families in the intervention arm of the trial, but because families change in different ways the average change on any single measure is very small. Preventive interventions producing small effects in a large population can be very valuable. One mm Hg reduction in blood pressure, although irrelevant to the individual, is well worthwhile at a population level. To identify an effect size change of less than 0.3 trials need to be very large and expensive and numbers increase almost exponentially as detectable effect sizes reduce. The cost of the huge trials necessary to pick up very small change on a single measure in complex interventions can become prohibitive and out of all proportion to the costs of the intervention, which in preventive services, especially at universal level, are necessarily low. The issue of multiple outcomes can be relevant in pharmaceutical trials where unexpected outcomes may occur, but it is particularly pertinent for interventions in complex situations where a multitude of factors influence, in the parenting programme example, parents' capacity to change. Those who study complex systems have identified unpredictable outcomes as one of their key characteristics (Cilliers, 1998). This fundamentally important observation does not seem to have had the influence it deserves on thinking about evaluation of family interventions. It certainly does not sit well with the expectations of clinical trials.

Results of trials based on change in one outcome may therefore miss important effects unless they are very large. Even when they show change on one outcome, however, the result is not always helpful for practitioners or policy makers because it is presented in the form of an average level of change. This can hide great individual level variation. The result shows that the intervention is, on the whole, valuable with regard to impact on that one outcome, but it does not provide information about whom it is most likely to benefit. Sub-group analyses can go some way to investigating differential benefits, but from a statistical point of view, subgroup analysis is fraught with problems (Brookes et al., 2001). Other types of research, for example qualitative approaches, can give this information as well as identifying the variety of ways, both expected and unexpected, in which people can change.

Readiness to change

Problems with blinding and placebo control are well recognised in RCTs of complex interventions and problems with identifying a primary outcome measure are also recognised, but what has been proposed

much less often is that participants in RCTs of preventive interventions may be less likely to benefit than the general population.

Trials of prevention strategies are known to attract atypical participants; they are more likely to be of higher social group and better educated, thus potentially more healthy and less able to benefit than the population as a whole (Britton, 1998). There is, however, another possible reason why randomised trials should recruit those least able to benefit. Most of the interventions in public health, social care or education involve behaviour change. In public health this may be a change in health related lifestyle, in early years it may be a change in approaches to parenting and in educational setting may be a change in pupil or teacher behaviour, for example with regard to discipline or bullying. Because behaviour change is such a major issue for public health and because it is challenging to achieve, it has been the subject of research and theoretical development by psychologists and others. One well respected model of behaviour change is Prochaska and Diclemente's transtheoretical model (Prochaska et al., 1992). This describes seven stages: precontemplation, contemplation, preparation, action, maintenance, termination and relapse. The message for practitioners is that their actions need to be related to their client or patient's stage of readiness to change (Hunt & Pearson, 2001). In precontemplation and contemplation the approach is to motivate change; in preparation to encourage personal choice, ownership and engagement; in action to encourage and support etc., etc.

These important differences are not taken into account in recruitment to RCTs. If they were, a potentially important problem with RCTs in this area of practice, might become manifest. People who are at the point of 'readiness to change', those who are most ready to benefit from an intervention, are unlikely to agree to take part in a study in which they might be randomised not to get the help they feel they need. In clinical terms participants who are 'ready to change' are not in 'equipoise' with regard to the potential value of the intervention, but in clinical trials it is not difficult to control access to the intervention and in this situation those most ready to change may want to join the trial. It is much more difficult to control access to a preventive intervention especially if there are equivalent interventions already established in the community. Even if a parenting intervention is new, parents will have access to a wide range of other parenting support initiatives which they can take up with greater or lesser degrees of enthusiasm. If this theoretical problem is real, those who are most likely to be recruited to preventive, behaviour change intervention trials are those who are least likely to be able to benefit. Because they are not ready to change, they are more likely to drop out of the intervention during the course of the trial, or not to engage fully in the programme. In clinical trial jargon this

would be called recruitment bias and would be held to effect external validity, meaning that the results of the trial cannot be held to be applicable to population groups who were not included in the trial. In clinical trials this is usually groups who are *less* likely to benefit from the intervention, the elderly, people with multiple problems or people who do not speak English (Britton, 1998), but the reverse may be true in preventive interventions. Recruitment bias in favour of those least likely to benefit would lead studies to find smaller effects than could be achieved if the programme were offered to people most likely to benefit. So it does not invalidate research on the programmes which do have trial evidence, it just means that the programme might be much more effective and cost effective than the result suggest.

Is bias with regard to readiness to change just a theoretical issue?

Is there any evidence that these are more than theoretical risks? We have recently completed a trial of the Family Links Nurturing Programme, a ten week, universal level parenting programme that meets criteria for effectiveness (Hutchings et al., 2004) but has not previously been evaluated in an RCT. Our trial was undertaken in deprived areas in South Wales where the programme was available in the context of Flying Start, the Welsh equivalent of Sure Start. The Welsh Assembly Government (WAG) was reluctant to finance a non-‘evidence-based’ programme and support was about to be withdrawn, from a programme which was popular with practitioners. So the WAG and the county councils collectively identified funds to finance the trial.

Any family living in a Flying Start area in four Welsh county councils was eligible to take part in the trial and families were invited through health visitors and Flying Start staff. Although it took longer than planned, we recruited the full number of participants (287) and achieved an exemplary level of 85% follow up overall. However non-attendance and drop out from the intervention was high. Only 50% of those recruited completed more than five sessions.

As part of the study, we interviewed ten parents who declined to take part in the research, but did take part in the programme. We also interviewed twenty two parents who had taken part in both research and programme asking them, amongst other things the reason why they chose to take part.

In the first group, who refused the research, one parent exemplified the readiness to change issue.

‘I was just desperate to go on the course.... ..’ It could have been six months to a year or something before I could have gone on it’ (the waiting list control condition). [ZL000]

In this group almost all parents said they chose to go on the course because they really needed it or because they felt there was a problem with their child(ren).

'I was having a lot of trouble with my boy, he was biting and not listening to me and just being really naughty to be honest' [INT2]

Only one parent said they chose to go because someone else was going or because someone else had put their name down.

'I only took part because I knew the people that were going'. [C270511]

In contrast, in the larger group who had agreed to take part in the trial, only five parents gave problems with their parenting or their children's behaviour as the reason for agreeing to take part:

'It sounded pretty good because I was having a big problem with my daughter at the time; she was really naughty at that time'. [N030510]

'[Child's name] was just being a child and just being a bit of a pain, and she was like, 'well this could help you' so hmmm, I thought, well we'll give it a go and see how it works and anything, hmm anything was an improvement. Anything that would help was gonna be good.' [TN127]

More parents gave general reasons for wanting to do the programme.

'Pretty much the same as I am with anything really. If there's anything out there that can help me do a better job, then I am up for reading, or going, or looking at it, yeah.' [FM120610]

'I think I was just interested to do something new especially as it was related to parenting.' [DS264]

'I don't know, it was just, see what you could learn, 'cause there was an NVQ you could do. So that was it really 'cause they said you could do an NVQ, so I applied.' TN130

Three quarters of parents in this group specifically mentioned helping others and ensuring the programme remained available in the locality as a reason for taking part.

'Because I just thought, you know that kind of programme, especially round this area where there are quite a few problems with children, it's quite handy. It'll be a shame to withdraw a programme when it is needed in the area. So that's obviously why.' CP06012011

About a quarter said they took part in the trial because they were asked to do it.

'I always get on with my health visitor and she asked me to [take part]' [JP311110]

Or because friends were doing it :

'my mate was doing so I thought I'd give it a go.' [TN125]

This is only a small study of one trial, but the results certainly suggest that recruitment bias could be a real issue. The parents we interviewed who refused the research were much more likely to express attitudes in line with readiness to change than were parents who took part in the trial. The latter group were quite likely to have agreed to take part for the benefit of others. These parents, particularly those who did it to support the continued availability of the programme in the local community, were less likely to feel the need to change themselves. In a real world setting, a universal offer of a parenting programme would be unlikely to attract this group and likely to attract those who wanted help with their parenting or their child's behaviour.

What does this mean for evaluating preventive services for children?

Differences between the size of effects observed in RCTs and non randomised controlled studies have in the past been attributed to bias in controlled studies (Mac Lehosé et al., 2000), but if recruitment bias is influential in preventive intervention trials, in the way suggested above, differences might be due to systematic underestimation of effects in randomised controlled trials. This is important for policy making and commissioners. Behaviour change and personal development are difficult to achieve, but small changes at population level can have a very large impact on health and social wellbeing. Very large trials are needed to pick up small effects and trials need to be even larger if they themselves reduce the effectiveness of the intervention. Their costs start to exceed the limits of most sources of research funding and raise questions about the opportunity costs of research.

Consider, in contrast, the findings from relatively inexpensive qualitative studies (Kane et al., 2007). These have reported parents as saying that during the course of a parenting programme they developed insights into the causes of their children's behaviour or relationship difficulties, 'ah ha moments',

entirely consistent with the content of the programmes, together with changes in family wellbeing that followed from these insights and their impact on family interaction and behaviour. These findings prove beyond reasonable doubt that the programme helped some families in the intended way. What they do not do is provide policy makers or commissioners with quantitative estimates of the effect on groups of families as RCTs do. If however, the quantitative estimates are inaccurate the latter may be misled. Consider also multi-method study designs including process, outcome, formative and summative evaluation methods (Snooks et al., 2011). These can describe and cost the impact of highly complex multifaceted interventions so policy makers and commissioners have a clear idea of the changes they might reasonably expect for a certain level of investment.

Many of the issues raised with regard to RCTs have been considered by others and alternative RCT designs have been proposed. Zelen's design (Adamson et al., 2006) in which potential participants are randomised before recruitment gets over the problem of excluding those ready to change from the intervention arm and could therefore be preferable in this context. Potential participants who refuse to take part in the control arm may, however, be different from those who refused to take part in the intervention arm so making the groups unbalanced. The matched pair design (Kratochwill et al., 2004) in which eligible families are first matched on key outcomes and then randomised within each pair, helps to deal with the problem of non-attendance because only the families who turn up to a programme and their matched pair are included in the analyses. The problem is that matching can only be done on a limited number of identifiable variables and may not truly match control and intervention arms. If it was possible to truly match families, observational studies could give results as valid as randomised controlled ones.

If the intervention under study is only available in the context of a trial, as is often the case for new drugs, many of these problems can be avoided, because those most ready to change are likely to 'opt in' on the chance that they may get the intervention. But in the real world of prevention, interventions which seem to be useful to families are made available through third sector organisations or on the internet or through books. It is often not possible to rigidly control these opportunities.

Policy makers and commissioners, like clinicians, value the best quality evidence. The randomised controlled trial has been offered as a method that offers such evidence, but in the context of children's preventive services that belief needs to be qualified with the need for critical appraisal of methods used in the particular context of the study, irrespective of study design. Clinicians in health care settings

who find it difficult to tolerate uncertainty order more investigations to try and reduce uncertainty. In doing so they greatly increase patient care costs and rarely improve outcomes. In the field of prevention, the RCT is in danger of becoming the policy makers' and commissioners' equivalent of the clinical investigation.

Because social, educational and public health interventions are complex and are offered in the context of the real, complex world, it is simplistic to believe that finding out what helps can be achieved with just one research approach. We need to become more sophisticated in evaluating complex interventions, and re-consider the weight given to the RCT. Considering evidence from good studies of whatever design might make policy makers and commissioners lives more complex and less certain but it might improve outcomes for children.

References

Adamson, J., Cockayne, S., Puffer, S., Torgerson, D.J. (2006) Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemporary Clinical Trials* **27**(4) pp.305-19.

Allen, G. (2011). Early Intervention: the Next Steps. An Independent Report to Her Majesty's Government. London: Cabinet Office [online]. Available: <http://www.dwp.gov.uk/docs/early-intervention-next-steps.pdf> (accessed 23rd July 2011)

Bonell, C.P., Hargreaves, J., Cousens, S., Ross, D., Haves, R., Petticrew, M., Kirkwood, B.R. (2011) Alternative to randomisation in the evaluation of public health interventions: design, challenges and solutions. *J Epidemiology and Community Health*, **65** (7) pp. 582-7.

Brookes, S.T., Whitley, E., Peters, T.J., Mulheran, P.A., Egger, M., Davey Smith, G. (2001) Subgroup analysis in randomised controlled trials: quantifying the risk of false positive and false negatives. *Health Technology Assessment*, **5**(33) pp.1-56.

Britton, A., Mckee, M., Black, N., McPerson, K., Sanderson, C., Bain, C. (1998) Choosing between randomised and non randomised studies. *Health Technology Assessment* **2**(13) pp. 1-124.

Cilliers, J. (1998) WHO Regional Office for Europe 1998: Health promotion evaluation: recommendations to policy makers: report of the WHO European Working Group on Health Promotion Evaluation Copenhagen WHO Regional Office. (document EUR/ICP/ IVST 05 01 03)

Cilliers, P. (1998) *Complexity and post modernism*. New York: Routledge.

Elias, M., Zins, J., Graczyk, P., Weissberg, R. (2003) Implementation sustainability and scaling up of social emotional and academic innovations in public schools. *School Psychology Review* **32** (3) pp. 303-319.

Hatch, T. (2000) What does it take to break the mold? Rhetoric and reality in new American schools. *Teachers College Record* **102** (3) pp 561-589.

Hunt, P. & Pearson, D. (2001) Motivating Change. *Nursing Standard* **16** (20) pp. 45-55.

Hutchings, J., Gardner, F. & Lane, E. (2004) Making evidence based interventions work in clinical settings; common and specific therapy factors and implementation fidelity. In Carole Sutton, David Utting and David Farrington (Eds): *Support from the start: Working with young children and their Families to reduce the risks of crime and antisocial behaviour*. London: Department for Education and Skills

Kane, G., Wood, V., Barlow, J. (2007) Parenting programmes: a systematic review and synthesis of qualitative research. *Child: Care Health and Development* **33** (6) pp. 784-793.

Kratochwill, T.R., McDonald, L., Levin, J.R., Bear-Tibbetts, H.Y., Demaray, M.K. (2004) Families and schools together: an experimental analysis of a parent mediated multi-family group program for American Indian children. *J School Psychology* **42** pp.359-383.

National Academy for Parenting Research (NAPR). Parenting Programme Evaluation Tool <http://www.kcl.ac.uk/iop/depts/cap/research/napr/parentingprogrammesevaluationtool.aspx> (accessed 23rd July 2011)

MacLehoise, R.R., Reeves, B., Harvey, I.M., Sheldon, T.A., Russell, I.T., Black, A.M.S. (2000) A systematic review of comparisons of effect sizes derived from randomised and non- randomised studies. *Health Technology Assessment* **4** (34) pp. 1-154.

Prochaska, J.O., DiClemente, C.C., Norcorss, J.C. (1992) In search of how people change. Applications to addictive behaviours. *American Psychologist* **47**(90) pp. 1102-1104.

Snooks, H.A., Bridie, A.E., Cohen, D., Nugent, M., Rapport, F., Skone, J., Meredith, A., Davis, T., O'Sullivan, D. (2011) Costs and effects of a 'healthy living' approach to community development in two deprived communities: findings from a mixed methods study. *BMC Public Health* **11** (1) pp.25.