



## ESSAY

# Good intentions and received wisdom are not good enough: the need for controlled trials in public health

Sally Macintyre

**Correspondence to**

Sally Macintyre, MRC/CSO  
Social and Public Sciences Unit,  
4 Lilybank Gardens, Glasgow  
G12 8RZ, Scotland, UK;  
sally@sphsu.mrc.ac.uk

Accepted 12 November 2010

Published Online First

8 December 2010

**ABSTRACT**

In the 1970s Archie Cochrane noted that many healthcare procedures and forms of organisation lacked evidence of effectiveness and efficiency, and argued for improved methods of evaluation, moving from clinical opinion and observation to randomised controlled trials (RCTs). His arguments gradually became accepted in medicine, but there has been considerable resistance among policymakers and researchers to their application to social and public health interventions. This essay argues that opposition to RCTs in public health is often based on a false distinction between healthcare and community settings, and sometimes on a misunderstanding of the principles of RCTs in health care. It suggests that just as in medicine, good intentions and received wisdom are not a sufficient basis for making public policy and allocating public funds for social or health improvement.

Archie Cochrane's book, 'Effectiveness and efficiency: random reflections on health services'<sup>1</sup> is well known for arguing the case for randomised controlled trials (RCTs) in health care. Antenatal care was one of the fields he singled out:

'this service is basically a multiphasic screening procedure, which by some curious chance, has escaped the critical assessment to which most screening procedures have been subjected in the last few years and there seems no reason why the same approach that has proved so useful elsewhere should not be used here'. (Cochrane, p 66)<sup>1</sup>

I was particularly interested in this comment because at the time I was involved in evaluating a modified schedule of antenatal care, using a before and after case study approach.<sup>2</sup> Our problem was that there was no clear counterfactual with which to compare the new system. The study, which cost a not insignificant amount, was therefore somewhat inconclusive, other than finding that the new system did not seem to be a complete disaster or kill mothers and babies. Opponents could use our findings to suggest the new system was worse, and supporters that it was better, than the old one. We noted that:

'Random allocation to different schedules would have maximised comparability between experimental and control groups, and simultaneous comparisons would have avoided the contaminating effects of secular change on the outcome measures used.' (Hall *et al*, p 115)<sup>2</sup>

The arguments against using an RCT were that it was unethical to experiment on pregnant women and their children, and that antenatal care was

a complex matter and therefore inherently unsuitable for an RCT. Interestingly, it was considered ethically permissible to experiment in a non-randomised way, in the process withholding what was currently considered normal care from pregnant women, and scientifically permissible to introduce and try to evaluate a complex system of care without any controls.

Subsequently, Cochrane's arguments were highly influential among perinatal epidemiologists, and a number of RCTs of maternity care in the UK were undertaken, including:

- ▶ Midwife or general practitioner-led care versus obstetrician-led care<sup>3-5</sup>
- ▶ Traditional versus reduced schedules of antenatal visits<sup>6</sup>
- ▶ Women holding or not holding their own obstetric records<sup>7</sup>
- ▶ Perineal management (restrictive vs liberal use of episiotomy)<sup>8</sup>
- ▶ Enema on admission<sup>9 10</sup>
- ▶ Postnatal support for mothers in disadvantaged inner-city areas.<sup>11</sup>

Many of these trials were initially thought to be impossible, unethical and/or impractical, but they nevertheless happened. They had a number of important features. First, they did not insist on slavish adherence to a standardised protocol. The trials of midwife versus shared care did not involve women being banned from seeing obstetricians or midwives, but rather the comparison of two general policies or principles.<sup>4</sup> Second, they involved multiple components and series of decisions, social interactions and behaviours. Third, many evaluated activities spanning many months. Fourth, they involved multiple outcomes as well as multiple inputs (eg, the antenatal care trials typically looked at antenatal admissions, non-attendances, numbers of antenatal visits, antenatal diagnoses, inductions of labour, satisfaction with care, etc). Fifth, they involved evaluations of processes and acceptability as well as outcomes. Sixth, randomisation meant that results were less likely to be biased by self-selection of patients or professionals, and were therefore more conclusive than previous case study approaches. However these (fairly typical) characteristics of healthcare RCTs often seem to be misunderstood by opponents of public health RCTs.

**PUBLIC HEALTH RESEARCH**

From the late 1990s a number of commentaries in the UK pointed to the lack of robust evidence to support social and public health policies in the UK.<sup>12 13</sup> One report noted:

'Although there is often evidence on the scientific justification for action and for some specific interventions, there is generally little evidence about the cost-effectiveness of public health and preventative policies or their practical implementation.' (Wanless, p 5)<sup>14</sup>

Why so little evidence? In the UK, many evaluations focus on inputs, throughputs and customer or professional satisfaction rather than on outcomes. Second, few policies or programmes are implemented in ways that facilitate robust evaluation of outcomes (eg, they often lack baseline data, comparison groups, clear objectives, and/or statistical power). Third, there is a general reluctance in the UK to subject social or public health policies to RCTs. There have been far fewer controlled studies in the UK than in the USA, which raises issues of generalisability across contexts (eg, the 'nurse family partnership' developed and extensively studied in the USA<sup>15</sup> may not be relevant in the UK where deprived first-time mothers already have access to antenatal and postnatal support via the NHS). These issues together militate against the production of robust evidence about effectiveness and efficiency.

Recently in the UK, members of parliament criticised policy-makers' approaches to evaluation:

'All too often Governments rush in with insufficient thought, do not collect adequate data at the beginning about the health of the population which will be affected by the policies, do not have clear objectives, make numerous changes to the policy and its objectives and do not maintain the policy long enough to know whether it has worked.' (House of Commons Health Committee, p 5)<sup>16</sup>

They were particularly critical of officials' reactions to suggestions that controlled trials should be used. Many of these responses suggested fundamental misunderstandings about the principles and practice of RCTs. For example, one senior civil servant rejected suggestions that a 'healthy towns' initiative be subjected to a controlled trial by saying:

'it would challenge any academic to come up with a randomised town' (House of Commons Health Committee, p 7)<sup>17</sup>

implying that randomisation means choosing one town at random and treating it as representative of all towns. In the next section I illustrate and discuss such common misconceptions, in particular that community trials are essentially different from healthcare trials.

### **SOME OBJECTIONS TO RCTs IN PUBLIC HEALTH** **Communities differ whereas individuals do not**

'A community intervention with a matched community control is far more feasible (than an RCT) but still challenging because, unlike individuals, communities vary widely in characteristics related to exposure to risk.' (Moller, pp 2–3)<sup>18</sup>

'It is unlikely that any complex intervention will work for everyone.'<sup>19</sup>

However, individuals also vary enormously in exposure to risk and response to interventions, which is why one needs sufficient sample sizes to capture variations in both experimental and comparison groups.

### **Communities and organisations are complex whereas individuals are not**

'Communities clearly differ. They also have attributes that are not reducible to those of individual members. These include cultures (eg, religious beliefs), structures (eg, employment patterns, and relationships (eg, contact between ethnic groups)).' (Pawson, p 52)<sup>20</sup>

However, individuals are also complex organisms, with characteristics that are greater than the sum of their component chemical parts. Also, all interventions in healthcare settings, even if of a highly standardised drug, involve social settings and social interactions, power dynamics, local cultures, motivations, behaviours, etc. This is certainly the case of perinatal trials mentioned above; the idea that a trial of midwife-led versus obstetrician-led antenatal care, involving interactions between pregnant women and healthcare professionals over several months, is not complex seems bizarre. Indeed as Oakley has suggested:

'It can be argued that the greater the complexity of the setting into which an intervention is introduced, the more need there is to ensure that factors that may affect the outcomes of interest are equally distributed between intervention and control groups.' (Oakley *et al*, p 175)<sup>21</sup>

### **Social/public health interventions, unlike surgical or drug interventions, do not do harm**

One reason for objections to RCTs of public health policies or programmes is the belief that, unlike surgical or pharmaceutical interventions, they are unlikely to do harm. This view privileges social and public health actions, and assumes that the plausibility of potential benefit is a sufficient basis for action.<sup>22</sup>

However, there are numerous examples of apparently plausible policies or programmes having no benefit or actually being harmful. For example, in the UK the risk of death from fire is associated with low socioeconomic status because of social differences in risk factors for fires and the ownership of smoke alarms, and the risk of death in a house fire is three times higher in homes without smoke alarms; so it is plausible that giving free fire alarms to deprived households might reduce excess fire deaths among them. However, an RCT found that giving out free alarms in a deprived community did not reduce injuries from fire, because few alarms had been installed or maintained. It concluded that issuing free smoke alarms may waste resources and be of little benefit unless alarm installation and maintenance is assured.<sup>23</sup>

The Scared Straight programme in the USA brings juvenile delinquents into prisons to meet life prisoners, who attempt to deter them from a life of crime. Criminologists and many stakeholders, including the general public, have been positive about the programme, which has *prima facie* plausibility. However, none of seven RCTs showed any benefit, and a meta-analysis showed that recidivism rates were higher among the experimental group.<sup>24</sup> If there had not been RCTs of fire alarms and Scared Straight, we might continue to implement these programmes on the basis of their plausibility.

### **Community trials are impossible**

It is often simply stated as a matter of agreed fact that RCTs are impossible in community settings. However, a wide range of community-based RCTs has already been undertaken or planned, including:

- ▶ The Mexican universal health insurance programme<sup>25</sup>
- ▶ The effects of hand washing on child health in squatter settlements in Karachi<sup>26</sup>
- ▶ The deterrent effect of police raids on crack houses<sup>27</sup>
- ▶ The effectiveness of toughened glassware in reducing injuries in bars<sup>28</sup>
- ▶ School breakfast clubs<sup>29</sup>
- ▶ Out-of-home day care for disadvantaged families<sup>30</sup>
- ▶ Community-level interventions to address social and structural determinants of health in 40 areas in London.<sup>31</sup>

**Public health trials are more difficult than healthcare trials**

Some opponents of public health RCTs seem to assume that RCTs in health care are easy to set up and do not involve problematical ethical issues (in contrast to what is required for community-based trials). However, similar difficulties have been overcome in healthcare RCTs; for example, for temporal lobe epilepsy,<sup>32</sup> arthroscopic surgery for osteoarthritis of the knee,<sup>33</sup> and transplantation of embryonic dopamine neurons for severe Parkinson's disease<sup>34</sup> (the latter two involving sham surgery in the control group). The trialists surmounted many of the difficulties often regarded as too difficult in community evaluations.

**RCTs require one to adhere strictly to protocol**

One objection to public health RCTs is that they force one to stick rigidly to protocol:

'A population outcome is the goal. Such a broad focus means that strict control of intervention, subject and analysis required for a true experiment or clinical trial is impossible. This send shudders down the spines of those brought up in the empirical tradition.' (Moller, p 2)<sup>18</sup>

However, healthcare RCTs do not require one to have such strict standardisation. The trial of perineal management did not insist that women in the experimental arm all had episiotomies and nobody in the control arm did,<sup>8</sup> and the RCT of surgical versus medical treatment for epilepsy did not closely control what the surgical or medical treatments were.<sup>32</sup> Intention-to-treat analysis of trial outcomes is recommended precisely because not everyone in the intervention group will actually receive the intervention or receive it in the same way, and not everyone in the control group will be deprived of the intervention.<sup>35</sup>

The term 'controlled' may lead to some confusion here, it sometimes being interpreted as meaning rigid fidelity to the programme, rather than some sort of comparison with a counterfactual such as a comparison group, which is what would happen if the intervention had not taken place.

**RCTs involve a single experimental and comparison unit**

'Even where matching populations have been found, a final comparison comes down to a single case with control design and a critical reviewer can easily dismiss results.' (Moller, p 3)<sup>18</sup>

This seems to assume that public health RCTs would only select one intervention and one comparison group; (see the earlier comment about 'a random town')<sup>17</sup> however, as has been pointed out:

'It is common to see reports of community intervention trials in which one intervention community is compared with one control committee. This is equivalent to a clinical trial with one patient in each treatment group.' (Hayes and Bennett, p 323)<sup>36</sup>

A key element of healthcare RCTs is that they involve sufficient numbers in both experimental and control groups to rule out the role of chance, and in many cases in public health it would similarly be possible to have sufficient numbers of experimental and control units (eg, schools, neighbourhoods or towns).<sup>36</sup>

**RCTs are expensive, and if they do not demonstrate a positive difference are failures**

'There are examples of failed, expensive trials. For example no intervention effect was observed among heavy smokers, the

primary target population of the COMMIT trial. Similarly CART demonstrated limited positive results, with most cancer related behaviour showing no intervention effects...' (Sanson-Fisher, p 158)<sup>37</sup>

This seems to conflate the notion of a failed trial (ie, one that is badly designed or fails to recruit sufficient numbers) with a failed policy or intervention (ie, one that shows modest or no benefit or actual harm). Both the fire alarm<sup>23</sup> and the Scared Straight trials<sup>24</sup> illustrate that RCTs can be successful and cost effective even if they show reasonably conclusively that an intervention has no, modest, or adverse effects.

**RCTs do not have long enough follow-up**

'In population health research many outcomes of interest are far into the future ... the practical difficulties in maintaining prospective randomisation for prolonged periods across entire populations are substantial' (Sanson-Fisher *et al*, p 157)<sup>37</sup>

However, there is no intrinsic reason for the issue of length of follow-up to be any different between an RCT and an uncontrolled evaluation. (The High Scope Perry RCT of early childhood intervention has now followed up the participants for 40 years).<sup>38</sup> Such comments may be confusing the long-term follow-up of outcomes with the long-term maintenance of the exposure to which people or communities were randomly assigned.

**RCTs have poor external validity**

'a disadvantage to using RCTs in population health research is the lack of generalisability, or low external validity' (Schweinhart *et al*, p 157)<sup>38</sup>

It is difficult to understand why non-RCTs in community settings, for example the numerous (and sometimes very expensive) non-randomised evaluations in the UK of area-based initiatives such as Sure Start<sup>39</sup> or Health Action Zones<sup>40</sup> should be regarded as any more generalisable than RCTs.

**RCTs are unethical**

It is often seen as unethical to conduct policy RCTs because they withhold potential benefits from the control group.<sup>22</sup> However, this assumes that well-intentioned policies will be beneficial, and that the direct or opportunity costs of implementing a policy are of no ethical concern. As was pointed out in the early days of healthcare RCTs, it seems perverse to see it as ethical to give or withhold programmes of unknown benefit to 100% of the population, but not to 50%.<sup>41 42</sup> This is particularly the case when the intervention has to be rationed anyway (eg, not every community can have a Sure Start local programme or be a Health Action Zone).

As members of parliament have pointed out:

'All the reforms we have discussed are experiments on the public and can be as damaging (in terms of unintended effects and opportunity cost) as unevaluated new drugs or surgical procedures. Such wanton large-scale experimentation is unethical, and needs to be superseded by more rigid rigorous culture of piloting, evaluating and using the results to inform policy.' (House of Commons Health Committee, p 66)<sup>16</sup>

**CONCLUSION**

It appears that many objections to social and public health RCTs are based on false comparisons with healthcare RCTs, the latter being seen to be simple, and to have standardised exposures and



outcomes. Many of these objections were raised earlier about healthcare RCTs, but have been overcome in the healthcare arena, often with considerable ingenuity, and with recognition of the complexity of humans and their contexts.

Do objections to social and public health RCTs matter? I believe they do, and that Cochrane would agree. For example, the government refused to allow a randomised evaluation of the Sure Start programme in England, which has led to considerable problems in interpreting the results, because although the researchers tried to find matching areas not receiving the intervention, and to control for any obvious socioeconomic and demographic differences between intervention and control areas, differences between areas receiving and not receiving the intervention, rather than the intervention itself, may influence any observed differences in outcome.<sup>16 43</sup> The results of an expensive evaluation of an expensive intervention can therefore be contestable rather than conclusive.

With rare exceptions, such as when a new universal policy such as banning smokeless fuel or indoor smoking is introduced and can be evaluated using interrupted time-series methods,<sup>44</sup> RCTs are both more possible than many objectors think, and more conclusive about the benefits and cost effectiveness of (usually publicly funded) policies and interventions.

So I think we should be braver and much more creative in arguing (whether with politicians, public health practitioners, research funders, potential recipients or ethics committees) for RCTs. As for the privileging of social and public health as not requiring robust randomised evidence, I believe that Cochrane's comments about psychiatry are equally applicable to public health:

'I cannot agree that colleagues, however distinguished, intelligent and hard-working, and who obviously believe they are doing good, should have a blank cheque to encourage the use of (XX) without bothering to measure the benefit and cost of what they are doing.'  
(Cochrane, p 59)<sup>1</sup>

**Acknowledgements** This is an abridged version of the Cochrane lecture given at the Society for Social Medicine annual conference in 2009. The author is grateful to Lyndal Bond, Iain Chalmers, Matt Egan, Mark Petticrew and Helen Roberts for their contributions to her thinking on this topic, in some cases over many years, and to the Society for Social Medicine committee for inviting her to give this lecture.

**Funding** This work is supported by the UK Medical Research Council, wbs U.1300.00.006.

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; not externally peer reviewed.

## REFERENCES

- Cochrane A. *Effectiveness and efficiency: random reflections on health services*. London: Nuffield Provincial Hospitals Trust, 1971.
- Hall M, Macintyre S, Porter M. *Antenatal care assessed: a case study of an innovation in Aberdeen*. Aberdeen: Aberdeen University Press, 1985.
- Hundley VA, Cruickshank FM, Lang GD, et al. Midwife managed delivery unit: a randomised controlled comparison with consultant led care. *BMJ* 1994;**309**:1400–4.
- Tucker JS, Hall MH, Howie PW, et al. Should obstetricians see women with normal pregnancies? A multicentre randomised controlled trial of routine antenatal care by general practitioners and midwives compared with shared care led by obstetricians. *BMJ* 1996;**312**:554–9.
- Turnbull D, Holmes A, Shields N, et al. Randomised, controlled trial of efficacy of midwife-managed care. *Lancet* 1996;**348**:213–18.
- Sikorski J, Wilson J, Clement S, et al. A randomised controlled trial comparing two schedules of antenatal visits: the antenatal care project. *BMJ* 1996;**312**:546–53.
- Elbourne D, Richardson M, Chalmers I, et al. The Newbury maternity care study: a randomised controlled trial to assess a policy of women holding their own obstetric records. *Br J Obstet Gynaecol* 1987;**94**:612–19.
- Sleep J, Grant A, Garcia J, et al. West Berkshire perineal management trial. *BMJ (Clin Res Ed)* 1984;**289**:587–90.
- Romney ML, Gordon H. Is your enema really necessary? *BMJ (Clin Res Ed)* 1981;**282**:1269–71.
- Drayton S, Rees C. They know what they're doing. *Nurs Mirror* 1984;**159**:4–8.
- Wiggins M, Oakley A, Roberts I, et al. Postnatal support for mothers living in disadvantaged inner city areas: a randomised controlled trial. *J Epidemiol Community Health* 2005;**59**:288–95.
- Macintyre S, Chalmers I, Horton R, et al. Using evidence to inform health policy: case study. *BMJ* 2001;**322**:222–5.
- Milward L, Kelly M, Nutbeam D. *Public health intervention research: the evidence*. London: Health Development Agency, 2001.
- Wanless D. *Securing good health for the whole population. Final report*. London: HM Treasury, 2004.
- Olds DL, Kitzman H, Hanks C, et al. Effects of nurse home visiting on maternal and child functioning: age-9 follow-up of a randomized trial. *Pediatrics* 2007;**120**:e832–45.
- House of Commons Health Committee. *Health inequalities (third report of session 2008–9)*. London: The Stationery Office, 2008.
- House of Commons Health Committee. *Oral evidence taken before the health committee Wednesday 19th of November 2008 (Rt Hon Alan Johnson MP and Mr Hugh Taylor CB)*. London: The Stationery Office, 2008.
- Moller J. Reconsidering community based interventions. *Inj Prev* 2004;**10**:2–3.
- James A. Could go further (rapid response to Craig et al 2008). *BMJ* 2008;**337**.
- Pawson R, Tilley N. *Realistic evaluation*. London: Sage Publications, 1997.
- Oakley A, Strange V, Toroyan T, et al. Using random allocation to evaluate social interventions: three recent UK examples. *Ann Am Acad Pol Soc Sci* 2003;**589**:170–89.
- Skrabaneck P. Why is preventive medicine exempted from ethical constraints? *J Med Ethics* 1990;**16**:187–90.
- DiGiuseppi C, Roberts I, Wade A, et al. Incidence of fires and related injuries after giving out free smoke alarms: cluster randomised controlled trial. *BMJ* 2002;**325**:995.
- Petrosino A, Turpin-Petrosino C, Buehler J. 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database Syst Rev* 2002;(2):CD002796.
- King G, Gakidou E, Imai K, et al. Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. *Lancet* 2009;**373**:1447–54.
- Luby SP, Agboatwalla M, Feikin DR, et al. Effect of handwashing on child health: a randomised controlled trial. *Lancet* 2005;**366**:225–33.
- Sherman LW, Rogan DP, Edwards T, et al. Deterrent effects of police raids on crack houses: a randomized, controlled experiment. *Justice Q* 1995;**12**:756–81.
- Warburton AL, Shepherd JP. Effectiveness of toughened glassware in terms of reducing injury in bars: a randomised controlled trial. *Inj Prev* 2000;**6**:36–40.
- Shemilt I, Harvey I, Shephstone L, et al. A national evaluation of school breakfast clubs: evidence from a cluster randomized controlled trial and an observational analysis. *Child Care Health Dev* 2004;**30**:413–27.
- Toroyan T, Roberts I, Oakley A, et al. Effectiveness of out-of-home day care for disadvantaged families: randomised controlled trial. *BMJ* 2003;**327**:906.
- Wall M, Hayes RJ, Moore D, et al. Evaluation of community level interventions to address social and structural determinants of health: a cluster randomised controlled trial. *BMC Public Health* 2009;**9**:207.
- Wiebe S, Blume WT, Girvin JP, et al. A randomized controlled trial of surgery for temporal-lobe epilepsy. *N Engl J Med* 2001;**345**:311–18.
- Moseley JB, O'Malley K, Peterson NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;**347**:81–8.
- Freed CR, Greene PE, Breeze RE, et al. Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med* 2001;**344**:710–19.
- Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992;**21**:837–41.
- Hayes R, Bennett S. Simple sample size calculations for cluster randomised trials. *Int J Epidemiol* 1999;**28**:319–26.
- Sanson-Fisher RW, Bonevski B, Green L, et al. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med* 2007;**33**:155–61.
- Schweinhart L, Montie J, Xiang Z, et al. *Lifetime effects: the High/Scope Perry Preschool Study through age 40*. Ypsilanti, MI: High/Scope Press, 2005.
- Belsky J, Melhuish E, Barnes J, et al. Effects of Sure Start local programmes on children and families: early findings from a quasi-experimental, cross sectional study. *BMJ* 2006;**332**.
- Barnes M, Bauld L, Benzeval M, et al, eds. *Health Action Zones: partnerships for health equity*. London: Routledge, 2005.
- Smithells R. Iatrogenic hazards and their effects. *Postgrad Med J* 1975;**15**:126–7.
- Chalmers I. Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date, replicable evaluations. *Ann Am Acad Pol Soc Sci* 2003;**589**:22–40.
- Rutter M. Is sure start an effective preventive intervention? *Child Adolesc Mental Health* 2006;**11**:135–41.
- Craig P, Dieppe P, Macintyre S, et al. *Developing and evaluating complex interventions: new guidance*. London: Medical Research Council, 2008.