

SOME measures are just plain common sense. If drug use is rife in prisons, deter offenders with random drug tests. If teenagers are straying into delinquency, give them a taste of what prison life is like to shock them into cleaning up their acts. And since newly qualified drivers cause many deaths and injuries, all children should be taught how to drive responsibly at school.

Policies like these are bound to make a difference, aren't they? Well, yes, but perhaps not quite in the way you might imagine. All are examples of bright ideas that have backfired in the real world. Delinquents given tours of prison are usually more likely to reoffend. Random drug tests seem to increase heroin use.

In the 21st century, you might expect governments to be pragmatic about achieving their aims, to do what works. This means

basing policies on hard evidence rather than on assumptions or ideology. Yet this seldom happens. Even when policies are tested before being rolled out to an entire area or country, the methods used to evaluate their effectiveness are often worse than useless.

But now more and more researchers are calling for social strategies to be assessed by the gold standard for establishing the effectiveness of any intervention: randomised controlled trials, long used to find out if new drugs are effective and acceptably safe.

"Relative to where we've got to in medicine, it's a disaster," says Lawrence Sherman, a criminologist at the University of Cambridge and the Jerry Lee Center of Criminology at the University of Pennsylvania, Philadelphia. "Anything involving the treatment of people from education to social services and crime prevention is mostly terra incognita."

Randomised controlled trials have become the norm in testing medical treatments over the past 50 years. Half the patients in a trial get a new therapy, and half get a conventional one or placebo – the control group. Patients are randomly assigned to each group to ensure any difference in outcomes reflects differences in the effectiveness of the treatments, not differences between patients.

Once a therapy's effectiveness is established, its cost-effectiveness can be compared with the alternatives. "It's really a step towards a rational society and a fulfilment of the 18th-century Enlightenment, but it's just taken root in a professional culture of medicine much more firmly than it has in social policy," says Sherman.

Assessing social policies using randomised controlled trials did start to take off in the US from the 1960s to 1980s. But the practice has

Science rules OK!

Why do we still run societies in such a crazy way?
Hazel Muir explores the rational alternative



declined, partly because policy makers became disenchanted when the trials did not endorse their brainwaves, according to Sheila Bird, a statistician at the Medical Research Council Biostatistics Unit in Cambridge, UK.

In some cases at least, this may have been because the trials were so small that any differences between the treatment and control groups were due to chance. In medicine, too, many early drug trials were too small to produce conclusive results.

For example, say you want to give smokers free nicotine patches because your groundwork suggests this might mean in five years' time 40 per cent rather than 30 per cent of them will have stopped smoking. A randomised trial to assess this effect would have to assign at least 800 people to either free patches or none to make sure that the trial has a good chance, about 8 chances

in 10, of detecting that improvement.

The fact that randomised trials have to be large to produce reliable results has led to the common misperception that they are always laborious, slow and expensive. Because politicians are keen to look confident and decisive, they instead tend to prefer simple evaluations of policies that are "doomed to success", according to Laurence Moore, a social scientist at Cardiff University, UK.

"They're almost designed to show that the idea is a good idea," he says. "It's well intentioned, but politicians are not open to the idea that a rigorous evaluation might help them get things better. Rigorous evaluations are perceived as threatening rather than supportive of better policy."

The upshot is that many policies are rolled out across entire countries without credible evidence that they do any good. Bird gives the

example of drug treatment and testing orders in the UK, which began in 1999 as an alternative to prison sentences for drug-addicted criminals. Because addicts often steal to feed their heroin habit, the idea is that getting them off drugs will reduce crime.

The treatment orders, lasting up to three years, oblige offenders to attend rehab, as well as having regular drug tests and reviews. But the evidence base for the orders is "vacuous", Bird says. Small pilot studies compared the reoffending rate for addicts treated in the community to the average for all released prisoners, rather than the reoffending rates for prisoners who were addicts.

"It is very remiss that we do not have a solid evidence base. If we knew how much these orders could reduce reoffending rates, judges might see them as a win-win situation," says Bird. "If that is not the case and the orders ▶



“It’s a step towards a rational society and a fulfilment of the 18th-century Enlightenment”

don’t work well, we need to find out why.”

Another barrier to rigorous trials is that social policies are sometimes assumed to be incapable of doing any harm. Take the case of mandatory drug testing in UK prisons, which began in 1995 and still operates in England and Wales. Prisoners are selected at random to give a urine sample that is tested for illegal drugs. Punishments for a positive test have included extra days added to a sentence, or loss of prison earnings.

But a 1995 pilot study in eight prisons rang alarm bells: the number of positive tests for opiates or tranquillisers shot up from 4.1 to 7.4 per cent during the study (*BMJ*, vol 312, p 1411). The thinking is that some prisoners switched from cannabis to heroin because cannabis use can be detected in urine for up to three weeks afterwards, while heroin use is detectable only for two or three days. “There was a very clear case that it might cause harm and that it was setting up new markets for heroin inside the prisons,” says Bird.

Likewise, teaching schoolchildren to drive responsibly is backfiring. A 2001 study reviewed large randomised trials in the US, Australia and New Zealand and showed there was no evidence that school-based driving lessons reduced road crashes. If anything, more teens had accidents because they passed their driving tests earlier. And since the 1970s, “Scared Straight” programmes in the US have tried to shock delinquent youths into mending their ways by giving them tours of prisons and discussions with remorseful “lifers”. But randomised trials show that these programmes do more harm than good.

Such policy disasters keep occurring partly because the very people who ought to be pushing for proper trials before policies are widely implemented often oppose them. Many social scientists wrongly think of

rigorous trials as meticulously controlled experiments that are difficult to apply in the real world, Moore says. They might say it’s impossible to do a controlled trial of a new teaching method, for instance, because no two teachers are the same and classrooms have different mixes of abilities, so you can’t standardise the experiment.

Stubborn ideology

In fact, it’s perfectly possible to do large “cluster” randomised trials that overcome this problem by comparing groups rather than individuals. For instance, pupils taught by the same teacher should be grouped together when evaluating different teaching methods. Moore’s team has used a cluster trial to test an anti-smoking project that cunningly exploits peer pressure, for instance. More than 10,000 pupils aged 12 to 13 in 59 English and Welsh schools took part.

In half the schools, researchers polled all the pupils to identify the most influential ones in their year. Typically, the top-scoring 30 pupils from each school were then invited to a two-day training course about the harmful effects of smoking and how they could encourage their friends not to take it up. Moore’s team found that pupils in these schools were 25 per cent less likely, on average, to start smoking over the following two years than those in schools where pupils only had conventional anti-smoking lessons (*The Lancet*, vol 317, p 1595).

This is not to say that randomised controlled trials are always appropriate. “There are evangelists for trials who basically say that everything should be trialled,” Moore says. “But the majority of published trials in the area of health behaviour change have been done on interventions that are never going to

make a difference.” Running trials is wasteful unless there is good reason to think an intervention is likely to work.

There is another barrier to evidence-based policy – just because the evidence is there does not compel anyone to use it. School driving lessons, Scared Straight programmes and many other policies continue despite solid evidence that they do not work. “What’s not there is the automatic transfer of knowledge to practice,” says Sherman. He points out that we have known since the 1850s that doctors can save lives by washing their hands more, yet in intensive care units the failure of doctors to wash their hands is still a major cause of death.

“Why would we expect elected officials to immediately change policy based on randomised trial research if we can’t get doctors to wash their hands?” Sherman believes that policy documents, and journalists, should be clearer about the quality of evidence for new policies to help the good ones make their mark.

Researchers and policy makers also need to talk to each other. “There are people in ivory towers doing lots of trials of hopeless interventions and then there are people in the real world actually trying to tackle the social realities of health inequalities – there’s a complete disjunction between the two,” Moore says. He adds that because policy makers were involved in his peer-led anti-smoking trials, the successful intervention got the go-ahead to be rolled out to all schools in Wales even before the results were published.

Another barrier to evidence-based policy can be stubborn ideology, according to opponents of controversial abstinence-only sex education programmes that have consumed almost \$1.7 billion of federal and state funding in the US since 1996. In addition, a third of US aid for HIV prevention programmes in the developing world must be spent on programmes promoting abstinence instead of contraceptives.

Among other requirements, the programmes must teach “that sexual activity outside of the context of marriage is likely to have harmful psychological and physical effects”. A 2004 report commissioned by a Democratic congressman concluded that four-fifths of the curricula contained false or misleading information, such as hugely exaggerating the risk of pregnancy or HIV transmission when condoms are used.

“The origin of this programme was not in science or research by any means, but in an ultra-conservative, ultra-religious ideology,” says James Wagoner, president of Advocates for Youth, a non-profit organisation in Washington DC that champions informed decision-making about



“You could see the abstinence-only movement as the sexual health equivalent of creationism”

sexual behaviour. “You could almost see the abstinence-only movement as the sexual health equivalent of creationism.”

Several studies, including a Congress-funded randomised controlled trial involving more than 2000 teenagers, showed the abstinence-only programmes were no more likely than conventional sex education to prevent or delay teenagers having sex, or reduce their number of sexual partners. Yet Congress continues to fund the programmes. Peer-reviewed studies of more than a dozen well-considered programmes for scientific sex education show these programmes can both make teenagers delay having sex and increase contraceptive use if they do have sex: “But how many of these would be eligible for federal funding? Zero,” Wagoner says.

Wagoner is optimistic, however, that the abstinence-only programmes will wane now that a growing number of states are rejecting the federal funds. “We’re now up to 16 states that have slapped a return-to-sender on these programmes,” he says. “That’s unparalleled – it’s a rarity that states refuse federal dollars. I think there’s a climate in the country now that is much more inclined to move forward and jettison this kind of nonsense.”

Everyone agrees that the appetite for evidence-based policy is growing. In the UK, the Economic and Social Research Council is working with several government departments to create a policy trials service over the next three years. It aims to boost the quality and impact of policy evaluation. “That is suggesting there is a demand to do more

trials, and there will be a unit in the UK that can make sure that when trials are done, they’re done well,” says Moore.

Criminology now has a whole sub-field devoted to randomised experiments, along with a dedicated journal, the *Journal of Experimental Criminology*, which has been going since 2005. Networking is improving too. Sherman is encouraged that senior police officers from the UK, US, India, Sweden and Australia will join researchers this summer at a conference in Cambridge, UK, on evidence-based policing. “You wouldn’t have even imagined this ten years ago,” he says.

Randomised controlled trials are finding solid evidence for the benefits of what’s called “restorative justice”, in which criminals meet their victims and discuss the impact of their crime. With his colleague Heather Strang, Sherman has reviewed many rigorous trials of restorative justice in North America, the UK, Australia and New Zealand and shown that with certain exceptions, it reduces repeat offending rates for violent criminals. It helps victims too, reducing symptoms of post-traumatic stress disorder in victims of violent crime by 40 per cent in one study in London.

“Restorative justice can work, but more research is needed to specify the conditions under which it works best,” says Sherman. “Soon there will be an extremely robust body of evidence about restorative justice – not answering all our questions, but demonstrating that it’s possible to take an innovation like restorative justice and subject it to rigorous research and improvement before just plunging right into it.”

Bird says we not only need more randomised trials, we need much bigger ones. “My intention for the criminal justice system in the 21st century is that it shouldn’t be making the errors we once made in medicine by doing small randomised studies. We should jump into the right ballpark to start with,” she says. “It took 50 years for medicine to get it right. Let’s have criminal justice get things right from day one.”

Social scientists and policy makers alike should recognise that large randomised trials are not necessarily expensive in an era when so much data about people is stored on computers anyway, Moore points out. For instance, it would cost next to nothing to arrange a huge randomised trial of classes teaching good-parenting skills in any country where databases routinely record children’s school exam results, teen pregnancies and any later criminal convictions.

In any case, while rigorous trials might sometimes be costly, the alternative is worse. Blindly rolling out policies that might silently fuel crime, ill-health or social inequality has got to be more expensive in the long run. ●