

When are randomised trials unnecessary?

Picking signal from noise

The relation between a treatment and its effect is sometimes so dramatic that bias can be ruled out as an explanation. **Paul Glasziou and colleagues** suggest how to determine when observations speak for themselves

Our knowledge of the effects of treatments comes from various sources ranging from personal clinical experience to carefully controlled trials. Although we are often wary of inferring the effects of treatments from evidence other than that from randomised controlled trials, we are all familiar with examples of situations in which confident inferences about treatments have been based on other kinds of evidence. For example, the first case series of puerperal sepsis treated with sulphonamides^{1 2} provided striking evidence that these new drugs had important benefits: although some patients died, the proportions surviving serious infections (puerperal sepsis, meningitis, etc) were substantially greater than predictions based on previous experience. These dramatic effects of sulphonamides were not observed in other conditions, however, and carefully controlled trials were required to distinguish confidently between moderate treatment effects and no material effects.²

To help us think about the circumstances in which randomised trials are unnecessary, we sought help³ in compiling a list of examples of treatments whose effects had been widely accepted on the basis of evidence from case series or non-randomised cohorts (box). We have considered three present day examples in more detail to help illustrate the basis for our conclusions:

Mother's kiss technique—A child presented to a clinic with a plastic bead lodged high in one nostril. The general practitioner asked the nurse for forceps, but she asked him whether he had thought of trying the mother's kiss technique.⁴ This entailed occluding the unblocked nostril while the mother blew into the child's mouth. The bead was easily dislodged and retrieved in this way, and mother and child were both delighted

Laser treatment of portwine stains—Portwine stains are present at birth. They can enlarge and change colour during childhood but are stable thereafter. The effects of a single laser treatment take about three months to be seen (after some initial inflammation has settled).⁵ Multiple treatments may be needed for optimum effects, but improvement is common after a single treatment

Fundoplication for heartburn—One option for patients with reflux causing heartburn is fundoplication, where the upper part of the greater curve of the stomach is wrapped around the oesophagus to mechanically prevent reflux. One of the early case series of laparoscopic Nissen's fundoplication showed dramatic results on both symptoms and objective findings.⁶ For example, 95% had abnormal pH and manometry results before

Paul Glasziou, professor, Centre for Evidence-Based Medicine, Department of Primary Health Care, University of Oxford, Oxford OX3 7LF

Iain Chalmers, editor, James Lind Library, James Lind Initiative, Oxford OX2 7LG

Michael Rawlins, honorary professor, London School of Hygiene and Tropical Medicine, London WC1E 7HT

Peter McCulloch, reader, Nuffield Department of Surgery, John Radcliffe Hospital, Oxford OX3 9DU
Correspondence to: P Glasziou paul.glasziou@dphpc.ox.ac.uk

surgery compared with 5% afterwards. In subsequent long term follow-up studies of symptoms, reflux was abolished in a similar percentage of patients and overall antacid use was reduced fivefold⁷

Prognosis: the background noise

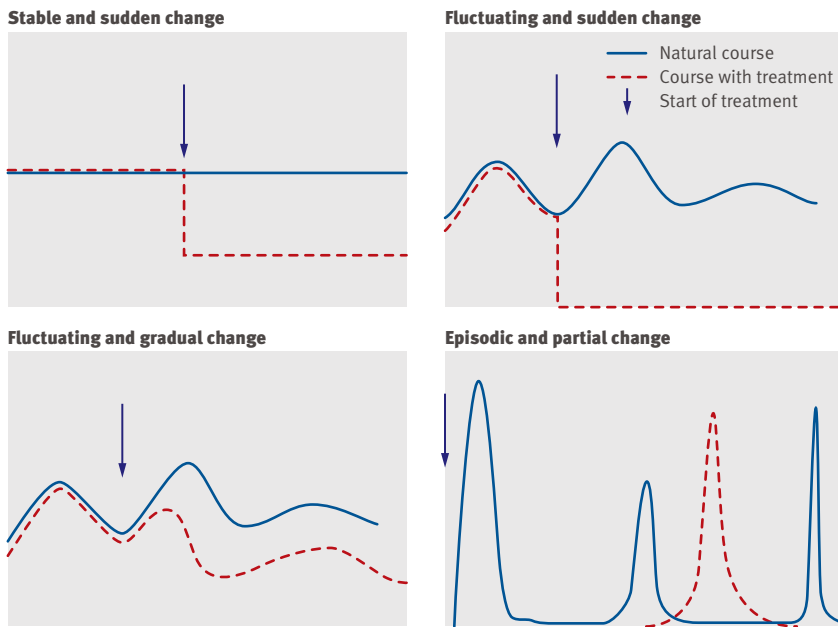
The first step in assessing a treatment effect is to look at the background noise. From the evidence of one case should we now adopt the mother's kiss technique as first line treatment for other children with nasal foreign bodies? The mother's kiss technique is a clear example of a rapid effect (seconds) in a stable condition. The size of the effect can be calculated as a relative rate: it takes less than 10 seconds to see the effect of the mother's kiss, compared with the hours beforehand (for 2 hours this is 720 periods of 10 seconds) with no movement of the foreign body. So the rate ratio of removal for a single case is:

$$\begin{aligned} \text{Rate ratio} &= \text{rate of progression during treatment} / \text{rate of progression during non-treatment} \\ &= (1/1) / (0.5/720) = 1440 \end{aligned}$$

(Note that we replaced the 0 cure rate with 0.5, a half correction that allows for a rate between 0 and 1, providing a more robust estimate and avoiding

Some historical examples of treatments with dramatic effects

- Insulin for diabetes^{w1}
- Blood transfusion for severe haemorrhagic shock^{w2}
- Sulphanilimide for puerperal sepsis^{w3}
- Streptomycin for tuberculous meningitis^{w4}
- Defibrillation for ventricular fibrillation^{w5}
- Closed reduction and splinting for fracture of long bones with displacement
- Salicin for acute rheumatism^{w6}
- Neostigmine for myasthenia gravis^{w7}
- Tracheostomy for tracheal obstruction^{w8}
- Suturing for repairing large wounds
- Drainage for pain associated with abscesses
- Pressure or suturing for arresting haemorrhage
- Ether for anaesthesia
- One way valve or underwater seal drainage for pneumothorax and haemothorax^{w9}
- Phototherapy for skin tuberculosis^{w10}
- Combination chemotherapy with cisplatin, vinblastine, and bleomycin for disseminated testicular cancer



Different degrees of signal: noise in single patient

division by zero. Note also that an occasional spontaneous cure—for example, from sneezing—would still result in a large rate ratio.)

This relative rate represents a large signal to noise ratio and is also significant ($P < 0.01$) because, under the null hypothesis, the chance that the cure occurred in the treatment period used out of 720 possible periods is $1/720$. However, the apparent effect is likely to be an overestimate as we are likely to note and report the successes rather than the failures.⁸ To generalise, we need data derived from several carefully assembled case series.⁹ A search yields only one report of a case series, in which the mother's kiss was successful in 15 out of 19 children.⁴ We think this is sufficient evidence to recommend use in practice without randomised trials. However, it clearly fails sometimes and it would be worth documenting why and doing randomised trials comparing techniques that are unlikely to have greatly different effects.

With stable or progressive conditions, rapid effects of treatment are easy to demonstrate—for example, the effects of removing a cataract on vision or of cholinesterase inhibitors for organophosphate poisoning. Many surgical procedures also fall into this category—for example, drainage of a pleural effusion or pneumothorax, any operation to arrest haemorrhage, repair of a hernia, and incision of a perianal haematoma.

To generalise further, we can try to predict the outcome (current prognosis) without treatment. This can be clear and easy for stable or progressive conditions but can be highly unpredictable in fluctuating or probabilistic conditions. Prognosis can be classified from most to least predictable as:

- Stable*—for example, portwine stain, lodged foreign body
- Progressive*—for example, otosclerotic deafness, cataract, many cancers
- Spontaneously remitting*—for example, colds, viral rashes

The recent examples of hormone replacement therapy and β carotene show how evidence from sources other than randomised trials can lead us badly astray

Fluctuating—for example, rheumatoid arthritis, eczema, and depression

Episodic—for example, migraine, asthma

Probabilistic (a possible future event)—for example, stroke.

Picking up the signal from the background noise

However, not all treatment effects in stable conditions are so easy to demonstrate. The prognosis and the treatment effect interact as noise and signal, and the ease of identification of treatment effects depends on the signal to noise ratio (figure). The effects of hearing aids on social functioning and quality of life, for example, are less immediate and predictable than the effect on hearing itself and are detected most reliably by parallel group randomised trials.¹⁰ Gradual or delayed effects, such as improvement in speech after hearing aids, are usually less obvious than immediate effects.

Consider the example of laser treatment for a portwine stain—a more gradual effect but with a stable condition. If the portwine lesion has been unchanged for 10 years and then improves three months after treatment, then the relative rate of improvement in three month intervals is:

$$\text{Rate ratio} = \text{rate during treatment} / \text{rate during non-treatment} = (1/3) / (0.5/120) = 80$$

(again using a half correction for the stable period).

This is relatively convincing, although any remaining doubt about whether the portwine stain had really changed could be resolved (without randomisation) by taking a photograph every three months over the 10 years and asking blinded examiners to select the post-treatment photograph with the best appearance. Similar examples include Paré's assessment, nearly four centuries ago, of the effects of a treatment for burns,¹¹ and Williams and colleagues' treatment of three yellow nails with topical vitamin E and three control nails with vehicle only.¹²

Such proof becomes more difficult when the condition is fluctuating or intermittent—for example, with inhaled corticosteroids for asthma or antidepressants to prevent migraine. Here, individual cases and experience are liable to be misleading as there is as much noise as signal. In these circumstances, we usually need randomisation and other measures to reduce biases in order to distinguish treatment effects from the effects of biases, unless the effect is very large, as in laparoscopic Nissen's fundoplication (our third example). Here the relative rate of abnormal manometry results before and after the fundoplication was $95\%/5\% = 20$ (exact numbers give a relative rate of 22 with 95% confidence interval 9.8 to 49). Long term follow-up several years after surgery shows a lasting reduction in the percentage of patients with reflux symptoms from 100% to around 5%,^{13 14} and a fivefold reduction in use of antacids.⁷ Given the size and rapidity of the change in these subjective and objective measures, fundoplication obviously works. Whether it works better than drugs or alternative operations is a different question, and one for which randomised trials are needed.

How large an estimate of a treatment effect is large enough?

How much difference between the treatment outcome (signal) and the natural outcome (noise) is enough? We know that confounding is common and often not obvious; indeed, this was the basis for inventing randomised trials. There is no unambiguous answer to this question: it will always remain a matter of judgment. However, it may be worth trying to develop a rule of thumb, such as that by which we conventionally accept $P=0.05$ as significant.

We suggest that a sufficiently extreme difference between the outcome ranges for treated and untreated patients might be defined by two rules: (a) that the conventionally calculated probability of the two groups of observations coming from the same population should be less than 0.01 and (b) that the estimate of the treatment effect (rate ratio) should be large. In our examples it was at least 20. Simulations have suggested that implausibly large associations, both between treatment and confounding factor and between confounding factor and outcome, are generally required to explain risks beyond relative rates of 5-10.^{15 16} One empirical study that compared randomly selected control groups in multicentre trials also found that, while modest confounding is very likely, such extremes are unlikely.¹⁷ We therefore suggest that rate ratios beyond 10 are highly likely to reflect real treatment effects, even if confounding factors associated with the treatment may have contributed to the size of the observed associations. However, further empirical work in other datasets is clearly desirable.

Possible additional evidence criteria

We have focused on the signal to noise ratio as a measure of the strength of the treatment effect. However, other factors are relevant in making inferences about treatment effects. Austin Bradford Hill proposed a list of factors strengthening confidence in inferences.¹⁸ The table shows how the causation guidelines he proposed might be applied to our three examples. The elements that are common to all three examples are the temporal relation, the strength of the relation (the effect size), and the plausibility, whereas several other criteria are not fulfilled.

Discussion and conclusions

Confident inferences about the effects of treatment are justified in several situations in which treatment effects are unlikely to be confused with the effects of biases. These include, in particular, mechanical interventions such as surgical procedures, where there is a rapid response on a stable background. A probabilistic approach based on the signal to noise ratio may help to define such situations. The strength of relation has already been incorporated in the process of grading evidence suggested by the GRADE collaboration.¹⁹

The recent examples of hormone replacement therapy and β carotene show how evidence from sources other than randomised trials can lead us badly astray. In both these cases, however, the signal to noise ratio was modest, with relative risks of around 2 (or 0.5,

Assessment of three treatments by Bradford Hill guidelines¹⁸

Criteria*	Mother's kiss for nasal object	Laser for portwine lesion	Fundoplication for heartburn
Temporal relation (treatment precedes effect)	Yes	Yes	Yes
Strength of relation (eg correlation or relative risk)	Very strong	Very strong	Very strong
Plausibility (based on current understanding of disease mechanism)	Yes	Yes	Yes
Consistency (across settings and methods)	No	Yes	Yes
Coherence (with knowledge of related treatments)	No	No	Yes
Dose-response relation	No	Yes	No
Specificity (treatment causes the effect and little else)	No	No	No

*We have omitted experiment because this is the topic of our discussion.

depending on which way the comparison is framed). Relative risks of this order would not meet our requirements for judging a treatment effect to be dramatic.

Although parallel group randomised trials will remain the principal means of obtaining reliable evidence about the average effects of treatments when effects are moderate, our three examples show some circumstances in which treatment effects can be inferred from well designed case series⁹ and non-randomised cohort studies. Further research is required to obtain better estimates of the plausible limits of bias in different types of non-randomised study designs.²⁰

We thank Abdelhamid Attia, Benjamin Djulbegovic, Hywel Williams, Jan Vandenbroucke, Olaf Dekkers, Dave Sackett, Jonathan Meakins, Ruth Gilbert, Amanda Burls, Ken Fleming, and the members of the Evidence-Based Health Care email list for help with examples and comments on earlier drafts of this paper.

Contributors and sources: All authors have been involved in both clinical trials and clinical practice and the links between these. PG and IC conceived the study; all authors contributed to compiling the examples used for analysis, and development of the concepts and writing of the paper. PG is guarantor.

Competing interests: None declared.

- Colebrook L, Kenny M. Treatment with prontosil of puerperal infections due to haemolytic streptococci. *Lancet* 1936;ii:1319-22.
- Loudon I. *The use of historical controls and concurrent controls to assess the effects of sulphonamides, 1936-1945*. James Lind Library (www.jameslindlibrary.org).
- Djulbegovic B. Non-randomized trials that changed medical practice. www.hsc.usf.edu/bdjulbeg/oncology/NON-RCT-practice-change.htm.
- Botma M, Bader R, Kubba H. A parent's kiss: evaluating an unusual method for removing nasal foreign bodies in children. *J Laryngol Otol* 2000;114:598-600.
- Goh CL. Flashlamp-pumped pulsed dye laser (585 nm) for the treatment of portwine stains: a study of treatment outcome in 94 Asian patients in Singapore. *Singapore Med J* 2000;41:24-8.
- Cuschieri A, Hunter J, Wolfe B, Swanstrom LL, Hutson W. Multicenter prospective evaluation of laparoscopic antireflux surgery. Preliminary report. *Surg Endosc* 1993;7:505-10.
- Bloomston M, Niels W, Rosemurgy AS. Symptoms and antireflux medication use following laparoscopic Nissen fundoplication: outcome at 1 and 4 years. *JLS* 2003;7(3):211-8.
- Vandenbroucke JP. Case reports in an evidence-based world. *J R Soc Med* 1999;92:159-63.
- Jenicek M. *Epidemiology: the logic of modern medicine*. Montreal: Epimed, 1995:136.
- Mulrow CD, Aguilar C, Endicott JE, Tuley MR, Velez R, Charlip WS, et al. Quality-of-life changes and hearing impairment. A randomized trial. *Ann Intern Med* 1990;113:188-94.
- Donaldson IML. *Ambroise Paré's account in the Oeuvres of 1575 of new methods of treating gunshot wounds and burns*. James Lind Library (www.jameslindlibrary.org).
- Williams HC, Buffham R, du Vivier A. Successful use of topical vitamin E solution in the treatment of nail changes in yellow nail syndrome. Full references 1-20 are on bmj.com

SUMMARY POINTS

Some treatments have such dramatic effects that biases can be ruled out without randomised trials

Dramatic effects can be defined by the size of the treatment effect (signal) relative to the expected prognosis (noise)

Real treatment effects are likely if the signal to noise ratio is large (above 10)

Large ratios may be due to the high proportion of patients improved or the rapidity of improvement