Why randomized controlled trials fail but needn't: 1. Failure to gain "coal-face" commitment and to use the uncertainty principle

David L. Sackett

Dr. Sackett is Director of the Trout Research & Education Centre at Irish Lake, Markdale, Ont.

From the underside

My pager went off as I was half-dozing, half-reading the newspaper. Yuri Gagarin, born the same year as me, had just become the first man to orbit the earth; Nelson Mandela, not yet imprisoned, had gone underground; and John F. Kennedy had just signed the bill creating the Peace Corps. An hour later, back on the charity ward and feeling the effects of too much acrid coffee and too little sleep, achalasia hit me when I found the lancet-shaped diplococci in her sputum. It was 3 am and the research protocol, appearing unannounced on our workroom wall, declared that, as the front-line clinician (the British call this "working at the coal face"), I had to enter her into the trial.

She was single, poor, the sole supporter of 3 children and now so sick after 10 days of cough, fever and sputum that she'd risked losing her menial job by leaving it early to struggle through Chicago's filthy March slush to our emergency room. My boss was testing one of the first synthetic penicillins against the then standard penicillin G in patients with pneumococcal pneumonia, and she fit the entry criteria. But she had classic signs of hepatization and already had suffered an episode of the euphoria and cyanosis we'd been taught was characteristic of bacteremia and interlobar spread. The last patient I'd seen this sick from pneumococcal pneumonia, a strapping 18-year-old basketball player, was also the first patient on whom I'd conducted fruitless open-chest cardiac massage.

With the fear, hopelessness and trust I'd come to expect from my patients, she consented at once to take part in the trial. But by the time I completed her entry form I knew what I had to do. Blocking the view of the ward nurse, I took the syringe containing the study drug from the refrigerator, loaded a second syringe with penicillin G and injected her with both.

I have never discussed this decision with anyone, nor admitted it until now. I don't know how many of my fellow house officers did the same thing for their sickest patients. I believe that my action was right in particular, wrong in general (I've never cheated since), and doubly preventable.

Effect

To the extent that other house officers doing their best to care for similar patients responded as I did, patients in both arms of this randomized controlled trial (RCT) would have been given penicillin G. As a result, it was a trial of the effect of *adding* synthetic penicillin *to* penicillin G, rather than the comparison of the 2 drugs that our boss had intended. In the absence of a negative interaction between the 2 penicillins, any lack of efficacy of synthetic penicillin when given alone would have remained undetected. The effect was a "false-equivalence" RCT.

I don't know how often this type of RCT failure occurs, and I've never encountered anyone who does know. Ken Schulz[1] collected anecdotes of attempts to "break the code" to find out what treatment a patient would be assigned to if they entered an RCT, from transilluminating a sealed envelope to rifling through the files of the principal investigator.

Causes

There were 2 underlying causes for the failure (at least in my contribution to it) of this RCT. The first was failure to gain "coal-face" commitment. I was an underling who had never been party to any discussion of the study question and how it would best be answered. It was not "my trial," and I had only a general stake, not a personal one, in guaranteeing that it generated a valid answer. The second cause was failure to use the uncertainty principle. I was certain that my patient needed penicillin G, and my responsibility to her welfare was in direct conflict with my responsibility to the internal validity of the RCT. Moreover, this possibility was neither acknowledged nor accommodated in the protocol posted on the wall. I could not serve both my patient and the RCT, and in serving the former I cheated the latter because this was preferable to serving the latter and thereby cheating the former, my patient. This conflict arose from my "knowing" that my patient needed penicillin G, and such convictions are just as relevant when they are wrong as when they are right.

Principles

Coal-face commitment

Only collaborators and patients who consider an RCT "theirs" should be expected to follow its protocol. Although this principle has to do with human behaviour, not research methods, it is a key determinant of the internal validity and efficiency of every RCT. Given the increasing variety and magnitude of the competing demands placed on front-line clinicians, no one should be surprised to discover these clinicians neglecting any task they deem nonessential. The more detailed the entry form and eligibility criteria for

> The uncertainty principle acknowledges that most clinicians and patients do have hunches about a treatment's effectiveness but that the boundaries ("confidence interval") around their hunches may run all the way from extremely effective (a wonder drug), across zero (ineffective) and into the realm of frank harm.

"somebody else's" RCT, the greater the risk the criteria will be ignored, misunderstood or misapplied by distracted clinicians who regard them as further intrusions into an overfull call schedule.

Uncertainty principle

Patient entry into RCTs should be governed by the uncertainty principle. What is "ethical" in one culture or era may be "unethical" in another; therefore, in this series I will avoid pronouncements on good and evil. However, when the set of ethical principles that are in vogue at a certain place and time impinge on the design and conduct of contemporaneous RCTs, I will examine the evidence on which the principles are based and their effects on RCTs.

> Equipoise is a "state of balance or equilibrium between two alternative therapies" such that "there is no preference between treatments, i.e., it is thought equally likely that treatment A or B will turn out to be superior." Some definitions require each individual clinician and patient to be free of any "hunch" or preference ("theoretical" equipoise), while others permit individual clinicians and patients to have hunches as long as they "recognize that their less-favored treatment is preferred by colleagues whom they consider responsible and competent" ("clinical" equipoise).

In the example I present in this essay, we need to consider both the waning North American principle of "equipoise" and the principle of "uncertainty" that is being increasingly adopted in most other parts of the world.

Equipoise is a "state of balance or equilibrium between two alternative therapies"[2] such that "there is no preference between treatments, i.e., it is thought equally likely that treatment A or B will turn out to be superior. At this point we may be said to be 'agnostic' … we would take odds of 1:1 on a bet."[3] In certain times and places, equipoise has been considered a prerequisite for an ethical RCT. And in some of these times and places, the individual clinician and patient had to be free of any "hunch" or preference ("theoretical" equipoise),[4] while in others, individual clinicians and patients could have a preference as long as they "recognize that their less-favored treatment is preferred by colleagues whom they consider to be responsible and competent"[5] ("clinical" equipoise).

Opponents of the equipoise construct (including me) argue that it has 3 fatal flaws. First, it is incapable of application: equipoise is lost as soon as the first pair of patients given the alternative treatments finishes the trial and the allocation code is broken. Second, it treats hunches (preferences) as point-estimates and ignores the uncertainty with which those hunches are held. Put another way, and linking the constructs of equipoise and uncertainty, equipoise demands a "confidence interval" of zero, whereas uncertainty permits and works with confidence intervals of 50%, 99%, or any other magnitude. Third, and as a result of the first 2, equipoise is almost never possessed by trialists or explored by ethics committees. For example, in the great majority of the more than 200 RCTs in which I have played a role, neither I nor my patients nor my collaborators nor the nonparticipants we encountered from the relevant profession were in equipoise, and our hunches frequently were strong ones (indeed, some poten-

tial clinical collaborators were so convinced that the experimental treatment was efficacious or useless that they refused to have anything to do with a trial of it). It isn't that we have no hunches and are indifferent to the alternative treatments in our RCTs, it's that we are uncertain about whether our hunches are correct.

The "uncertainty" construct rejects the indifference of equipoise and builds on the notion that clinicians and patients are often uncertain whether their hunches about a treatment's effectiveness are true. That is, although their hunch about a specific treatment may be that it is probably effective, the boundaries ("confidence interval") around that hunch may run all the way from extremely effective (a wonder drug), across zero (ineffective) and into the realm of frank harm. When the uncertainty boundaries of a group of clinicians and patients include or cross zero (such that they recognize that the treatment they prefer might, in fact, be useless or even harmful), it is time for a trial, and that trial is ethical. Similarly, in equivalence trials such as the one that opened this essay, uncertainty exists as long as the confidence interval around the hunch that one of the treatments is actually superior includes or crosses zero. As I'll show you in the following section, this uncertainty principle helps decision-making not only by individual clinicians and individual patients, but also by trialists and trial monitors.

Preventive strategies

Achieve commitment to the trial in the front lines

Application of the "coal-face" principle (only clinicians and patients who consider an RCT "theirs" should be expected to follow its protocol) should begin with the first draft of the question to be answered by the RCT (fighting through the question to be posed by an RCT deserves much of the total effort expended on it, and will be discussed in greater detail in future essays in this series). The question should be shown, discussed and argued over with an ever-widening array of clinical and methodological collaborators, and ultimately with the front-line physicians, research assistants and study nurses who will effect its success or failure. In multicentre RCTs, part of each centre's responsibility should be to educate and involve those who are admitting and following the study patients and responding to their questions and concerns throughout the trial. The benefits of this time-consuming activity are 4. First, the attendant discussion and debate improves the specificity and clinical usefulness of the question. Second, when this process draws in other scientists, including bench researchers, it

improves the science used to answer the question (and explain it). Third, these discussions permit collaborators and front-line participants to "buy in" to the trial and develop both the ownership and commitment that are essential for the successful assembly, care and follow-up of study patients and for adherence to the study protocol. Finally, the discussions provide the forum in which to understand the uncertainty principle and to gain confidence in the front lines in its ability to preserve patient choice and clinical judgement while protecting study validity.

Apply the uncertainty principle when entering patients into RCTs

My contribution to the failure of the penicillin trial would have been prevented if its eligibility criteria had incorporated the general principle of uncertainty as it applies to the individual patient. I find this incorporation best articulated by Richard Peto and Colin Baigent:[6]

A patient should not be entered if the responsible clinician or the patient are for any medical or non-medical reasons reasonably certain that one of the treatments that might be allocated would be inappropriate for this particular individual (in comparison with either no treatment or some other treatment that could be offered to the patient in or outside the trial).

I was reasonably certain that the synthetic penicillin was inappropriate for my patient. Had the uncertainty principle been in effect, my patient would never have entered the trial, and its internal validity would have been protected. Yes, one less patient would have entered the trial. But the consequent loss in the study's precision could have been made up by prolonging its recruitment phase, while the loss in its validity was irreparable.

There are several supplemental benefits to this preventive strategy. First, making patients equal partners in the application of the uncertainty principle legitimizes

---

Preventing RCT failure

When the success or failure of the RCT depends on the recruitment of patients who are appropriate in type and number, achieve commitment to the trial in the front lines (the "coal face"). This will
- improve adherence to the protocol,
- improve the study question,
- improve the science and
- provide a forum for incorporating values into research.

When clinicians and patients have hunches about efficacy (i.e., always), apply the uncertainty principle when entering patients into RCTs. This will
- recognize and reinforce clinical judgement and patients' values and autonomy,
- reduce complex eligibility criteria and paper work,
- act synergistically with the deliberations of the trial monitors and
- protect internal validity regardless of whether hunches are correct or wildly wrong.

their hunches, respects their autonomy and reinforces the need for their informed consent. Second, its application can reduce complexity, confusion and waste in the generation and application of eligibility and ineligibility criteria. When exclusion criteria try to anticipate all of the real-world situations in which a reasonable clinician might not want to invite an eligible patient to join an RCT, they swell in size and complexity, confound the patient's risk and responsiveness with the clinician's responsibility and can result in unnecessarily strict entry criteria (e.g., age) and decreased patient numbers. Third, the application of the uncertainty principle to individual patients acts synergistically with the deliberations of the trial monitors, who examine the accumulating unblinded results. Their prime duty is to monitor the boundaries of uncertainty at the group level, alerting the trialists when it shrinks (for all patients or sensible subgroups) to the point where the more effective treatment becomes clear (and if this clarity emerges for some prespecified subgroups but not others, the trial can be stopped for the former but continued for the latter, all on the basis of the uncertainty principle). Finally, while respecting the clinical judgement that keeps some patients and clinicians from entering RCTs, the uncertainty principle is impervious to the validity of their hunches; even when the hunches are wildly wrong, acting on them will not damage the internal validity of the trial result.

References

1. Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456-8.
2. Singer PA, Lantos JD, Whitington PF, Broelsch CE, Siegler M. Equipoise and the ethics of segmental liver resection. *Clin Res* 1988;36:539-45.
3. Lilford RJ, Jackson J. Equipoise and the ethics of randomisation. *J R Soc Med* 1995;88:552-9.
4. Hellman S, Hellman DS. Of mice but not men — problems of the randomized trial. *N Engl J Med* 1991;324:1585-9.
5. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987;317:141-5.
6. Peto R, Baigent C. Trials: the next 50 years. *BMJ* 1998;317:1170-1.

Reprint requests to: *Dr. David L. Sackett, Trout Research and Education Centre at Irish Lake, RR 1, Markdale ON N0C 1H0; sackett@bmts.com*