HEROES AND MARTYRS OF QUALITY AND SAFETY

Quality improvement research: are randomised trials necessary?

D Neuhauser, M Diaz

Have you ever heard the criticism that a quality improvement project was not scientific because it was not carried out as a randomised controlled trial (RCT)? These critics have got it wrong and here is why. This answer requires a review of the scientific method. (There are lots of good references to the scientific method. See <http://www.wikipedia.org> on "the scientific method".)

At its core, science is based on generalisable theory. This theory leads to hypotheses which can be tested experimentally over and over again. If the test fails, the theory must be rejected or modified. Replication, not peer review, is the standard of scientific evidence.

Look at the pen in your hand. There is scientific theory called gravity from which you can hypothesise that if you let go of your pen it will fall to your desk top. Drop your pen and you have tested this hypothesis and supported the theory of gravity.

If instead, your pen flies into the air you may have thrown it, be on a space ship or in Harry Potter's magic arts class. The first two possibilities are well defined and predictable by the full mathematical theory of gravity. The third possibility puts us in the realm of imagination which we can enjoy, but not replicate in reality. In a way, Harry Potter's Hogwarts School is like pure mathematics. Both are built on assumptions that lead logically to imaginary behaviour or theory which are beyond our real world of experimentation.

Scientific theory simplifies reality, tells us what is important, and allows us to generalise to other circumstances. It does not matter for the simple theory of gravity whether your pen is blue or black or even if it is a pencil or a penny. Theory defines what is not relevant, and what is relevant (height above sea level for gravity) and therefore what must be controlled for. This theory of gravity is tested millions of times a year in high school science classes all over the world, but it is still just a theory awaiting refinement by astrophysicists.

QUALITY IMPROVEMENT RESEARCH

To be a science, quality improvement studies need to be embedded in theory and be generalisable. If you make something work better at your hospital, what parts of your experience can be applied elsewhere? If it is unique to your organisation it is not generalisable.

Susan Cotey is a diabetes educator at Huron Hospital in Cleveland, Ohio. She gives out graph paper to elderly diabetic patients who live in the most impoverished part of her city. She uses a self-help book designed specifically for her patients. Each patient gets a copy. She asks them to plot their blood sugar measures over time, connect the dots and bring their graphs in to small discussion groups of similar patients who share their experience and learn about diabetes self-management (diet, exercise, weight control). Nearly every patient brings in their graph. The large majority of patients improve their diabetic control. This hospital has made diabetes management a centre of its healthcare mission. Figure 1 shows the graph paper format

used to create run charts for Sue's patients. Patients are also given daily diary tables to fill out, bring in and discuss (fig 2).

Sue's work is described above in generalised terms. This is good science reporting to the degree that others could replicate her work. If your clinic cares for diabetic patients you could start handing out graph paper and diary tables to patients and invite them to come back in a few weeks for a group education activity. You could see if your patients used the graph paper, came back and lowered their blood sugar levels. In short, Sue's method is very easy to replicate. You could try it even if you are not in Cleveland. It might work or it might not. This is an easily testable hypothesis.

THREE LEVELS OF GENERALISABILITY

Sue Cotey's diabetes care programme can be described at three levels of abstraction associated with different degrees of generalisability.

1. Low level of generalisability

The programme can be described in specific detail. Five to 10 elderly diabetic patients meet three times over six weeks at the Huron Hospital to learn about diabetes, diet and exercise and share experiences. This detailed level of description is important for purposes of replication. However, a patient's unique blood sugar levels may be generalisable only to this patient's past, present and future health status.

2. Intermediate level of generalisability

Group meetings, the use of graphed data over time and patient education is a description at a greater level of generalisability. Stated this way these methods could be used for patients with

Susan Cotey

Sue Cotey RN CDE is a certified diabetes educator with a passion for her work and her patients. Since 2002, she has counselled over a thousand diabetic patients from the East Cleveland area her hospital serves. Her office started in a back corner of the hospital. Her recognised success in attracting patients, helping them learn, improve and measure their own progress led the hospital's leadership to make her work central by moving her clinic to prime space in the rebuilt lobby of the hospital. Sometimes half the public buses in the city have signs promoting her programme. Her office is in a high traffic area where patients stop in to chat. The small waiting area is a friendly social place. If Sue Cotey is a "one of a kind" care giver and the success of her programme is totally dependent on her charisma then her experience is not reproducible. However, that hypothesis can be tested. The science of quality improvement has to be based on processes that can be used by others.

To contact Susan Cotey use scotey@ccfseast.org

Figure 1 Graph form used by patients to create their own blood sugar run charts. The darker horizontal band indicates the in-control range of values.

hypertension. In fact, Chris Hebert, MD did this with successful results.¹ This level of generalisability is applicable across a lot of patients and healthcare organisations.

3. High level of generalisability

Personal involvement, paying attention, readiness for change, just in time education, social support from small groups, and rapid feedback are conducive to change. Stated this way, these findings are generalisable to all human behaviour. However, at this level of abstraction you do not have enough information to repeat what Sue has done.

A good quality improvement publication should describe the programme and results at all three levels of abstraction in enough detail so others can replicate it, and broadly enough so

that it is applicable to other care programmes and human behaviour. If only the highest level of abstraction is used, the reader cannot repeat the study.

THREE TYPES OF QUESTIONS

Depending on the question asked (or if you will: how the hypothesis is stated) the test of the theory and the research design is different. These questions or hypotheses are:

1. Does Sue Cotey's approach to diabetes control reduce blood sugar levels for most patients (over time)?
2. Is this diabetes treatment approach better than other methods (concurrent comparison)?
3. Does the use of graph paper by patients improve care (reductionism)?

Date: _____

Time	Blood sugar reading before meal	Medication taken	Carbohydrates eaten (list portion size)	Exercise (Y/N)	Exercise type	Other factors affecting your blood sugar
Breakfast						
Lunch						
Dinner						
Before bed						

Figure 2 The diabetic patient's diary table for one day.

This paper builds on ideas previously developed in this series:

- For the debate between clinical judgment, laboratory science and clinical research see Best M, Neuhauser D. Pierre Charles Alexandre Louis: master of the spirit of mathematical clinical sciences. *Qual Saf Health Care* 2005;**14**:462–4.
- For the problem of replicating historical trials, see Neuhauser D, Diaz M. Daniel: using the Bible to teach quality improvement methods. *Qual Saf Health Care* 2004;**13**:153–5.
- For a comparison between randomised controlled trials and statistical process control, see Diaz M, Neuhauser D. Pasteur and parachutes: when statistical process control is better than a randomized clinical trial. *Quality and Safety in Health Care* 2005;**14**:140–3.

These are three different questions with important research consequences. The first deals with improvement over time and can be answered without reference to concurrent comparisons. This is the core approach of quality improvement and has important differences to current clinical research. It suggests the value of using run and control chart methods and the constant improvement of the processes involved. This is the domain of statistical process control.

The second question requires a concurrent comparison. The comparison requires controlling for other characteristics which might affect the results (being in a space ship and testing gravity). This is the way the question is asked for RCTs. This approach is also used for benchmarking.²

The third question is reductionist. Does one aspect of the intervention (graph paper use) affect the outcome? To answer this question you need to replicate Sue's work in identical groups except for the use of graph paper by the experimental group. In Sue's case the whole effort may make a difference even if the use of graph paper alone may have little observable impact.

Many, if not most, quality improvement projects are combined bundles of activities that all together can produce improved results over time. If so, it does not make sense to pull apart their components.

Without a battery the car does not go. Therefore, batteries make the car go. Without tires the car does not go. Therefore, tires make the car go. Of course you also need both an engine and gas. Reductionist reasoning in system contexts has its problems.

THREE PROBLEMS WITH RANDOMISED CONTROLLED TRIALS

The widespread use of RCTs is a relatively new phenomena of the last 25 years. In 1980 there were 21 RCTs reported in the *New England Journal of Medicine* in an entire year. Today the Cochrane Collaborative lists about 500 000 of them (see <http://www.cochrane.collaborative.org> (The Cochrane Library)). At their simplest they ask the question: is drug X better than usual care for Y patients? That is to say it is a controlled comparison with current practice. Such trials have three important flaws.

1. After a decade or two, they cannot be replicated for two reasons: (a) we cannot replicate 1980s medicine—by 2007 the control arm of these trials has changed; (b) if the original trial showed a difference, a replication may well be considered unethical. We cannot deny a beneficial drug or impose a harmful one. Because of the recent growth of RCTs, this is just

starting to be a problem. This vast body of trials is on its way to obsolescence.

2. These RCTs are largely devoid of generalisable theory. The question “is drug X better?” does not allow us to generalise beyond that drug. However, RCTs may be generalisable in two ways. There could be a theory based on molecular biology or pathophysiology or both which would allow us to make predictions about new similar drugs. These trials may be based on animal studies in controlled laboratory conditions. These animal studies can be replicated in the future.

The use of randomisation is an admission of our ignorance. We do not know what causes variation in disease outcomes. For the physicist studying gravity the parameters that matter are known and controlled for so that the only effect is the experimental one. A good study repeated over and over will get exactly the same results. Such studies do not use randomisation.

3. Because replication is at the heart of science, other things equal, expensive science is bad science. Expensive studies are by definition harder to repeat. Small simple, easily repeated, generalisable quality improvement projects can be very good science. Sue Cotey's work is an example.

THREE REASONS TO USE STATISTICAL PROCESS CONTROL

1. Are we improving? This is a question we should be asking and answering more frequently. Run and control charts are at the heart of this method. RCTs tell us what drug works, but not whether patients who could benefit get the drug. Quality improvement focuses on implementation.

2. For the first time we are collecting vast amounts of information about the process of care and its outcomes due to the steady diffusion of computers. Health system plans can now account for, in real time, how many diabetic patients they care for and how many have had a haemoglobin test in the last six months. They can know the process of care and outcomes. They are just learning how to use this information to make improvements. For chronic conditions, these information systems will have to allow patients to enter their own data so that there will be “real time” data about patient's health.

3. The focus is on system management. Evidence-based management asks if managerial decisions are based on data and research or just opinion. Just as medicine and nursing has moved away from decisions based solely on senior clinical opinion to evidence-based care, health management needs to do the same.

WHY RANDOMISED CONTROLLED CLINICAL TRIALS?

The evidence RCTs provide about current practice is better than unaided clinical opinion. “In my experience as a great grandmother, chicken soup is the best cure for a child's cold.”

The financing structure of healthcare worldwide requires such trials to identify what treatments should be paid for now by national health systems or health insurers. This is a short-term answer to a short-term managerial and policy problem.

Depending on how the hypothesis is stated they can be the best way to answer a question. An example of a good RCT is the Rand Insurance Trial.³ In this trial some families were given health insurance with large deductibles. Such middle class families used a lot less care without demonstrable effect on their health. These results have led to greater use of large deductibles by employer-based health insurance plans in the US. The virtue of this trial is that it was embedded in economic theory and as such is widely generalisable both in the 1990s and decades from now, because economic behaviour is not expected to change.

SPC and RCTs answer different questions and fill different needs. SPC needs to be used more as a foundation of evidence-based health management. Quality improvement efforts need much better quantitative documentation over time which shows the relation of process to outcome. RCTs need to be embedded in generalised replicatable theory. Otherwise it is a scientific house without a foundation. With some thought and a focus on theory, generalisability and replication, quality improvement versus RCTs is an unnecessary argument.

.....

Authors' affiliations

D Neuhauser, M Diaz, Epidemiology & Biostatistics, Medical School, Case Western University, Cleveland, Ohio, USA

Correspondence to: Professor D Neuhauser, Epidemiology & Biostatistics, Medical School, Case Western University, 1900 Euclid Avenue, Cleveland, Ohio, USA; dvn@case.edu

Accepted 23 November 2006

REFERENCES

- 1 **Hebert C**, Neuhauser D. Improving Hypertension Care with Patient Generated run Charts: Physician, Patient and Management perspectives. *Qual Manage Health Care* 2004;**13**:174–7.
- 2 **Lorig KR**, Ritter PL, Laurent DD, *et al.* Internet-based chronic disease self management: a randomized trial. *Med Care* 2006;**44**:964–71.
- 3 **Newhouse J**, the Insurance Experiment Group. *Free for all? lessons from the health insurance experiment*. Cambridge Mass: Harvard University Press, 1993.

BNF for Children 2006, second annual edition

In a single resource:

- guidance on drug management of common childhood conditions
- hands-on information on prescribing, monitoring and administering medicines to children
- comprehensive guidance covering neonates to adolescents

For more information please go to **bnfc.org**