

Education and debate

Individual response to treatment: is it a valid assumption?

Stephen Senn

Most drug trials assume that patients respond consistently to treatment, but the assumption is rarely tested. If patients vary randomly in their response to a drug rather than some patients never responding, searches for a genetic basis for non-response are futile

Department of
Statistics, University
of Glasgow,
Glasgow G12 8QQ
Stephen Senn
professor of statistics
stephen@
stats.gla.ac.uk

BMJ 2004;329:966-8

Imagine a trial with 1000 representative patients, chosen from a population of patients with erectile dysfunction, until now resistant to treatment. Each is given the opportunity of trying a new treatment once. Seven hundred succeed in gaining an erection; the other three hundred fail. How should we interpret these results?

One common interpretation is that the treatment works for 70% of patients 100% of the time and for 30% of the patients 0% of the time. However, nothing in the data forbids a radically different interpretation—namely, that the treatment works in 100% of the patients 70% of the time. In the first case, ability to succeed on treatment is a permanent feature of the patient. In the second case, individual response cannot be predicted: the patients are indistinguishable from each other regarding response to treatment. They sometimes respond and they sometimes do not. Intermediate cases between these two extremes are, of course, also possible.

Examples of confusion

Most clinical trials do not permit us to distinguish between these two extreme cases or indeed any intermediate case. Yet many trialists plump for the first explanation—that of individual response to treatment—rather than pure random variability. In fact, you do not have to search far in the pages of the *BMJ* to find examples of the unstated assumption of individual response to treatment dictating the interpretation of clinical trials. I shall consider two from the *BMJ* and a third example from elsewhere.

My first concerns a statement of Allen Roses, which was reported by Richard Smith, the former editor of the *BMJ*, as follows:

[The] worldwide vice president of genetics at GlaxoSmith Kline, is reported on the front page of the *Independent* (8 December, p 1) as saying: 'Our drugs don't work on most patients'. . . He is an enthusiast for pharmacogenomics and hopes that greater understanding of genetics will mean that we will be able to identify with a "simple genetic test" people who will respond to drugs and design drugs for individuals rather than populations. We have, however, been hearing this tune for a long time, and it's hard to see the business model for individually tailored drugs.¹

The last sentence here is the wisest. We have, indeed, been hearing this for a long time. Neither



Is individual response to a drug consistent?

Smith nor Roses, nor indeed anybody else, would be in a position to tell whether the drugs concerned work moderately well for all patients or extremely well for some and not at all for others, for the simple reason that GlaxoSmithKline, like all other drug companies, runs almost no trials that would be capable of identifying one explanation from the other.

My second example comes from an article in the *BMJ* in 1998 in which Guyatt et al claimed: "A method for estimating the proportion of patients who benefit from a treatment when the outcome is a continuous variable has been developed."² The method is most simply illustrated for a crossover trial. It consists of calculating for a given patient the difference between treatment and control and comparing this with some agreed standard, say a clinically relevant difference, to judge whether the observed difference is important.

But the same problem arises as with our erectile dysfunction example. Consider a crossover trial in asthma comparing salmeterol with salbutamol in which it is judged that a difference in forced expiratory volume in one second of 200 ml is clinically relevant. The trial is run, and 24 out of 32 patients exhibit a difference at least as great as this, whereas eight do not. The 24 are labelled as responders and the other eight as non-responders, and we conclude that salmeterol produces a clinically relevant superior response to salbutamol for three out of every four patients, or at least this is what Guyatt et al would invite us to believe.

In fact, things are not so simple. The table shows two extreme possibilities when we repeat the whole experiment again, so that we now have two comparisons of salbutamol and salmeterol for every patient. In the first case, we have perfect correlation between the responses in the two crossover trials and in the second we have independence. The two cases are radically different and are identifiable by virtue of the fact that the effect of each treatment has been measured in more than one period. It is the pattern of joint responses that permits the identification of the case. The margins on the table are the same in both cases and do not permit identification.

In fact, there is a further difficulty. Suppose the average effect in such a trial is greater than the clinically relevant difference of 200 ml. If there is any within patient variability due to either pure measurement error or random temporal fluctuation in the state of the patient, then the observed difference between measurements on the different drugs for a given patient may be less than the clinically relevant difference, even though the true difference is greater than it. Suppose that this within-patient standard deviation is 100 ml. Then it can be calculated (see box 1) that even if the true effect were constant and equal to 250 ml for every single patient, and hence greater than the clinically relevant difference of 200 ml, on average 36% would randomly fail to show such a clinically relevant difference between treatments.

My third example comes from a re-analysis of the β blocker heart attack trial (BHAT) by Horwitz et al.³ Here the title of the article says it all: "Can treatment that is helpful on average be harmful to some patients?" As the authors put it:

The 31 centers were divided into 21 dominant centers (mortality rates higher for placebo than propranolol) and 10 divergent centers (higher mortality rates for patients randomised to propranolol). Overall, compared to placebo, propranolol reduced the risk of dying for the "average" patient from 9.8 to 7.2%. Results for patients in dominant centers (RR = 0.50) were significantly different from those in divergent centers (RR = 1.33).

This use of a significance test on groups of centres that are identified only by result is, of course, quite illegitimate. The authors continued:

We conclude that differences in results across centers of a multicenter RCT may reflect important distinctions in the clinical conditions of enrolled subjects. These distinctions help to identify subgroups of patients in which treatment that has an average overall benefit may be harmful for some patients.

This may, of course, be true in general, but unfortunately for Horwitz et al it is not true of the BHAT study.

Pattern of responses in two crossover trials using the same patients with models of perfect correlation and independence

First crossover	Second crossover		Total
	Responders	Non-responders	
Perfect correlation			
Responders	24	0	24
Non-responders	0	8	8
Total	24	8	32
Independence			
Responders	18	6	24
Non-responders	6	2	8
Total	24	8	32

Box 1: Calculation of proportion who will appear to show a clinically relevant difference in asthma trial

If the within-patient measurements are independent, the variance of the difference between them will be twice the variance of an individual measurement. Hence the standard deviation of the differences will be the square root of twice the individual standard deviation. In our example, the standard deviation of the difference will be $\sqrt{2} \times 100 \text{ ml} \approx 141 \text{ ml}$.

If the FEV₁ values are normally distributed, then we can calculate the probability that a given difference will be less than the clinically relevant difference from tables of the standardised normal distribution. Since the clinically relevant difference is 200 ml and the mean is 250 ml, the standardised difference becomes $(200 - 250) / 141 \approx -0.35$, and the probability of a standard normal deviate being less than this is 0.36. Hence 36% of all patients will fail to show a clinically relevant difference when the two drugs are compared once, even though on average in the long run they would show such a difference.

The study was re-analysed by Senn and Harrell using a random effects model, and this analysis produced a result that even they did not expect: there was no variation between centres above and beyond that ascribable to random variation (box 2).^{4,5}

Lesson for pharmacogenomics

The lesson for those in pharmacogenomics is the following. To the extent that the purpose of such research is to identify genetic factors governing individual response to treatment, it is founded on a largely untested assumption—namely, that such consistent individual responses exist. Patient by treatment interaction (that is to say individual response to treatment) provides an upper bound to gene by treatment interaction (differential response by genetic subgroups) because patients differ by more than their genes.⁶ Kalow et al therefore suggest that when the disease is chronic and crossover trials with repeat administration are possible (that is, a series n-of-1 trials) they should be carried out to identify disease and treatment combinations in which individual response is important as a preliminary step before looking for genetic factors.⁷ Unless patient by treatment interaction exists, it is pointless looking for gene by treatment

Box 2: Multicentre trials: why random differences between centres are inevitable

In multicentre trials, not only are centres far too small to show significance alone (hence the need for a multicentre trial), they are even too small to guarantee that an effect reversal (whereby the poorer treatment is observed to perform better) cannot occur. The more centres there are, the more likely it is that some will show an effect reversal, for two reasons. The first is that the more centres, the smaller the fraction the average centre has of the number of patients required to show reliable results and the second is that the more centres there are the more chances there are that at least one will buck the trend. In the BHAT study, the centres were so small, the event so rare, and the treatment effect so small that about 10 out of 31 effect reversals were expected by chance alone. There was no evidence of any differential response between centres. The moral is that extreme care must be taken in examining variation between centres.

Summary points

For most clinical trials it is impossible to identify which patients will usually respond to treatment

It is often assumed without proof that such individual response is important

The difficulties of interpretation are exacerbated if continuous measurements are dichotomised

Since individuals differ by more than their genes, genetic variability cannot exceed individual variability

Carefully designed repeated period crossover trials have a useful role in identifying individual, and by extension genetic, response

interactions, and patient by treatment interaction can be examined only by repeated crossover trials.⁸ Such trials can then be analysed using random effect models in a way that will permit the resolution of variability into various sources: the overall effect of treatment, variability between patients, variability within patients, and patient by treatment interaction.⁹

The pharmaceutical industry has rarely, if ever, carried out the sort of trial that would permit identifica-

tion of patient by treatment interaction. Thus statements that the drugs don't work on most people are based on mere supposition; the drugs may work moderately well for all people. To identify those drug and disease combinations for which individual response is important and hence for which genetic factors may be, it will be important to plan and analyse carefully.

Contributors and sources: SS used to work in the pharmaceutical industry and his research interests include design and analysis of clinical trials, in particular crossover trials.

Competing interest. SS is a consultant to the pharmaceutical industry.

- 1 Smith R. The drugs don't work [editor's choice]. *BMJ* 2003;327(13 Dec). <http://bmj.com/cgi/content/full/327/7428/0-h?>
- 2 Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ* 1998;316:690-3.
- 3 Horwitz RI, Singer BH, Makuch RW, Viscoli CM. Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. *J Clin Epidemiol* 1996;49:395-400.
- 4 Senn SJ, Harrell F. On wisdom after the event. *J Clin Epidemiol* 1997;50:749-51.
- 5 Senn SJ, Harrell FE Jr. On subgroups and groping for significance. *J Clin Epidemiol* 1998;51:1367-8.
- 6 Senn SJ. Individual therapy: new dawn or false dawn. *Drug Info J* 2001;35:1479-94.
- 7 Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics* 1998;8:283-9.
- 8 Senn SJ. Applying results of randomised trials to patients. N of 1 trials are needed. *BMJ* 1998;317:537-8.
- 9 Senn SJ. Author's reply to Walter and Guyatt. *Drug Info J* 2003;37:7-10. (Accepted 11 August 2004)

The PROGRESS trial three years later: time for a balanced report of effectiveness

Richard Wennberg, Camilla Zimmermann

Has the use of the phrase "perindopril based blood pressure lowering regimen" resulted in an oversimplistic and hence inaccurate interpretation of the results of the PROGRESS trial creeping into the literature and clinical practice?

Any doctor who keeps up to date with the stroke literature will be familiar with PROGRESS—the perindopril protection against recurrent stroke study,¹ a clinical trial which had resulted, by February 2004, in more than 35 ancillary publications. The aim of the study was to "resolve clinical uncertainty about the efficacy and safety of routine blood-pressure-lowering therapy for individuals with a history of stroke or transient ischemic attack."¹ The trial found that blood pressure lowering was safe and effective, but three years after its publication, the optimal antihypertensive regimen for secondary stroke prevention remains unclear. The title of the study, together with its results as reported, may lead the unwary reader to conclude that perindopril used alone protects against recurrence of stroke. Although this was actually not a finding of PROGRESS, the study design and data presentation obfuscate this fact rather than making it clear.

Design of the trial

The PROGRESS trial had a "flexible" design, which meant that not all patients received the same antihypertensive regimen. Patients who had had a stroke or transient ischaemic attack in the previous five years were

eligible. They were also required to have "no definite indication (such as heart failure) for treatment with an ACE inhibitor and no definite contraindication (such as previous intolerance) to such treatment." The entry criteria did not include blood pressure, but treatment with agents other than angiotensin converting enzyme inhibitors was recommended to patients with uncontrolled hypertension before they entered the trial. Patients were then randomised, on a double blind basis, to either active treatment or placebo. All active treatment patients received treatment with perindopril. Indapamide was added "at the discretion of the individual physician" if a patient had "no specific indication for or contraindication to treatment with a diuretic."¹

After a mean follow up of 3.9 years, the active treatment group as a whole had significantly fewer strokes and major vascular events. However, a prespecified subgroup analysis showed that although participants treated with the combination of perindopril plus indapamide had a significantly lower stroke risk than patients who received double placebo (43% risk reduction; 95% confidence interval 30% to 45%), patients treated with perindopril alone had a stroke risk that was not discernibly different from placebo (5% risk reduction, -19% to 23%). Even though there was

Department of Medicine,
University of Toronto, University Health Network,
Toronto, ON,
Canada M5T 2S8

Richard Wennberg
associate professor
Camilla Zimmermann
lecturer

Correspondence to:
R Wennberg
Richard.Wennberg@uhn.on.ca

BMJ 2004;329:968-71