

COMMENT

STATISTICS A call to police the whole data-analysis pipeline, not just *P* values **p.612**

SPRING BOOKS Does Nicholas Stern's global vision admit ground truth? **p.614**

SPRING BOOKS Metaphor pile-up obscures the meaning of junk DNA **p.615**

SPRING BOOKS Grind, politics and dirty tricks in life of polio-vaccine pioneer **p.620**

ILLUSTRATION BY GREG CLARKE



Time for one-person trials

Precision medicine requires a different type of clinical trial that focuses on individual, not average, responses to therapy, says **Nicholas J. Schork**.

Every day, millions of people are taking medications that will not help them. The top ten highest-grossing drugs in the United States help between 1 in 25 and 1 in 4 of the people who take them (see 'Imprecision medicine'). For some drugs, such as statins — routinely used to lower cholesterol — as few as 1 in 50 may benefit¹. There are even drugs that are harmful to certain ethnic groups because of the bias towards white Western participants in classical clinical trials².

Recognition that physicians need to take individual variability into account is driving huge interest in 'precision' medicine. In January, US President Barack Obama announced a

US\$215-million national Precision Medicine Initiative. This includes, among other things, the establishment of a national database of the genetic and other data of one million people in the United States.

Classical clinical trials harvest a handful of measurements from thousands of people. Precision medicine requires different ways of testing interventions. Researchers need to probe the myriad factors — genetic and environmental, among others — that shape a person's response to a particular treatment.

Studies that focus on a single person — known as *N*-of-1 trials — will be a crucial part of the mix. Physicians have long done these in an ad hoc way. For instance, a doctor

may prescribe one drug for hypertension and monitor its effect on a person's blood pressure before trying a different one. But few clinicians or researchers have formalized this approach into well-designed trials — usually just a handful of measurements are taken, and only during treatment.

If enough data are collected over a sufficiently long time, and appropriate control interventions are used, the trial participant can be confidently identified as a responder or non-responder to a treatment. Aggregated results of many *N*-of-1 trials (all carried out in the same way) will offer information about how to better treat subsets of the population or even the population at large. ►

► Formalizing and scaling up the *N*-of-1 approach means solving various practical problems. These include exploiting the diversity of health-monitoring devices, developing new ones and identifying appropriate disease biomarkers, such as tumour DNA circulating in the bloodstream. It will also require a cultural shift on many levels — in regulatory agencies, in pharmaceutical companies and, most of all, in the clinic.

A WORLD OF DIFFERENCE

Discovering that an intervention works well in certain groups happens relatively rarely and often by chance. Researchers typically get disappointing results with a drug in large, population-based trials. This leads them to conduct ad hoc post-trial analyses, to try to identify the factors that cause some of the people in the trial to seem to be responsive³.

For instance, the drug Gleevec (imatinib) was found to double survival rates of leukaemia patients⁴ with a chromosomal abnormality in their tumours called the Philadelphia translocation. Similarly, it turns out that Erbitux (cetuximab) improves the survival of people with colorectal cancer whose tumour cells carry a mutated *EGFR* gene but not a mutated *KRAS* gene⁵.

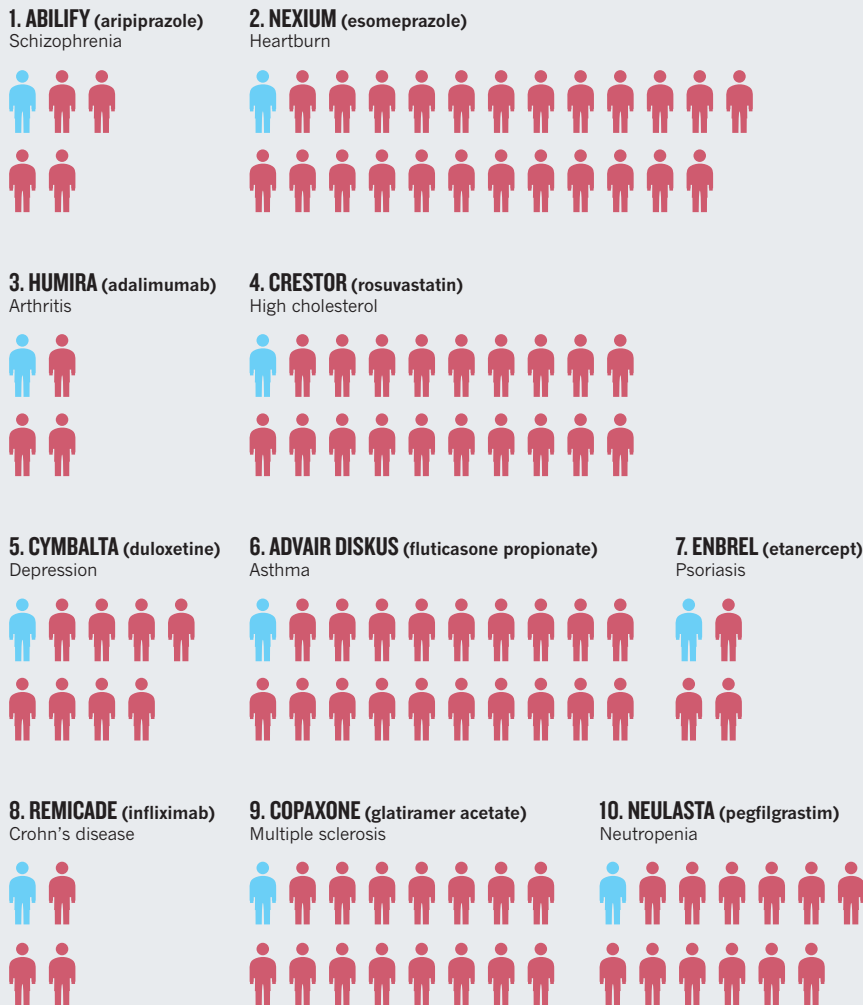
This approach to discovery is inefficient at best. Conventional phase III trials involve thousands of people. The intervention being tested is often given at random to one group while another group receives a sham treatment, such as a sugar pill or the standard treatment that physicians would give such patients. Because scant data are collected on factors such as genetics, lifestyles and diets, the results of these trials often indicate the need for yet another study to validate the effectiveness of the intervention among the apparent responders and to establish the underlying mechanisms.

Various trial designs have been developed that better account for variability between patients. Basket trials, which have mainly been used for cancer, test the effectiveness of an intervention on the basis of its mode of action, regardless of what disease it was designed to treat. For instance, the US National Cancer Institute's MATCH Trial, expected to launch in May, will use genetic markers from tumours to assign 1,000 people who have different types of late-stage cancer to different treatment 'baskets'. Baskets will include around 30 patients, all of whom share certain genetic anomalies; participants will be given one of around 25 drugs. The effectiveness of the different matches of treatment to genetics will be compared as the trial unfolds.

In umbrella trials, researchers test the effectiveness of multiple drugs in a study of a single disease. This approach is being used in the ongoing Stand Up To Cancer and Melanoma Research Alliance Dream Team clinical trial, in which I am involved, run by Yale University

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).



Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr78f.

in New Haven, Connecticut, and the Translational Genomics Research Institute in Phoenix, Arizona. We are giving genomically guided treatment to around 70 people out of 100 with late-stage melanoma — choosing from an array of 40 or so drugs (the rest of the participants are receiving the standard treatment usually given to such patients). Here, the effectiveness of matches between drugs and genetics will be compared with standard care. Lastly, adaptive trials aim to match interventions to patients while the study is ongoing, on the basis of patient responses.

Even these trial designs may not be personalized enough. Among people who share, say, a particular mutation known to be targeted by a specific drug, many other factors can contribute to any one person's responsiveness. This is particularly true for those with cancer. The drug vemurafenib, for instance, was approved in the United States to treat late-stage melanoma in people whose tumours

carry the BRAF(V600E) mutation. But some tumour cells develop other anomalies that make them resistant to the drug. Thus clinicians considering whether to prescribe vemurafenib may need to take into account a whole slew of biomarkers.

COMPARING TREATMENTS

In *N*-of-1 trials, all sorts of relevant data will need to be collected for one person, as frequently as possible — perhaps every day or periodically over months or years. The usual design and statistical safeguards could be employed, such as blinding patients and experimenters to the drugs being tested, and the use of control interventions (such as periods of standard care). In addition, appropriate crossover designs, in which different interventions are administered to the same person alternately (possibly with 'wash-out' periods in between to allow the drugs' effects to wear off) would enable experimenters to

compare the effect of different treatments in the same person.

If done properly, claims about a person's response to an intervention could be just as well supported by a statistical analysis as by analyses designed to assess population-level responses on the basis of classical clinical trials⁶.

An example of this approach is a study⁷ in Australia, which measured reported pain levels, swelling and other symptoms associated with osteoarthritis and chronic pain in 132 people taking different drugs over three years. For each person, measurements were taken every 2 weeks for 12-week periods, when the patient was either off or on a particular drug. By comparing the data collected before and after the different treatments, the researchers showed that, although initially costly, the formalized *N-of-1* trials resulted in more-effective prescriptions.

Sometimes *N-of-1* trials will be neither appropriate nor feasible. For instance, the costs are probably too high for public-health studies that investigate the effect of a population-wide intervention, such as adding fluoride to drinking water. Making objective claims about individual responses requires taking appropriate measures (of tumour progression, say) repeatedly and efficiently. Yet it is not always clear what to measure. Only a fraction of the thousands of proposed biomarkers have been shown to be useful in the clinic.

But in many instances, an *N-of-1* approach is ideal. Such studies are already being done for some rare diseases by necessity, but often without the use of sophisticated trial designs and without necessarily collecting the appropriate information to make hypotheses about the drug's mechanism. Many experimental drugs are also administered in 'compassionate use' settings. And many widely used drugs are provided to combat diseases for which they were not approved ('off label' prescription), for people who fail to respond to all other treatments. Examples include uses of the drug mexiletine to treat the rare muscle disease non-dystrophic myotonia, and experimental treatments for the Ebola virus.

Well-designed *N-of-1* trials could also be useful in the early stages of clinical drug development or in repurposing drugs — for exploring the molecular and physiological effects of a new compound (or of an old compound in a new context). Likewise, studies investigating the safety and appropriate dosages of drugs could take an *N-of-1* approach. Currently, phase I and II clinical trials usually involve giving different amounts of a drug to a small group of healthy volunteers. Better would be to tailor dosages to individuals' metabolic profiles.

N-of-1 trials could be designed to guide clinicians in detecting disease onset. For

instance, US physicians generally view levels of a blood protein called CA125 greater than 30 or 35 as an indication of ovarian cancer. However, a level of 20 or 25 may be a cause for concern if the person's average CA125 levels hovered around 10 or 15 over the previous year⁸. Establishing personal thresholds for uncovering disease onset is the goal of the registered clinical trial known as the Tanner Project (www.tannerproject.org), in which I am involved.

By looking for commonalities across multiple *N-of-1* studies — in which the same types of data are collected using the same procedures — researchers should be able to draw inferences about the effectiveness of an intervention in certain subsets of the population, such as in people sharing particular genetic features, as well as in the whole population.

Various teams are developing and testing algorithms to match interventions, or a combination thereof, to individuals on the basis of their genetic make-up, biochemistry, diet and other factors. For instance, matching drugs to tumour profiles is a key goal of the Stand Up To Cancer umbrella trial.

MAKING IT HAPPEN

There are significant barriers to making *N-of-1* trials commonplace. Regulatory agencies, researchers and clinicians are rightfully wary of moving away from classical clinical trials. Pharmaceutical companies tend to focus on drugs that are likely to be used by thousands or millions of people. What is more, tailoring treatments to patients is costly. For example, the cancer-care company Foundation Medicine in Cambridge, Massachusetts, charges patients between US\$5,000 and \$7,500 to sequence their tumours and to use the results to advise on treatments. And there is a lot of work to be done on biomarkers, monitoring devices, study designs and data-analysis methods.

A key component will be transforming everyday clinical care into solid *N-of-1* trials. In my view, the time is ripe for three reasons.

First, there is a growing interest in 'omics' assays that expose people's unique characteristics at the molecular level. Researchers and clinicians are assaying people's blood metabolites (their metabolome) and the microbes in their bodies (their microbiome) as well as their DNA and RNA⁹. Second, cheap and efficient devices that collect health data are becoming available, such as the Apple Watch, continuous glucose monitors and portable electroencephalogram (EEG) monitors. Lastly, governments and life-sciences funding bodies worldwide are increasingly

supporting a more targeted approach as well as patient engagement in medicine, such as through the US Patient-Centered Outcomes Research Institute, established in 2010.

I am confident that, ultimately, governments, regulatory agencies and pharmaceutical companies will support sophisticated, well-designed *N-of-1* trials. These could save the millions of US dollars that are spent on inappropriate interventions, the management and treatment of persistent or recurring diseases, and on conventional phase III trials (which can cost between \$100 million and \$700 million per drug). Regulatory agencies such as the US Food and Drug Administration are beginning to recognize the importance of individual responses¹⁰. And sufficient financial or market incentives provided by governments could persuade pharmaceutical companies to broaden their focus away from 'blockbuster' drugs — especially given the poor rates of return on drug discovery in recent years.

Key to making precision medicine mainstream is the ongoing shift in the relationship between patients and physicians. A major advantage of the *N-of-1* approach over classical trials is that patients are no longer guinea pigs, whose involvement in a study may help only future generations. In *N-of-1* trials, the effectiveness of different treatments are vetted for the actual participants. Indeed, members of hundreds of patient-advocacy groups, for instance for rare genetic diseases, are eager to be involved in studies to test candidate drugs.

Physicians are having to become more acutely aware of the unique circumstance of each patient — something most people have long called for. ■

Nicholas J. Schork is director of human biology at the J. Craig Venter Institute in La Jolla, California, USA. He is also professor at the University of California, San Diego, and at the Translational Genomics Research Institute (TGen) in Phoenix, Arizona, USA. e-mail: nschork@jcvi.org

1. Mukherjee, D. & Topol, E. J. *Prog. Cardiovasc. Dis.* **44**, 479–498 (2002).
2. Currie, G. P., Lee, D. K. & Lipworth, B. J. *Drug Saf.* **29**, 647–656 (2006).
3. Uryniak, T. et al. *Stat. Biopharmaceut. Res.* **3**, 476–487 (2011).
4. Druker, B. J. et al. *N. Engl. J. Med.* **344**, 1038–1042 (2001).
5. Karapetis, C. S. et al. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
6. Kravitz, R. L. & Duan, N. *Design and Implementation of N-of-1 Trials: A User's Guide* (Agency for Healthcare Research and Quality, 2014).
7. Schuffham, P. A. et al. *J. Gen. Intern. Med.* **25**, 906–913 (2010).
8. Drescher, C. W. et al. *J. Clin. Oncol.* **31**, 387–392 (2013).
9. Chen, R. et al. *Cell* **148**, 1293–1307 (2012).
10. US Food and Drug Administration. *Paving the Way for Personalized Medicine: FDA's Role in a New Era of Medical Product Development* (US Department of Health And Human Services, FDA, 2013).