

# What Ever Happened to N-of-1 Trials? Insiders' Perspectives and a Look to the Future

RICHARD L. KRAVITZ, NAIHUA DUAN, EDMUND J. NIEDZINSKI, M. CAMERON HAY, SASKIA K. SUBRAMANIAN, and THOMAS S. WEISNER

*University of California, Davis; Columbia University; New York State Psychiatric Institute; Miami University; University of California, Los Angeles*

**Context:** When feasible, randomized, blinded single-patient (n-of-1) trials are uniquely capable of establishing the best treatment in an individual patient. Despite early enthusiasm, by the turn of the twenty-first century, few academic centers were conducting n-of-1 trials on a regular basis.

**Methods:** The authors reviewed the literature and conducted in-depth telephone interviews with leaders in the n-of-1 trial movement.

**Findings:** N-of-1 trials can improve care by increasing therapeutic precision. However, they have not been widely adopted, in part because physicians do not sufficiently value the reduction in uncertainty they yield weighed against the inconvenience they impose. Limited evidence suggests that patients may be receptive to n-of-1 trials once they understand the benefits.

**Conclusions:** N-of-1 trials offer a unique opportunity to individualize clinical care and enrich clinical research. While ongoing changes in drug discovery, manufacture, and marketing may ultimately spur pharmaceutical makers and health care payers to support n-of-1 trials, at present the most promising resuscitation strategy is stripping n-of-1 trials to their essentials and marketing them directly to patients. In order to optimize statistical inference from these trials, empirical Bayes methods can be used to combine individual patient data with aggregate data from comparable patients.

---

*Address correspondence to:* Richard L. Kravitz, Division of General Medicine, UC Davis School of Medicine, 4150 V. Street, Suite 2400 PSSB, Sacramento, CA 95817 (email: rkravitz@ucdavis.edu).

The Milbank Quarterly, Vol. 86, No. 4, 2008 (pp. 533–555)  
© 2008 Milbank Memorial Fund. Published by Wiley Periodicals Inc.

**Keywords:** N-of-1 trial, Bayesian methods, treatment effects heterogeneity, clinical effectiveness, research policy, personalized medicine.

**M**OST REVIEWS OF EVIDENCE-BASED MEDICINE EMPHASIZE THE role of well-conducted randomized controlled trials (RCTs) as the gold standard for evidence-based practice. However, RCTs provide estimates of how well a treatment works *on average* for patients in the trial. They do not predict how a particular individual will respond. Making such predictions is at the core of clinical therapeutics and requires a more individualized approach. One way to develop more refined predictions of therapeutic benefit is to perform within-individual randomized controlled trials, otherwise known as *n-of-1 trials*.

N-of-1 trials are single-patient trials of treatment effectiveness and safety. In this article, we argue that they could be used more widely to facilitate a modern clinical care and research enterprise that accounts for the heterogeneity of treatment effects across individual patients (Greenfield et al. 2007; Kravitz, Duan, and Braslow 2004). In so doing, however, we acknowledge the checkered history of the n-of-1 movement and consider why it has failed to gain traction with the clinical and research communities. Our analysis is informed by a review of the literature as well as telephone interviews with some leading n-of-1 pioneers and proponents. We conclude by suggesting that recent trends in drug development, discovery, and marketing have created a new opening for n-of-1 trials. In the end, an alliance of drug manufacturers, managed care organizations, patients, and physicians could revive the movement, bringing increased precision to clinical care while enhancing research translation.

## A Brief History of N-of-1 Trials

N-of-1 trials are crossover experiments conducted with a single patient. They are potentially relevant to guiding the therapy of chronic, nonfatal conditions, especially when the goal of treatment is symptom control. The central idea is that by carefully and fairly observing a patient's response to two or more treatments, a physician and patient can determine

which of the treatments is likely to work best for that patient over the long term.

N-of-1 trials typically have three elements: randomized assignment to repeated treatment episodes (often in pairs, e.g., AB-BA-BA); blinding of patient and clinician to the treatment delivered during each episode; and systematic, blinded measurement of outcomes. N-of-1 trials have a long history in psychology and education. B.F. Skinner was an early proponent, and Murray Sidman's classic text regarded single-subject trials as an accepted part of the psychologist's experimental arsenal (Sidman 1960). Then, beginning in the 1970s, several articles appeared describing and critically evaluating the use of single-patient experiments in psychotherapy, rehabilitation, and special education (Barlow and Hersen 1984; Kazdin 1982; Kratochwill 1978). Apart from an occasional application to psychotherapy, however, n-of-1 trials were essentially undiscovered in clinical medicine.

### *The Birth of a Movement*

That changed in 1986 when Gordon Guyatt and his colleagues reported their experience in the *New England Journal of Medicine* (Guyatt et al. 1986). Guyatt's story, obtained during a telephone interview in early 2007, is revealing. As he tells it, the Department of Clinical Epidemiology and Biostatistics at McMaster University in the early 1980s was a heady place:

The department was multidisciplinary and very tightly integrated. So there were . . . statisticians and psychologists and people with behavioral backgrounds, physicians and epidemiologists getting together on a regular basis. And for a while, one of the psychologists would say, "Oh, that would be very interesting for an n-of-1 trial." And we said, "Thank you very much" and would go on. Then at one point it clicked, and we started to get out the psychology literature and found three textbooks full of n-of-1 designs from a psychology perspective. . . . It was totally old news.

The McMaster group moved quickly, meeting regularly to discuss how the n-of-1 concept could be adapted to clinical medicine. The need for certain modifications was obvious. Blinding, for example, is essentially impossible in psychotherapy studies, but relatively easy to

implement in drug trials. In contrast, certain design elements popular with psychologists (e.g., interrupted time series) are overly complex for medical settings. After a few months of meetings, the group was ready to launch its first trial. As Guyatt recalled:

I had this seventyish-year-old patient who was an asthmatic, and he was really struggling. He was taking theophylline, he was on a beta-agonist, and he was taking oral steroids. So I thought, he's on all this medication. Let's find out how much good he's getting from his theophylline. We were planning to do three pairs [of comparisons between theophylline and placebo]. Well, after just two pairs, there was such a dramatic difference between one period and another that I said, "Look, there's just no question. We know what's going on here." We broke the code, and the guy was much worse taking his theophylline. And so we took him off his theophylline, and he was much better, dramatically better. . . . As a first example, that vividly conveyed the power of the tool.

The McMaster group described the n-of-1 concept and reported on the results of the theophylline trial in the *New England Journal of Medicine* (Guyatt et al. 1986). In this and a subsequent publication (Guyatt et al. 1988), the group laid out the rationale for single-patient trials, provided guidelines for their implementation, and described the establishment of a formal n-of-1 trial service within their teaching hospital. A pharmacist prepared capsules of active drug and matching placebo and retained the treatment "code"; all other team members (in addition to the patient and treating physician) were blinded to the treatment allocation. Outcomes were monitored at predetermined intervals, and the trial continued as long as the patient and physician agreed that more information was needed. (Facilitating a dialogue and partnership between patient and physician was a critical, if understated, component of the service.) A tutorial service was also established to teach clinicians interested in conducting their own n-of-1 trials.

By decade's end, the McMaster group had launched seventy trials (each involving one patient) and completed fifty-seven. Of the fifty-seven completed, fifty provided a definite clinical or statistical answer, and fifteen resulted in a major change in therapy (Guyatt et al. 1986, 1990b). The results generated tremendous enthusiasm in the academic medical community. The Canadians did not have to proselytize. According to Guyatt, "The idea sold itself."

### *An Early Adopter*

One place where the idea took root quickly was the University of Washington (UW). Eric Larson, now a professor of medicine at UW and the executive director of the Group Health Center for Health Studies in Seattle, was an assistant professor in the 1980s. He remembers attending a meeting of the American Federation for Clinical Research in Washington, D.C., where Guyatt presented his work.

I was sitting there as a junior faculty member, and Gordon Guyatt got up and presented his n-of-1 trial abstract. And I was just amazed. I thought this is fabulous. We spend so much effort on precision in diagnosis, yet we have very little precision about treatment. And it looked to me like this was a method that might do it. And the case that he presented was very dramatic, a pulmonary case that just defied explanation until you realized that the person probably had aspiration, and therefore the theophylline was making him worse.

Larson returned home and started talking up the idea. Together with a pharmacist, a statistician, and one of the medical chief residents, he obtained a small grant from the National Center for Health Services Research, developed a clinical n-of-1 trial service, and ran it for two years. It was a productive period: Larson and his colleagues ran thirty-four single-patient trials and published their results in the *Journal of the American Medical Association* (Larson, Ellsworth, and Oas 1993). Of the thirty-four trials, sixteen favored active treatment and eighteen supported a decision to discontinue treatment (or not start it in the first place). In their report, Larson's group grappled with a number of issues that continue to stir debate, including how many episodes (periods) of treatment to test, which randomization schemes to use, how to select outcomes, and how to determine "success."

With the help of statistician Alice Arnold, Larson's group specified the number of treatment periods using an informal Bayesian approach based on the degree of uncertainty concerning drug effectiveness (from the standpoint of both the literature and the referring physician). Drugs with a high probability of success (and less variability in response) required fewer treatment periods than did drugs with a more equivocal record. As for randomization, Larson's group chose a modified, non-pairwise approach, such that each treatment period was individually randomized to treatment or control. The only constraint was that all

patients receive each of the two alternative treatments at least twice during six treatment episodes. For outcome measures, the team favored 7-point Likert scales linked to the outcomes that patients deemed most important (e.g., ease of climbing the bedroom stairs). Superiority of treatment versus control was defined as a 0.5 point or greater absolute difference (on a 7-point scale) in the mean outcome.

Most clinicians reported little inconvenience from participating, and 85 percent said they would refer other patients to the service. Most patients found the experience useful (79 percent) and would participate in another such trial (63 percent). Resource use was estimated at \$400 to \$500 in 1990 dollars (\$673 to \$841 in 2008).<sup>1</sup> Sustainability was a problem, however. After the grant ran out, the single-patient trial service at the University of Washington dissolved. As Larson asked: “The question really is—how many patients are there that really want to know this? And how many doctors are there . . . to promote this to patients? . . . There are an awful lot of people who just want you to tell them what to take, and they’ll do it.”

The McMaster and University of Washington experiences shared a common trajectory. At each institution, the recipe for initial success included dynamic leadership, a multidisciplinary team, and a focused investment of resources. At UW, Larson had the added advantage of being the associate dean for clinical affairs from 1989 on. In both places, the effort proved difficult to sustain: demand was lukewarm, the business case was weak, and other research opportunities beckoned. Nevertheless, the n-of-1 idea refused to die. Long after shutting down their formal operations, both Guyatt and Larson continued to receive phone calls from investigators interested in establishing their own single-patient trial services. During most of the 1990s, few came to fruition. By late in the decade, however, single-patient trials began to experience a modest renaissance.

### *The Second Generation*

Located on the arid northeastern shore of Australia, the University of Queensland is one of Australia’s premier research institutes. In 1999, C. Jane Nikles and her colleagues established a single-patient trial service focusing on attention deficit/hyperactivity disorder (ADHD) and osteoarthritis (Nikles, Clavarino, and Del Mar 2005; Nikles et al.

2000a, 2000b, 2006, 2005). The service began as a regional resource in Queensland but rapidly evolved into a national mail-order operation that was effectively promoted through television advertising. This direct-to-consumer approach generated extensive interest and circumvented the recruiting obstacles confronting the other n-of-1 trial services. The investigators developed standardized outcomes measures and established mechanisms for remotely enrolling patients, obtaining informed consent, distributing active treatments and placebos, collecting and analyzing outcomes data, and providing feedback on results. By 2007, the Queensland group had conducted several hundred single-patient trials. Aside from minor funding contributions from hospitals and pharmaceutical companies, this service has been subsidized by federal grants.

More recently, Sunita Vohra and colleagues set up an n-of-1 service for pediatric patients at the University of Alberta. The service focuses principally on complementary and alternative medicine (CAM). The first consultations examined the use of probiotics in diarrhea-predominant irritable bowel syndrome, melatonin for sleep in attention deficit disorder, acupuncture for chemotherapy-induced vomiting, and homeopathy for eczema. Conceived in 2003 and initially funded in 2004, the University of Alberta service took another two years to become fully established, as the leadership sought approval from the Health Research Ethics Board (HREB), the College of Physicians and Surgeons of Alberta, and Health Canada. An important component of the negotiation concerned whether n-of-1 trials are “research” (and therefore subject to human subjects regulations concerning approval, consent, and oversight) or “clinical care” (exempt from such regulations but still subject to oversight by the usual clinical authorities). Vohra was relentless in arguing that n-of-1 trials were an enhanced form of clinical care. As she described it:

The n-of-1 approach can be used in a research study or as a clinical service. We're a clinical service. We use research methods; we use randomization and blinding; but we accept consults. Because it's a clinical service, we're also very comfortable individualizing things for the patient. We don't feel the need for rigid inclusion and exclusion criteria and a rigid way of determining this many periods, that many pairs. We think that part of what a clinical n-of-1 service offers is the idea of bending research methods around the patient, bending the outcome assessment around the patient, and so we try to remain highly flexible as you would in any clinical setting that is focused on the patient's needs.

Like the n-of-1 trial services at McMaster and the University of Washington, Vohra's service is based at an academic medical center and faces many of the same threats to sustainability, including difficulty maintaining physicians' enthusiasm and stable funding. However, several features of the University of Alberta service provide grounds for optimism: the service has a relatively narrow clinical focus on pediatric CAM and appears to have the support of local clinicians, who view it as a valuable resource for answering certain clinical questions. In addition, by virtue of its hard-won classification as a clinical service (rather than research), it is exempt from many of the more onerous mandates that would otherwise govern its conduct. And at least for the time being, it has energetic and committed leadership.

Whether through skill, luck, or sheer tenacity, Vohra succeeded where others failed in convincing the HREB that employing research techniques (randomization, blinding, objective outcome assessment) does not automatically translate into "research." Ethics boards have tended to insist that single-patient trials are research and thereby subject to case-by-case review of individual protocols and strict requirements for written informed consent. Part of the reason may be that the directors of n-of-1 trial services generally try to publish their findings; intent to publish is part of the Common Rule definition of research promulgated by the National Institute of Health. (Of course, some common forms of publication such as case reports and case series do not necessarily produce "generalizable knowledge," and the same could be said of single-patient trials. But this may be too subtle a distinction for most ethics boards.)

Janine Janosky, a statistician and director of a primary care research network based at the University of Pittsburgh, has collaborated in dozens of single-patient trials; in every single case "they have been considered research." Janosky thinks that whatever the trigger—randomization, the use of formal outcomes measures, the use of the word *trial*—the effects are powerful. "I've tried different arguments. Tried calling it quality assurance, tried another approach. But I haven't been able to convince anybody." In approaching the HREB, Vohra's team carefully laid out the components of the n-of-1 approach and argued successfully that well-designed and well-executed n-of-1 trials can be simply an enhanced form of clinical care.

For the most part, n-of-1 trials have remained firmly tethered to academic medical centers, which are able to leverage the necessary clinical, pharmacy, and statistical expertise. But it was only a matter of time



before enthusiasts with a more entrepreneurial bent caught the fever. Almost a decade after Larson's seminal publication and around the same time that Nikles was writing up the Queensland experience, Fred Huser and John Rothman developed Opt-e-Scrip, with the mission of bringing n-of-1 trials to the masses. Applying their extensive experience to pharmaceutical product development, they patented a technology that became the foundation for a commercial n-of-1 service (Reitberg 2001, 2003). Founded in 2001, the service provided comprehensive support for conducting n-of-1 trials for patients with allergic rhinitis, osteoarthritis, or gastroesophageal reflux. Prefabricated kits (costing from \$75 to \$310, depending on the drugs being tested) included blinded treatments and questionnaires that were completed by the patient and returned to the company for analysis. Upon receiving the patient's data, Opt-e-Scrip prepared an individualized report and delivered it promptly to the treating physician. Working with Wilson Pace, a Denver-based family physician-researcher and director of the American Academy of Family Physicians National Research Network, the company launched a collaborative clinical trial examining the concept, to conduct a comparison of n-of-1 trials to standard practice. The trial was aborted, however, when Opt-e-Scrip was able to raise only half of the capital required. Dr. Pace is still a believer, though: "They're just undercapitalized. You know, they've got the right idea; they just need somebody with a little more capital to step up, and then I think they'll get it right."

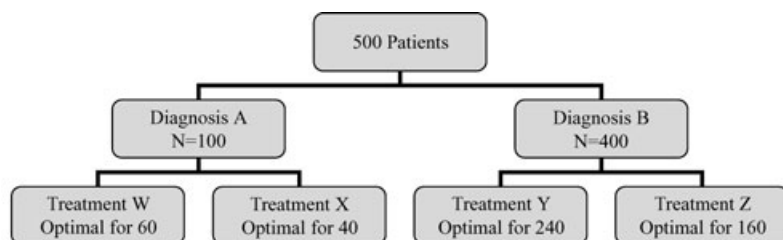
## The Promise of Therapeutic Precision

N-of-1 trials require substantial time, effort, and resources. Still, it is reasonable to wonder why society is willing to invest billions of dollars for technology that helps us get the right diagnosis but has not embraced an approach that can dramatically increase the likelihood that a patient gets the right treatment. As Eric Larson described the situation: "We send people to get X-rays for diagnostic precision, but we don't pay nearly as much attention to therapeutic precision. We'll put somebody on a drug for five years and not think too much of it."

Larson's point about therapeutic precision can be illustrated with a hypothetical example. Assume that among 500 patients presenting with a particular constellation of symptoms and signs, 100 have Diagnosis A and 400 have Diagnosis B. Assume further that among the 100 patients

with Diagnosis A, 60 will respond better to Treatment W than to other treatments, and 40 will respond better to Treatment X (with no ties allowed). Of the 400 patients with Diagnosis B, 240 will respond better to Treatment Y and 160 will respond better to Treatment Z (see figure 1, top panel).

Now consider four treatment strategies (figure 1, bottom panel). The proportion of patients treated successfully is least using the “One Size



Strategy	Description	Number of Successes (%)
One size fits all	All patients assumed to have the condition with the highest prior probability, absent any diagnostic testing; all patients treated with the therapy that works best (on average) for patients with the most prevalent condition.	240 (48)
Diagnostic precision (testing)	All patients undergo a diagnostic test with both sensitivity and specificity equal to 90%. Based on the results, patients are assigned to the most likely diagnostic category and treated with the therapy with the highest average effectiveness for that condition.	270 (54)
Therapeutic precision (single patient trial)	All patients assumed to have condition with the highest prior probability, absent any diagnostic testing; all undergo an n-of-1 trial with sensitivity and specificity of 80%, comparing treatments reserved for that condition.	320 (64)
Doubly precise	Patients undergo diagnostic testing followed by an n-of-1 therapeutic trial.	360 (72)

FIGURE 1. Distribution of a Hypothetical Panel of Five Hundred Patients by Diagnosis and Treatment Response (Top Panel) and Probability of Success (Bottom Panel)

Fits All Strategy” (48 percent), greater using the Diagnostic Precision Strategy (54 percent), greater still employing a Therapeutic Precision Strategy incorporating n-of-1 trials (64 percent), and best of all using a Double Precision Strategy (testing to elucidate the diagnosis followed by an n-of-1 trial to establish optimal therapy) (72 percent).

By directly addressing the heterogeneity of treatment effects across individuals, n-of-1 trials support the clinician’s natural instinct to individualize care, an approach that may benefit the patient but that finds little support from evidence-based guidelines, rules, regulations, and incentives aimed at diminishing practice variation (Walter et al. 2004). But as the example suggests, a Therapeutic Precision Strategy could yield significantly better patient outcomes, with or without concomitant efforts to enhance diagnostic precision.

### *Balancing Individual and Aggregate Data*

Of course, as Guyatt and Larson grasped intuitively when they varied the number of n-of-1 treatment episodes according to the clinical context, the results of an n-of-1 trial should be interpreted in light of aggregate data available from other sources (Senn 1991, 1993, 1998, 2001, 2003; Senn, Bakshi, and Ezzet 1995). In the most pristine case, when data from comparable patients participating in similar n-of-1 trials suggest that one treatment is clearly superior to another, with consistent effects across individuals, only a small number of treatment episodes will be required to confirm the superiority of that treatment for an individual. The reason is that aggregate data from other patients can be borrowed to supplement the individual’s own results. N-of-1 trial services provide an opportunity to combine individual and aggregate data using empirical Bayes methods (Schluter and Ware 2005; Senn, Bakshi, and Ezzet 1995; Zucker et al. 1997). In particular, the empirical Bayes estimate for the individual treatment effect is usually a weighted average of the estimated individual treatment effect based entirely on individual data from the patient’s own n-of-1 trial plus the estimated average treatment effect based on aggregate data pooled across the results of n-of-1 trials among comparable patients. The weight is determined by the variance components within individuals and between individuals. If, for an n-of-1 trial with a large number of crossovers, the *within-individual* variation is small (suggesting little random variation in outcome over time), the

empirical Bayes estimate will be dominated by the individual's own data. If the *between-individual* variation is small, the empirical Bayes estimate will be dominated by the aggregate data. In this situation, the average treatment effect is a good proxy for the individual treatment effect. A similar approach could conceivably be used to incorporate aggregate data based on the research literature to guide the interpretation of an individual patient's n-of-1 trial results, assuming that the individual patient data and literature-based aggregate data are derived from the same population, thereby allowing for a common prior distribution, which is a prerequisite for the empirical Bayes methods.

### Barriers to Implementation and Dissemination

Given their early success and demonstrable potential to improve therapeutic precision, why have single-patient (n-of-1) trials not been adopted more widely? Robust uptake would require the buy-in of health care organizations, clinicians, and patients, each with particular reasons for not running to the party. For want of a credible business case, health care organizations have not energetically invested in n-of-1 trial support services. Academic medical centers, for their part, have been perfectly willing to establish n-of-1 trial services with the support of extramural funding but have been reluctant to invest their own resources. One solution is for n-of-1 trialists to streamline operations, create economies of scale, and identify partners who will reap some of the clinicoeconomic benefits of n-of-1 trials. Potential partners could include managed care organizations, health insurers, and large employers.

For clinicians, the barriers are conceptual, practical, psychological, and structural. Most of our informants noted difficulties in educating doctors about the clinical value of single-patient trials and convincing them that participation is worth the effort. Single-patient trials harness clinical care to some of the approaches of clinical research (e.g., randomization, blinding, standardized measures). The dual nature of n-of-1 trials (incorporating research and clinical care) can challenge the most acute intelligence, even (and perhaps especially) experienced academic clinician-scientists. The confusion extends to the ethical realm, where questions are raised about the appropriateness of performing "research" on treatments already known to be effective. Few appreciate that in

contrast to traditional RCTs, n-of-1 trials can still be conducted ethically even in the presence of certainty about average effectiveness. In addition, every patient enrolled in an n-of-1 trial stands to benefit personally, if only by learning that an expensive or poorly tolerated therapy is doing less good than previously believed.

Conceptual barriers aside, n-of-1 trials make a number of practical demands. Clinicians must explain the idea to their patients, see them regularly throughout the trial, and evaluate the results jointly at the trial's end. With clinical agendas expanding and reimbursement flat (Bodenheimer 1999, 2006), taking the time to sell an unfamiliar concept—let alone follow through on the logistics—might be untenable. But beyond a simple calculus of time and effort, there are probably deeper psychological reasons for physicians' resistance. While clinicians understand that the practice of medicine involves uncertainty, they are not necessarily comfortable with it. Recommending that a patient enroll in an n-of-1 trial acknowledges this uncertainty, suggesting that the physician does not really know what is best. It is much easier—and arouses far less cognitive dissonance—to simply write a prescription. As Guyatt summed it up: “[Physicians] tend to have ways that they've learned to operate where they're comfortable, and it's hard to move outside these ways. It's hard to do something extra that's going to take time, and energy, and initiative.”

More broadly, the ideology and structure of medical practice in the United States favor making the correct diagnosis over recognizing patient variability and perfecting therapeutic regimes over time. As Frank and Zeckhauser suggest,

Our assessment is that there is much ready-to-wear treatment supplied by physicians. In many cases this appears to be a sensible response to a complex decision making environment. Nevertheless, there is also evidence of the supply of ready-to-wear treatment that is based on idiosyncratic physician-specific preferences or severe biases in the application of heuristics. (Frank, Zeckhauser, and National Bureau of Economic Research 2007, pp. 44–45)

Patients might seem natural enthusiasts for n-of-1 trials, the only form of research that truly focuses on “patients like me” (Brookes et al. 2007). Understanding and addressing patients' perspectives on n-of-1 trials is critical, because in the absence of external funding, institutions

will not support n-of-1 trial services without a solid business case that reflects pent-up demand, and clinicians will not refer patients to n-of-1 trial services unless they perceive that their patients will not only benefit from trial enrollment but also be grateful for the opportunity. The most ready explanation for the apparent lack of consumer demand is that with the exception of the Queensland program, they have not been asked. N-of-1 trials make significant demands on patients: they may need to complete special consent procedures, make extra office visits, and fill out surveys and diaries. Patients also incur the opportunity cost of taking a potentially ineffective therapy for half the trial, which may last for weeks. Finally, patients who enroll in n-of-1 trials risk relinquishing placebo effects: since 30 to 40 percent of intervention effects may result from placebo (Benson and Friedman 1996), learning that a long-cherished treatment “doesn’t work” might be harmful. Whether patients are aware of these factors is a subject worthy of additional research.

The Australian experience (as well as the remarkable success of direct-to-consumer advertising of prescription drugs in the United States) (Wilkes, Bell, and Kravitz 2000) suggests a possible role for direct marketing of n-of-1 trials to patients, who may value knowing that the costs, risks, and side effects of a particular medicine are “worth it.” As patients continue to assume a larger share of pharmaceutical costs, they may find that n-of-1 trials offer a way to reduce expenditures while preserving access to the treatment that is best for them. In addition, marketing n-of-1 trials directly to patients could be more effective than trying to work solely through beleaguered physician intermediaries. N-of-1 marketing could be organized as stand-alone campaigns or combined with other efforts to enhance health literacy and numeracy (Schwartz, Woloshin, and Welch 2007). Properly executed, a consumer-oriented n-of-1 campaign would not circumvent physicians but, rather, would prepare patients to discuss the n-of-1 approach with their doctors in the spirit of shared decision making (Makoul and Clayman 2006).

A more comprehensive approach to addressing these questions is suggested by the literature on diffusion of innovation (Rogers 2003). Rapidly adopted innovations have a clear relative advantage, are compatible with current practice, are relatively simple, can be tried out on a limited basis, and have results clearly observable to others. By this metric, n-of-1 trials succeed on some levels and fail on others. As we have discussed, n-of-1 trials offer a clear benefit in terms of enhanced therapeutic precision, but they also convey costs in terms of patient,

clinician, and system time and resources. While technically compatible with both traditional “opinion-based care” and the more recently heralded “evidence-based medicine,” single-patient trials may be perceived as a paradigm shift and thus incompatible with current practice. The innovation is relatively simple (the required technology—some placebo capsules, a few survey forms, a personal computer—is fairly inexpensive), but perhaps not simple enough. The innovation is scalable—in fact, the history of n-of-1 trials to date can be thought of as a kind of demonstration project that is ripe for expansion. Finally, the results are certainly observable to others, as evidenced by a litany of published reports of n-of-1 trials. But as Gordon Guyatt epitomized the issue, “It turned out to just be too much trouble for people.”

### What Would Make N-of-1 Trials “Tip”?

Malcolm Gladwell’s bestselling book *The Tipping Point* argues that products and ideas “tip,” that is, become part of the popular consciousness and culture when they fill a clear need and are championed by the right people (whom he calls connectors, mavens, and salespeople). N-of-1 trials are nowhere near the tipping point. To deliver on their potential will require either a shift in the ratio of perceived benefits to burdens (in Rogers’s parlance, an increase in their relative advantage) or a new set of stakeholders for whom the gains more clearly outweigh the losses. Fortunately for those who think the methodology offers something unique and important, both are possible.

One way to tip the cost-benefit scales would be to create systems that minimize the time, effort, and resources required from patients, doctors, and health care organizations. As described earlier, Opt-e-Scrip attempted to do just that, making n-of-1 trial enrollment as simple as obtaining the patient’s consent, writing a “prescription,” and faxing it to Opt-e-Scrip’s corporate headquarters. While the company succeeded in convincing the American Medical Association to establish a Current Procedural Terminology (CPT) code for “personalized medicine tests” (0130T), its ultimate success remains in question. The Queensland experience (involving mail-order trials throughout eastern Australia) has been more consistently positive but owes much to sustained government funding and an effective television advertising campaign that appealed to consumers.

“Tipping” requires both a core of enthusiasts and structures and mechanisms to convey that enthusiasm. Practice-based research networks (PBRNs) could conceivably provide both. First established in the 1970s, these loose affiliations of primary care clinicians have grown because of the funding initiatives launched under the auspices of the Agency for Healthcare Research and Quality (AHRQ) and the National Institutes of Health Clinical and Translational Science Awards (CTSAs). PBRNs are the culmination of a long research tradition in primary care pioneered by figures such as James Mackenzie, Will Pickles, John Fry, and Curtis Hames and emphasizing the careful observation of individual patients pieced together to build generalizable knowledge (Green and Hickner 2006). Having established their interest in asking and answering questions that arise in practice, PBRN members might be more likely than the average clinician to embrace n-of-1 trials as an opportunity to apply research methods in the care of individual patients.

It is also conceivable that physicians could be converted as the conceptual wall between clinical care and research in single-patient trials slowly cracks. After all, the therapeutic trial (trying out a medication for days to weeks to “see if it works”) is already an integral part of medicine. Both physicians and patients may come to see that the best predictor of a long-term response to therapy is a weighted average of aggregate experience combined with the results of a carefully conducted n-of-1 trial. Marketing such trials to clinicians will likely involve convincing them that these studies are a simple extension of what they already do. Ironically, community-based practitioners may be easier to engage than research-oriented academics, who tend to fixate on the “low power” of n-of-1 trials and have trouble mustering enthusiasm for an approach that does not, by itself, produce generalizable knowledge.

Another possible way to make n-of-1 trials more acceptable to practitioners is to strip them down to their essentials. In clinical practice, physicians use evidence and experience to generate a list of treatment options. Patients and physicians move down the list based on trial and error. By dropping the requirement for blinding and relaxing some of the randomization procedures, it might be possible to retain some of the benefits of n-of-1 trials with fewer burdens while improving on the informal therapeutic trials that patients and doctors conduct every day. For example, in a single-patient head-to-head trial of (generic) omeprazole versus Nexium<sup>®</sup> for acid reflux, considerable information might be gleaned by simply alternating the two medications (without blinding)



every fortnight for a total of eight to twelve weeks and asking patients to keep detailed symptom diaries. Research is needed to determine whether the reduction in costs and burden and the gain in acceptability from such diluted designs would be worth the reduction in scientific rigor.

While reform from within holds some promise, three ongoing changes in the external scientific, business, and policy environments may also augur well for the n-of-1 movement. First, the escalating cost of drug trials means that pharmaceutical companies are under increasing pressure to make the right clinical research design choices up front. It is no longer tenable to make consequential decisions about entry criteria, dose, comparators, trial duration, and outcomes based on underpowered Phase I and Phase II studies. Notwithstanding legitimate concerns about the handling of within-patient variation (Senn 1993), for some conditions and treatments a relatively small number of n-of-1 trials could yield considerable information about the heterogeneity of drug onset (interval between treatment administration and treatment effect) and offset (interval between cessation of treatment and cessation of effect), the treatment effects themselves, and the side effects (Guyatt et al. 1990a).

Second, the expanding market for biopharmaceuticals has raised the stakes for health care payers and purchasers. These agents include a number of large molecular entities targeted to cancer, multiple sclerosis, Crohn's disease, rheumatoid arthritis, and more. They often are hugely effective but very expensive. As new biologics are developed and new indications are identified, pressure will grow to direct these therapies to the patients deriving the greatest benefit. In response to the high costs of biologics, insurers and managed care organizations have invoked "stepped care," a strategy that requires physicians and patients to try more established and less costly treatments before turning to biologics (Hochberg, Tracy, and Flores 2001; Rich et al. 2004). The stepped-care approach to chronic disease management has both clinical and economic appeal, but opposition has arisen from patient advocacy groups, drug manufacturers, and even legislators (Kravitz, Duan, and White 2008). As additional, very expensive chronic agents come to market, payers will begin to demand evidence that they work, not just on average but in the individual patient. A recent cost-minimization analysis showed that offering an n-of-1 trial option as an alternative to stepped care for costly biologics could be cost saving (Kravitz, Duan, and White 2008). This creates a niche for n-of-1 trials. One can imagine a situation in

which patients will be told, “We’ll pay for the drug, but only if you enroll yourself in an n-of-1 trial to demonstrate that it works for you.” Such incentive schemes will raise ethical questions, but they are no more ethically compromised than many current practices, including formulary restrictions and stepped-care requirements.

Third, the emerging field of personalized medicine is resulting in therapies that are “custom built” for individual patients (Hoffman 2007). Ultimately, these therapies will need to be tested in the individuals for whom they were designed. One can readily imagine designer therapies for chronic, symptomatic conditions being vetted in single-patient trials. Thus, future support for n-of-1 trials may come from an unlikely alliance of old-school pharmaceutical manufacturers (who need to find more efficient ways to test drugs), public and private payers (who need to control costs and waste), and biotechnology companies (who need to demonstrate the efficacy and safety of custom-designed therapies in the individuals for whom they were designed).

### Harnessing the Demand for Therapeutic Precision: Marketing Directly to Patients

In a 2004 *Wired* magazine article, Chris Anderson coined the term the *long tail* to describe how the Internet was suddenly making it possible to make big profits in very small markets. The tail in question is the long, sloping right-hand side of a highly skewed histogram, where relatively few units of any one product or service are sold. Traditionally, companies have focused on the “fat” part of the curve where most of the buyers are; they could not afford to offer items with limited appeal. That all changed with the Internet, which allows companies like Amazon.com (books), Apple (music), and Netflix (movies) to move their products for a mere fraction of the former cost. At the present time, consumer demand for therapeutic precision in any one clinical area is not high. But in the aggregate, enough patients may have enough interest in knowing which medicine is really best for them to make a commercial n-of-1 trial service (or a nonprofit version sponsored by PBRNs) viable.

Consider, for example, Mr. Ransom, a forty-five-year-old businessman with chronic heartburn. His symptoms have been well controlled with Nexium<sup>®</sup>, a powerful prescription stomach acid blocker, but his insurance company will no longer pay for it. He checks out the prices at

drugstore.com and finds that he can obtain Nexium for \$4.98 per pill. Based on a close reading of the evidence, his doctor thinks that over-the-counter generic omeprazole (\$0.70 a pill) might work just as well. The patient tries it for a week. He seems to experience more heartburn than before but wonders whether this is due to stress at work or possible dietary indiscretions. Is it worth the extra \$1,562 per year to stay on Nexium for the long term? One day while shopping for shaving supplies at his local HealthCo Pharmacy, he sees a sign: "Is Your Prescription Working? Inquire Inside." Intrigued, he talks with the pharmacist and learns that for a small fee, HealthCo will provide a personalized test of drug effectiveness: an n-of-1 trial. The pharmacist calls the doctor, obtains a verbal order, and gives the patient a bag containing six sequentially numbered bottles, each containing seven identical-appearing pills. Mr. Ransom is instructed to take one pill each day from the "bottle of the week," recording his heartburn symptoms (and any potential side effects) on a special Scantron form. He goes home, follows instructions, and, six weeks later, returns to the pharmacy. The symptom-rating forms are scanned in, and a report is prepared. While there was some fluctuation in symptoms, overall the patient did as well on omeprazole as on Nexium<sup>®</sup>. The report is faxed to his doctor. During an office visit the following week, Mr. Ransom and his physician review the n-of-1 trial results together and decide to proceed with omeprazole.

If replayed across the long tail of individuals with chronic conditions who are uncertain about the value of their current medication regimen or the potential benefit of new agents, this scenario could generate substantial demand for single-patient trials. In other words, to emerge as a sustainable model for clinical care, n-of-1 trials may need to cast off some of their academic trappings and focus on appealing to what patients want and need.

## Conclusions

Over the past twenty years, n-of-1 trial services have been launched with great fanfare, only to recede into the academic woodwork once external funding support was withdrawn. Yet these trials have the potential to play an important role in the promotion of evidence-based, patient-centered care. The survival of the n-of-1 movement depends on finding a successful business model. Stripping n-of-1 trials down to

their essentials, marketing trials directly to patients as a component of good clinical care, and incorporating such trials into the work of PBRNs would represent an important start. In the longer term, drug makers, managed care organizations, and the biotechnology industry could become allies in the successful dissemination of this uniquely valuable bridge between evidence-based medicine and individualized care.

## Endnote

1. Adjusted for inflation using Inflation Calculator. Available at <http://stats.bls.gov/cpi/> (accessed August 29, 2008).

## References

- Barlow, D.H., and M. Hersen. 1984. *Single Case Experimental Designs: Strategies for Studying Behavior Change*. 2nd ed. New York: Pergamon Press.
- Benson, H., and R. Friedman. 1996. Harnessing the Power of the Placebo Effect and Renaming It “Remembered Wellness.” *Annual Review of Medicine* 47:193–99.
- Bodenheimer, T. 1999. The American Health Care System—Physicians and the Changing Medical Marketplace. *New England Journal of Medicine* 340:584–88.
- Bodenheimer, T. 2006. Primary Care—Will It Survive? *New England Journal of Medicine* 355:861–64.
- Brookes, S.T., L. Biddle, C. Paterson, G. Woolhead, and P. Dieppe. 2007. “Me’s Me and You’s You”: Exploring Patients’ Perspectives of Single Patient (N-of-1) Trials in the UK. *Trials* 8:10.
- Frank, R.G., R. Zeckhauser, and National Bureau of Economic Research. 2007. *Custom Made versus Ready to Wear Treatments; Behavioral Propensities in Physicians’ Choices*. Cambridge, Mass.: National Bureau of Economic Research.
- Green, L.A., and J. Hickner. 2006. A Short History of Primary Care Practice-Based Research Networks: From Concept to Essential Research Laboratories. *Journal of the American Board of Family Medicine* 19:1–10.
- Greenfield, S., R. Kravitz, N. Duan, and S.H. Kaplan. 2007. Heterogeneity of Treatment Effects: Implications for Guidelines, Payment, and Quality Assessment. *American Journal of Medicine* 120:S3–9.

- Guyatt, G.H., A. Heyting, R. Jaeschke, J. Keller, J.D. Adachi, and R.S. Roberts. 1990a. N of 1 Randomized Trials for Investigating New Drugs. *Controlled Clinical Trials* 11:88–100.
- Guyatt, G.H., J.L. Keller, R. Jaeschke, D. Rosenbloom, J.D. Adachi, and M.T. Newhouse. 1990b. The N-of-1 Randomized Controlled Trial: Clinical Usefulness. Our Three-Year Experience. *Annals of Internal Medicine* 112:293–99.
- Guyatt, G.H., D. Sackett, J. Adachi, R. Roberts, J. Chong, D. Rosenbloom, and J. Keller. 1988. A Clinician's Guide for Conducting Randomized Trials in Individual Patients. *Canadian Medical Association Journal* 139:497–503.
- Guyatt, G.H., D. Sackett, D.W. Taylor, J. Chong, R. Roberts, and S. Pugsley. 1986. Determining Optimal Therapy—Randomized Trials in Individual Patients. *New England Journal of Medicine* 314:889–92.
- Hochberg, M.C., J.K. Tracy, and R.H. Flores. 2001. “Stepping-Up” from Methotrexate: A Systematic Review of Randomised Placebo Controlled Trials in Patients with Rheumatoid Arthritis with an Incomplete Response to Methotrexate. *Annals of the Rheumatic Diseases* 60(suppl. 3):iii51–54.
- Hoffman, E.P. 2007. Skipping toward Personalized Molecular Medicine. *New England Journal of Medicine* 357:2719–22.
- Kazdin, A.E. 1982. *Single-Case Research Designs: Methods for Clinical and Applied Settings*. New York: Oxford University Press.
- Kratochwill, T.R. 1978. *Single Subject Research: Strategies for Evaluating Change*. New York: Academic Press.
- Kravitz, R.L., N. Duan, and J. Braslow. 2004. Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages. *The Milbank Quarterly* 82:661–87.
- Kravitz, R.L., N. Duan, and R.H. White. 2008. N-of-1 Trials of Expensive Biological Therapies: A Third Way? *Archives of Internal Medicine* 168:1030–33.
- Larson, E.B., A.J. Ellsworth, and J. Oas. 1993. Randomized Clinical Trials in Single Patients during a 2-Year Period. *Journal of the American Medical Association* 270:2708–12.
- Makoul, G., and M.L. Clayman. 2006. An Integrative Model of Shared Decision Making in Medical Encounters. *Patient Education and Counseling* 60:301–12.
- Nikles, C.J., A.M. Clavarino, and C.B. Del Mar. 2005. Using N-of-1 Trials as a Clinical Tool to Improve Prescribing. *British Journal of General Practice* 55:175–80.
- Nikles, C.J., P.P. Glasziou, C.B. Del Mar, C.M. Duggan, A. Clavarino, and M.J. Yelland. 2000a. Preliminary Experiences with a Single-Patient Trials Service in General Practice. *Medical Journal of Australia* 173:100–103.

- Nikles, C.J., P.P. Glasziou, C.B. Del Mar, C.M. Duggan, and G. Mitchell. 2000b. N of 1 Trials. Practical Tools for Medication Management. *Australian Family Physician* 29:1108–12.
- Nikles, C.J., G.K. Mitchell, C.B. Del Mar, A. Clavarino, and N. Mc-Nairn. 2006. An N-of-1 Trial Service in Clinical Practice: Testing the Effectiveness of Stimulants for Attention-Deficit/Hyperactivity Disorder. *Pediatrics* 117:2040–46.
- Nikles, C.J., M. Yelland, P.P. Glasziou, and C. Del Mar. 2005. Do Individualized Medication Effectiveness Tests (N-of-1 Trials) Change Clinical Decisions about Which Drugs to Use for Osteoarthritis and Chronic Pain? *American Journal of Therapeutics* 12:92–97.
- Reitberg, D.P. 2001. *Method and Kit for Treating Illness*. Edited by USPTO. USA: Opt-e-Scrip.
- Reitberg, D.P. 2003. *Method and Kit for Treating Illnesses*. Edited by USPTO. USA: Opt-e-Scrip.
- Rich, S.R., I.C. Coleman, R. Cook, D.S. Hum, B. Johnson, T. Maves, W.J. Mazanec, J.R. Miller, W.J. Proveaux, H.S. Rossman, and W.H. Stuart. 2004. Stepped-Care Approach to Treating MS: A Managed Care Treatment Algorithm. *Journal of Managed Care Pharmacy* 10:S26–32.
- Rogers, E.M. 2003. *Diffusion of Innovations*. 5th ed. New York: Free Press.
- Schluter, P., and R. Ware. 2005. Single Patient (N-of-1) Trials with Binary Treatment Preference. *Statistics Medicine* 24:2625–36.
- Schwartz, L.M., S. Woloshin, and H.G. Welch. 2007. The Drug Facts Box: Providing Consumers with Simple Tabular Data on Drug Benefit and Harm. *Medical Decision Making* 27:655–62.
- Senn, S.J. 1991. Controlled Trials in Single Subjects. *British Medical Journal* 303:716–17.
- Senn, S.J. 1993. Suspended Judgment N-of-1 Trials. *Controlled Clinical Trials* 14:1–5.
- Senn, S.J. 1998. Applying Results of Randomised Trials to Patients. N of 1 Trials Are Needed. *British Medical Journal* 317:537–38.
- Senn, S.J. 2001. Individual Therapy: New Dawn or False Dawn. *Drug Information Journal* 35:1479–94.
- Senn, S.J. 2003. Author's Reply to Walter and Guyatt. *Drug Information Journal* 35:1479–94.
- Senn, S.J., R. Bakshi, and N. Ezzet. 1995. N of 1 Trials in Osteoarthritis. Caution in Interpretation Needed. *British Medical Journal* 310:667.
- Sidman, M. 1960. *Tactics of Scientific Research; Evaluating Experimental Data in Psychology*. New York: Basic Books.
- Walter, L.C., N.P. Davidowitz, P.A. Heineken, and K.E. Covinsky. 2004. Pitfalls of Converting Practice Guidelines into Quality Measures:

- Lessons Learned from a VA Performance Measure. *Journal of the American Medical Association* 291:2466–70.
- Wilkes, M.S., R.A. Bell, and R.L. Kravitz. 2000. Direct-to-Consumer Prescription Drug Advertising: Trends, Impact, and Implications. *Health Affairs* 19:110–28.
- Zucker, D., C. Schmid, M. McIntosh, R. D'Agostino, H. Selker, and J. Lau. 1997. Combining Single Patient (N-of-1) Trials to Estimate Population Treatment Effects and to Evaluate Individual Patient Responses to Treatment. *Journal of Clinical Epidemiology* 50:401–10.

---

*Acknowledgments:* The authors gratefully acknowledge the following n-of-1 trial pioneers who participated in interviews for this article: Gordon Guyatt, Eric Larson, Sunita Vohra, Jane Nikles, Janine Janosky, Fred Huser, and Wilson Pace. The research for this article was supported by a grant from Pfizer, but the authors assume sole responsibility for its content.