EXHIBIT L

Specialty Conference

N-of-1 Clinical Trials

A Technique for Improving Medical Therapeutics

Discussant
ERIC B. LARSON, MD, MPH, *Seattle*

This discussion was selected from the weekly Grand Rounds in the Department of Medicine, University of Washington School of Medicine, Seattle. Taken from a transcription, it has been edited by Paul G. Ramsey, MD, Associate Professor of Medicine, and Philip J. Fialkow, MD, Professor and Chair of the Department of Medicine.

ERIC B. LARSON, MD*: Most clinicians have a keen interest in therapeutics and especially therapeutic efficacy. In fact, medical therapeutics can be viewed as a series of therapeutic experiments as follows:

A				*B*
Initial State	→	Therapy	→	Subsequent State

The patient comes to the physician in an initial state, *A*, and is offered treatment. The patient then assumes a subsequent state, *B*.[1] If *B* is more desirable, we typically judge that therapy was effective. If *B* is no different or is less desirable, we judge that therapy made no difference or was ineffective. Although this account seems straightforward, such simple assertions may not be true because of confounding factors.[2]

Effectiveness may be overestimated because of several factors. First, a patient can recover spontaneously coincident with treatment, an especially well-known occurrence for self-limited conditions. Second, patients commonly present when their symptoms are worse, especially patients with a chronic disease. Coincidental treatment appears to cause the problem to subside when the patient has simply returned spontaneously to the average, so-called baseline state of a chronic disease. This has been referred to as "regression toward the mean."[3] A third factor that may lead to an overestimation of effectiveness is a placebo effect. For some therapies, as much as 30% or more of the benefits may be due to the well-known placebo effect.[4] Finally, the expectation of a beneficial response and a willingness-to-please effect[5] are related to the placebo effect. In many patients, the simple "expectation" that a treatment will be beneficial may often be sufficient to promote a beneficial effect. The willingness-to-please effect results from the so-called obsequiousness bias[5] in which a patient gets better to please an expectant physician.

Similar confounding forces can obscure therapeutic effectiveness. Coexistent illness can coincidentally exacerbate the underlying problem. Chronic diseases have spontaneous exacerbations, and when these occur coincident with treatment, it appears that therapy is ineffective. Malingering or a secondary gain in which the patient experiences benefit from not getting better can make a patient resistant to the true effect of treatment. An age-related (physiologic) decline superimposed on a beneficial treatment effect may combine to cancel each other. Finally, if an incorrect diagnosis has been made, treatment will appear to be ineffective. For example, if a patient's symptoms or signs represent the upper or lower limits of a normal variation, then the treatment received, although usually effective, is ineffective in the misdiagnosed case.

Randomized Clinical Trials

Fortunately, randomized clinical trials (RCTs) have been used to evaluate medical therapeutics since the late 1940s.[6] Because such trials help eliminate the confounding factors outlined above, they have become the gold standard by which clinicians judge therapeutic efficacy. An RCT allocates consecutive patients to different treatments or randomly allocates the order of treatment in crossover experiments. When done carefully with enough patients, the randomization eliminates bias that might confuse the interpretation of the therapeutic experiment.

Unfortunately, many of a clinician's day-to-day treatment decisions cannot be based on the results of randomized trials. Table 1 shows examples of situations or problems in which RCTs may not be appropriate for making therapeutic choices. Unavailability of randomized clinical trials may be encountered in the case of a rare or unusual disease. Randomized trials may also not be available for some older treatments and for newer or novel treatments. Because RCTs have been widespread only since 1970, older treatments were often not evaluated by them. Newer or novel treatments, especially those devised by clinicians for single patients, are typically not subjected to randomized trials.

Even when there are good randomized trials showing efficacy, several factors limit their generalizability to a specific patient. For example, the patient might be outside the eligibility requirements for entry into an RCT. Eligibility criteria for most trials are so restrictive that less than 10% of patients with the disease in question may be accepted. Not surprisingly, the patients who are excluded are the ones in whom therapeutic dilemmas and an evaluation of therapeutics are often the most troublesome. Thus, their omission

*Professor of Medicine, Department of Medicine, and Medical Director, University of Washington School of Medicine, Seattle.

(Larson EB: N-of-1 clinical trials—A technique for improving medical therapeutics [Specialty Conference]. West J Med 1990 Jan; 152:52-56)

This work is supported by grant 1 RO3 HS 6006 from the National Center for Health Services Research and Health Care Technology Assessment.
Reprint requests to Department of Medicine, RG-20, University of Washington School of Medicine, Seattle, WA 98195.

TABLE 1.—*Limits of Randomized Clinical Trials (RCTs) for Care of Individual Patients*

RCT unavailable or impossible
Good RCTs show benefit but may not be generalizable
Eligibility criteria too restrictive
Some patients are nonresponders
Side effects
Good RCTs show *no* benefit but may not be generalizable
Atypical patients
Treatment response is idiosyncratic

from RCTs allows investigators to assess efficacy with fewer complicating factors. Another problem arises from the fact that even though a randomized trial has shown efficacy, not all patients will benefit from treatment. In addition, some patients may experience enough side effects that the net effect of treatment is harmful. The single patient who does not have a beneficial response experiences that event with 100% certainty even when generalizations based on populations studied by RCTs indicate the net effects are likely to be beneficial.

There are also limits to the generalizability of RCTs that show no apparent benefit. Good randomized clinical trials may not show any net benefit, but an individual patient may still benefit from treatment, especially if the treatment has biologic plausibility. Some RCTs have inadequate sample sizes and, hence, inadequate statistical power to show efficacy.[7] An individual patient could also be an atypical responder, or responsiveness to treatment may be idiosyncratic and difficult to demonstrate by an RCT.

In summary, even though randomized clinical trials are widely used for assessing therapeutic efficacy, their results may not apply to single patients or they may be unavailable for certain treatments, thus leaving clinicians in a quandary about therapeutic efficacy. Because of this quandary, there is increasing interest in single-patient experiments. A number of terms have been used to describe single-patient experiments, including N-of-1 trials, single-patient clinical trials, single-case analysis, crossover and self-controlled research designs, and single-patient RCTs. The field has an interesting history and holds great promise for improving the science of medical therapeutics.

Case Reports

Because case reports can be useful ways to illustrate valuable clinical lessons, I will present three single-case analyses in the order of my exposure to them. The first, a "case report" presented at the American Federation of Clinical Research meetings in 1985, was the case that piqued my interest in single-patient trials.[3] The second, a classic case that occurred at the interface of the developing science of statistics and popular culture, is intriguing for both its contents and the statistical power of its design.[8] The final case illustrates a single-case clinical trial that, although not random and only "single blinded," was convincing and influential.[9]

The first case was reported by Guyatt and co-workers from McMaster University, Hamilton, Ontario.[3] The patient, a 65-year-old man with uncontrolled asthmatic bronchitis, was becoming progressively more disabled by dyspnea with even simple daily activities. His therapeutic regimen eventually consisted of albuterol inhaler, ipratropium bromide, theophylline, and daily doses of prednisone. The clinician and the patient were uncertain whether the theophylline or ipratropium therapy was beneficial. Both suspected that theophylline was helpful and ipratropium was not. To optimize the therapeutic regimen, a single-patient trial was designed. Either theophylline or placebo, in a random order, was given for ten-day crossover periods. Three 10-day crossover pairs were planned. The end points included dyspnea, the need for albuterol inhaler, and the amount of sleep disturbance. During the first period, the patient did better than during the second ten days of the crossover trial. The same pattern then appeared during the second crossover period. The trial, which was originally scheduled to go for three crossover periods (about 60 days), now seemed too long to both the clinician and the patient. Both agreed that the trial should be terminated, presumably to allow the patient to resume taking theophylline. They were surprised when the placebo was associated with scores indicating increased well-being. Based on a review of the literature and the patient's course, it was determined that the seemingly anomalous results were most likely explained by gastroesophageal reflux (a xanthine side effect) and aspiration.[10] The theophylline therapy was stopped, and subsequently an N-of-1 trial of ipratropium revealed the beneficial therapeutic effects of its use. Eventually the patient was treated with a regimen of albuterol and ipratropium. He then tolerated a prednisone taper so that he could comfortably complete most of his activities of daily living on a regimen of 10 mg of prednisone every other day.

The second "case report" is not a medical case but represents a particularly famous single-case experiment. The case was an important one in the development of principles of experimentation and illustrates some useful points about randomization and statistical power. In 1935, R. A. Fisher, a British statistician whose name is most often linked with multiple-subject experiments, reported an example of how to conduct an experiment with a single subject and used that example to explain basic notions that underlie all experiments. This was the "lady tasting tea experiment."[8]

The case involved a tea-drinking English woman who claimed that she could tell whether the tea was added to the milk or the milk was added to the tea. Four cups of tea were prepared one way and four cups the other way, and the eight cups were then presented to her in a random sequence. She was told in advance that she was to identify the four cups that were prepared each way. The lady correctly identified all eight cups, and the P value was determined by the randomization test procedure. The null hypothesis was that her response at any treatment time was the same as it would have been at that time if any of the other cups had been presented. There are $8!/4!4! = 70$ ways in which eight cups can be presented with respect to milk first or tea first, given that four cups were milk first and four tea first. Thus, Fisher computed the P value as 1/70 because only 1 of the possible sequences of 4 *M*s and 4 *T*s correctly matched the woman's responses ($P = .014$).

An important feature of this experiment, in contrast to the first case report, is that the randomization occurred in blocks of eight treatments, not blocks of two as in the typical crossover experiment. Thus, the statistical power was considerably greater.

The third case report is a more primitive example of a single-patient trial.[9] Nonetheless, it also shows the value of single-patient experimentation. The report entitled "Inter-

nal-Mammary-Artery Ligation for Coronary Insufficiency—An Evaluation" was based on a presentation made in 1957 to the New England Surgical Society. This topic would later be investigated in a widely quoted article from the University of Washington describing a randomized, single-blind trial that compared a sham operation with internal mammary ligation.[1] Ralph Adams, MD, in the 1958 paper,[9] reported four cases, one of which was of a 60-year-old man admitted "three days after occurrence of his known episode of coronary thrombosis."

> His case was well known to the hospital because of previous attacks of deep thrombophlebitis, pulmonary embolism and hypercholesterolemia, and prior episodes of coronary occlusion. Precordial pain was intense and he was apprehensive that he would die. He was a highly educated man, well informed for a layman, on medical matters and in a position of considerable community responsibility. Admission was for the specific purpose of altering internal mammary circulation in the hope of giving him some cardiac protection. He was told . . . that this procedure was currently being widely discussed and, in some quarters, enthusiastically recommended. He was also informed that the hospital was in the process of evaluating the procedure as definitely as possible. These background facts led him to request that the operation be tried in the hope that he might be helped. . . .
>
> At operation, on the day of admission, a short incision was made in the second intercostal space lateral to each sternal border and each internal mammary artery was exposed. A silk ligature was placed about each artery but neither was tied. Thus, only a first-stage operation had been done, consisting of a skin incision and encirclement but not ligation of the internal mammary arteries.
>
> On awakening from the brief and light anesthetic, the patient reported that he was free of pain. He has had no pain since that date. An electrocardiogram on the day after operation showed no detectable change from preoperative tracing. Two days after the operation the ligatures from the internal mammary arteries were tied. Subsequent electrocardiographic tracings gave no evidence of improvement.

The author goes on to describe follow-up, which included no recurrence of symptoms, and states that

> in this case, there was not a fair chance to assay the relief of symptoms to be obtained by internal mammary artery ligation because the patient lost all symptoms after the first portion of a staged procedure that he believed to be the completed operation.

Adams reported what we would call a nonrandomized single-patient crossover experiment. A sham operation was followed by a real operation—dramatically showing what many might now call a placebo effect of internal mammary exposure.

Formation of an N-of-1 Clinical Trial Service

Before establishing a single-patient trial service, we contacted Dr Gordon Guyatt, who has actively investigated single-patient trials. He provided us with great encouragement and a summary of the experience of an N-of-1-trials service at McMaster University.[2] Most of his trials had been in the subspecialties of pulmonary medicine and rheumatology. Of the first 42 trials done at the center, 29 gave definitive results. In 11, active treatment was found to be effective, in 17 it was ineffective, and in 1 it was harmful (the theophylline case). Eight other trials gave less definitive results. Five were judged unsuccessful, three because, despite definitive outcomes, the results did not lead to action (G. Guyatt, written communication, June 1987).

Based on this encouraging report, we submitted a small grant proposal to the National Center for Health Services Research. Our research group, which includes Allan Ellsworth, PharmD; Jim Nuovo, MD (family medicine); Ina Oppliger, MD (rheumatology); Gerald van Belle, PhD; and Alice Arnold, MS (biostatistics), is now funded to establish and evaluate a single-patient trial service. We have announced our intentions to workers in other specialties and are currently receiving patients.

Because the objective of the "N of 1" experiment is to find the best treatment for a particular patient, we and others believe that some of the ethical questions asked of the standard randomized trial no longer apply.[2] For example, does the potential benefit to other patients outweigh the possible risk to this patient? Nonetheless, three ethical requirements do apply. First, a patient's free and informed consent should be requested after the clinician has described every feature of the trial that would materially affect the patient's decision to take part, including the reported effectiveness and safety of alternative treatments, the treatment targets to be used, and the duration and number of treatment periods to be executed. The second ethical requirement is that a patient must be free to withdraw at any time without loss of care. The third is that the same degree of confidentiality applied in other clinical situations must apply to the study results. One of our first tasks as an N-of-1 clinical trial service was to approach the Human Subjects Committee (Institutional Review Board) and seek approval for pending single-patient trials. They have developed an expedited approval process that facilitates the prompt institution of clinical trials.

When to Do a Clinical Trial

Perhaps the most germane issue in single-patient trials is when to do them. That is, when is a patient most likely to benefit from the results of a single-patient trial? The most important issue here is whether there is doubt about efficacy. Doubt may occur because neither the patient nor the physician is certain an existing treatment is working. In this setting, a patient with a chronic disease may be doing poorly or not improving on a medication regimen that could also be causing side effects, as exemplified by the theophylline case.

Another instance when efficacy may be in doubt is during the institution of a new treatment. Here the patient is being offered a new drug and the question is, "Will it work?" The clinician may be uncertain when the literature is equivocal about the drug, the risk-to-benefit ratio is less favorable, or the patient is reluctant to comply with presumably efficacious treatment.

For patients with rare or unusual conditions, the use of the single-patient trial may not only benefit the patient but also add to knowledge about the management of unusual conditions. The literature contains numerous examples of single-patient experiments where treatments of conditions like familial Mediterranean fever and narcolepsy were evaluated with N-of-1 trials.

Doubt about efficacy may be a motivating factor for a single-patient trial also when a patient insists on a treatment as necessary or effective in contradiction to medical advice or practice. The single-patient trial can be used when the physician is unable to convince the patient otherwise. In this case, a negative clinical trial should not surprise the physician but may be convincing to the patient.

After determining whether therapeutic efficacy is in doubt and deciding whether one wishes to demonstrate efficacy or a lack thereof, the clinician will need to consider other questions that affect the feasibility and worth of a single-patient trial. First is whether a treatment will likely be long term. Given the time required to conduct such a trial, single-patient trials of short-term therapies tend not to be

worth the effort required of the patient, and they are less likely to have value for the individual patient unless the patient will require the short-term treatment repeatedly.

Several questions related to the pharmacokinetics of a possible therapeutic agent affect the logistics and ease of doing single-patient trials.[12] The ideal treatment for single-patient trials is one that can be rapidly started and stopped. Thus, outcomes can be assessed starting relatively early in the trial, and there is little or no carryover between treatment periods. When these criteria are not met, carryover or period effects may complicate the interpretation.[13] These effects may require trials that are much more time consuming (for example, involving washout periods) or involve special design modifications. In general, single-patient trials are less likely to be useful for curative treatments (so-called period effects) or for long-acting treatments (due to carryover effects).

How to Do a Clinical Trial

There are three critical components of the single-patient trial: randomization, blinding of patient and physician to treatment assignment, and defining and quantitating the outcomes. The last, establishing explicit criteria for evaluating the efficacy of treatment, is a feature of the single-patient trial that is also important for medical therapeutics in general.

Randomization is necessary to minimize systematic biases that will occur related to the order of treatment and to permit double blinding to occur. Randomization is usually accomplished in a crossover style, that is, in blocks of two. If, however, it is predetermined that four, six, or eight trials will be done, the statistical power of the trial is improved considerably by randomization in larger blocks.[13] For example, when six trials are planned, the possible P values range from .125 for the paired experiment in which three crossover pairs occur ($[1/2]^3$) to .03 when all six trials are randomized independently ($[1/2]^6$). Intermediate values are possible when constraints are added.

Blinding is a key element to minimize observer-induced bias. In most single-patient trials, the patient records symptoms and, in some cases, signs. Ideally both patient and physician are blind to the treatment assignment. Records of assignment are kept with one of the trial service staff and, if a drug is involved, the pharmacist who has prepared the treatment packages.

Single-patient trials require that the goals of treatment be explicitly identified at the time the patient enters the trial. Ideally, three to five key variables are determined. The variables may reflect disease activity or symptom severity. Usually the most important variables measure patient functioning, reflecting the value of treatment for the patient. In the ideal case, outcomes would include the measurement of a physical sign, a subjective or objective rating of performance in conjunction with, for example, a laboratory measurement reflecting disease activity. The patient's goals must be assayed to be certain that the measures of performance are compatible with the patient's wishes, especially regarding quality of life.

Systematic measurement of a limited number of variables is important for a successful single-patient trial. We typically use self-administered questionnaires that rely on 7-point Likert scales or tabulate the frequency of events. We also teach patients to measure biologic variables like the forced expiratory volume in one second, peak flow, and walk time. We have found it easier to use 7-point Likert scales than visual analog scales. In the standard crossover design, the patient can be asked to state a preference for one treatment period compared with the other.

There are other issues that must be solved when designing a clinical trial. A critical question is the duration of treatment. In general, we believe the old adage, "shortest is easiest." Treatment often takes longer than expected, however, because time is required for peak effects to develop or for treatment effects to dissipate. For drug regimens that are rapidly started and stopped, treatments can be shorter and a random block design of six or eight trials of active drug and placebo can be evaluated in less than two weeks.

A special case occurs when a drug is being used to minimize or prevent attacks or exacerbations of a recurrent disease. To determine duration, the frequency of exacerbation needs to be estimated. Given a reasonable estimate of the frequency, the duration can be based on the "rule of 3s." This rule states that if an event occurs once every x days, the duration of observation must be three times x days to be 95% certain to observe one event. In the case of familial Mediterranean fever where an attack may occur once every two weeks, the treatment period would need to last six weeks to be reasonably certain to observe an effect.

Another question that affects the duration of the trial is how many pairs or trials are needed. The answer to this is the tautology, "as many as are needed." In some trials, we have recommended that a single pair may provide an adequate demonstration of efficacy. Such a demonstration lacks statistical power, but the demonstration of effect may be so compelling as to convince both patient and physician that efficacy is no longer in doubt. On the other hand, when the probability of a treatment being effective is about 50% before the

TABLE 2.—*Posterior Probabilities as Function of Prior Probabilities and Likelihood Ratio*

Prior Belief Treatment Is Effective, P	Likelihood That Treatment Is Better Than Spontaneous	Patient Improves	Posterior Probability, P
.01	3	Yes	.030
	5	Yes	.051
	1/3	No	.003
	1/5	No	.002
.10	3	Yes	.25
	5	Yes	.55
	1/3	No	.032
	1/5	No	.022
.50	3	Yes	.75
	5	Yes	.83
	1/3	No	.25
	1/5	No	.17
.80	3	Yes	.92
	5	Yes	.95
	1/3	No	.57
	1/5	No	.44
.90	3	Yes	.96
	5	Yes	.98
	1/3	No	.75
	1/5	No	.64
.95	3	Yes	.98
	5	Yes	.99
	1/3	No	.86
	1/5	No	.79

trial, and there are major risks of side effects, anything short of a statistical certainty may not be satisfactory. In the case of a paired crossover trial, the binomial distribution suggests that after four trials, the probability of treatment being repeatedly favored over placebo is .5 after the first trial, .25 after the second trial, .125 after the third trial, and .0625 after the fourth trial, which is $(1/2)^4$.

In general, the issue of "statistical" certainty—the mythical $P < .05$—is less critical in single-patient trials. An interesting perspective is added by assaying the clinician's estimate of the likelihood of success in that patient (the prior probability) and determining the estimated likelihood that the treatment is efficacious based on the literature. Using a Bayesian analysis, a posterior probability based on the patient outcome in a single-patient trial can be calculated as shown in Table 2 (G. van Belle, written communication, June 1987). These posterior probabilities show the effect that a single-patient trial can have on a clinician's level of certainty that treatment will be helpful for a patient.

Conclusion

We formed the trial service to simultaneously establish, demonstrate, and determine the value of single-patient trials in clinical practice and to help do the clinical trials. Our involvement ranges from being limited consultants providing study drugs and simply reviewing the protocol, to providing detailed, in-depth consultation regarding the value of a clinical trial in a particular patient, developing a study design, interviewing the patient, developing target outcomes, printing forms, preparing placebo drug and outcome forms, and doing follow-up. In all cases, we provide an interpretation of the results of the trial and are anxious to learn how the trial was used in clinical decision making and practice.

In summary, single-patient clinical trials can be used to improve the efficacy of treatment—especially long-term treatments and treatments with uncertain efficacy or a risk of serious toxic effects. Examples of suitable conditions for study are numerous, including common problems such as chronic obstructive lung disease, osteoarthritis, recurrent headache and other chronic pain syndromes, "fibrositis" or fibromyalgia, and agitation in demented patients. We have done trials in these common conditions and have also investigated more unusual and complex problems such as progestational drug side effects, treatment of the "restless" leg syndrome, and treatments of orthostatic hypotension. The principal benefits are an increased certainty for patients and their physicians that a treatment is worth pursuing because it is effective or should be abandoned because of an absence of a net benefit.

REFERENCES

1. Feinstein AR: Clinical biostatistics—II. Statistics versus science in the design of experiments. Clin Pharmacol Ther 1970; 11:282-292
2. Guyatt G, Sackett D, Taylor DW, et al: Determining optimal therapy—Randomized trials in individual patients. N Engl J Med 1986; 314:889-892
3. Sackett DL: Clinical diagnosis and the clinical laboratory. Clin Invest Med 1978; 1:37-43
4. Brody H: The lie that heals: The ethics of giving placebos. Ann Intern Med 1982; 97:112-118
5. Sackett DL: Bias in analytic research. J Chronic Dis 1979; 32:51-63
6. Cochrane AL: Effectiveness and Efficiency: Random Reflections on Health Service. London, Nuffield Hospitals Trust, 1972
7. Freiman JA, Chalmers TC, Smith H, et al: The importance of beta, the type II error, and sample size in the design and interpretation of clinical trials. N Engl J Med 1978; 299:690-695
8. Edgington ES: Statistics and single case analysis. Prog Behav Modif 1984; 16:83-119
9. Adams R: Internal-mammary-artery ligation for coronary insufficiency. N Engl J Med 1958; 258:113-116
10. Berquist WE, Rachelefsky GS, Kadden M, et al: Effect of theophylline on gastroesophageal reflux in normal adults. J Allergy Clin Immunol 1981; 67:407-411
11. Cobb LA, Thomas GI, Dillard DH, et al: An evaluation of the internal-mammary-artery ligation by a double-blind technique. N Engl J Med 1959; 260:1115-1118
12. Porta MS: The search for more clinically meaningful research designs: Single-patient randomized clinical trials. J Gen Intern Med 1986; 1:418-419
13. Kazdin AE: Single-Case Research Designs: Methods for Clinical and Applied Settings. New York, Oxford Press, 1982

The *n*-of-1 Randomized Controlled Trial: Clinical Usefulness

Our Three-Year Experience

Gordon H. Guyatt, MD; Jana L. Keller, BSc; Roman Jaeschke, MD; David Rosenbloom, DPharm; Jonathan D. Adachi, MD; and Michael T. Newhouse, MD

Objective: To review the feasibility and effectiveness of *n*-of-1 randomized controlled trials (*n*-of-1 trials) in clinical practice.

Design: Individual trials were double-blind, randomized, multiple crossover trials. The impact of *n*-of-1 trials was determined by eliciting physicians' plans of management and confidence in those plans before and after each trial.

Setting: Referral service doing *n*-of-1 trials at the requests of community and academic physicians.

Object of Analysis: All trials were planned, started, and completed by the *n*-of-1 service.

Measures of Outcome: The proportion of planned *n*-of-1 trials that were completed and the proportion that provided a definite clinical or statistical answer. A definite clinical answer was achieved if an *n*-of-1 trial resulted in a high level of physician's confidence in the management plan. Specific criteria were developed for classifying an *n*-of-1 trial as providing a definite statistical answer.

Main Results: Seventy-three *n*-of-1 trials were planned in various clinical situations. Of 70 *n*-of-1 trials begun, 57 were completed. The reasons for not completing *n*-of-1 trials were patients' or physicians' noncompliance or patients' concurrent illness. Of 57 *n*-of-1 trials completed, 50 provided a definite clinical or statistical answer. In 15 trials (39% of trials in which appropriate data were available), the results prompted physicians to change their "prior to the trial" plan of management (in 11 trials, the physicians stopped the drug therapy that they had planned to continue indefinitely).

Conclusion: We interpret the results as supporting the feasibility and usefulness of *n*-of-1 trials in clinical practice.

Annals of Internal Medicine. 1990;112:293-299.

From McMaster University, Hamilton, Ontario. For current author addresses, see end of text.

Randomized controlled trials are usually required to establish valid evidence of drug efficacy (1-3). However, there remain a number of clinical situations in which treatment decisions cannot be based on such trials. For example, guidance is unavailable for treating conditions that have not been investigated with randomized controlled trials; some conditions are so rare that even multicenter collaborative trials are not feasible. Further, even when a relevant randomized controlled trial generates a definite answer, its result may not apply to an individual patient. First, if the patient does not meet the eligibility criteria, extrapolation may not be appropriate; second, regardless of the overall trial results, some patients appear to benefit from the experimental therapy and some do not (4). To maintain the methodologic safeguards provided by randomized controlled trials and avoid the disadvantages of large-sample multicenter studies, we have developed a corresponding methodology for examining the intervention effect in individual patients.

Experimental studies (5-7) of single subjects have long been part of psychologic research. The methodology is known as single case or single subject research, $n = 1$, or, *n*-of-1 randomized controlled trials (hereafter referred to as *n*-of-1 trials). We have previously described how *n*-of-1 trials may be used in medical practice to determine the optimum treatment of an individual patient (4). More recently, we have provided detailed guidelines (8) for clinicians interested in conducting their own *n*-of-1 trials. Results pertaining directly to the patient involved are available immediately after the patient has completed the trial.

In 1985, we designed an *n*-of-1 service to facilitate clinicians' involvement with *n*-of-1 studies in our community (9). We have a formal referral service for *n*-of-1 studies and a tutorial service that teaches clinicians how to run their own trials. We describe our 3-year experience with providing the *n*-of-1 service in our community. We examined a spectrum of conditions and interventions in which *n*-of-1 trials were done and studied the outcome of each trial. The questions we asked were as follows: Are *n*-of-1 trials able to provide clinically useful information? Do clinicians change their management plans as a result of *n*-of-1 trials? Does physicians' confidence in management decisions change as a result of *n*-of-1 trials?

Methods

Criteria for Doing an *n*-of-1 Trial

After a clinician and a patient expressed interest in conducting an *n*-of-1 trial, we assessed the suitability of the underly-

ing condition and potential therapeutic intervention. We have previously reported a set of criteria (8) that should be satisfied before an *n*-of-1 trial is attempted; these criteria were applied to patients' presentation to the *n*-of-1 service. In short, in addition to the effectiveness of treatment being in doubt, the disorder should be chronic and relatively stable. The treatment, if effective, should be continued long-term, and the patient should be eager to collaborate in designing and participating in the *n*-of-1 trial. In addition, the treatment or treatments must have a rapid onset and termination of action, and an optimal treatment duration should be known and practical. In each case, the choice of medication and the dosage were selected on the basis of the attending physician's clinical judgment.

Conduct of Individual *n*-of-1 Trials

If our initial assessment of the clinical situation indicated that an *n*-of-1 trial was indicated, we prepared an individualized trial package. To assess drug efficacy, we administered individualized questionnaires that examined the severity of symptoms that were identified by patients as part of their disease and important in their daily life. These questionnaires consisted of four to seven items (symptoms), and severity of symptoms was usually measured on a 7-point scale. For example, if shortness of breath while shopping was a symptom identified as part of the illness and important in daily life, the patient was asked: Please indicate how short of breath you have been while shopping during the previous 2 or 3 days, by choosing one of the options from the scale below:

1. Extremely short of breath
2. Very short of breath
3. Quite a bit short of breath
4. Moderately short of breath
5. Mildly short of breath
6. A little short of breath
7. Not at all short of breath

Either the referring physician or a physician-member of the *n*-of-1 service saw the patient after each treatment period. The trial design was based on pairs of active drug and placebo, high dose and low dose, or first drug and alternate drug combinations; the order of administration within each pair was determined by random allocation. We recommended that at least three pairs of treatments be completed. Medication was prepared by one of the participating pharmacies. If active medication and matching placebo were available from the manufacturer, they were used; if not, the medication was crushed and put in capsules, and matching placebo capsules were prepared. The pharmacy held the code, and all other members of the team were blind to allocation. Treatment targets were monitored on a regular, predetermined schedule throughout the trial. If a patient felt much worse at any time during the trial, the current treatment period was terminated and, without breaking the code, the next treatment period was begun. The trial continued as long as the clinician and patient agreed that they needed more information to get a definite answer about the efficacy of the treatment or until the patient or clinician decided for any other reason to end the trial.

At the study's conclusion, the results were reported to the patient's physician. Mean values for all measures for each treatment period, the mean differences between treatment and control periods, the 90% confidence interval (CI) around the differences, and the probability of differences seen being due to chance (using a one-sided paired *t*-test of the difference in score) were reported (8). We also examined each treatment's magnitude of effect. Our previous experience with the symptom questionnaires that used a 7-point scale suggested that an improvement of 0.5 points per question corresponds to a noticeable improvement in the patient's well-being (10). For instance, if there were six questions, a total change of 3 or more points was considered clinically important.

To assess the impact of the *n*-of-1 trial on the physician's management plan, we asked each physician how he or she would treat the patient without an *n*-of-1 trial and, when *n*-of-1 trial results became available, how he or she intended to treat the patient. Management plan options included continuing the drug therapy, withdrawing the drug, or "other." We also investigated the level of the physician's confidence in his or her management plan, both before and after the *n*-of-1 trial, again using a 7-point scale. The physicians were asked the following: How comfortable do you feel now about your treatment plan?

1. Totally comfortable, certain it's the right thing for the patient
2. Almost totally comfortable, very likely it's the right thing for the patient
3. Quite comfortable, likely that the treatment plan is best for the patient
4. Not totally comfortable, but treatment plan is very likely to be as good as alternatives
5. Mildly uncomfortable, some uncertainty whether treatment plan is best for the patient
6. Moderately uncomfortable, feeling that the treatment plan may not be the best for the patient
7. Extremely uncomfortable, uncertain about treatment plan and, if wrong, patient may suffer

Review of 73 *n*-of-1 Trials

Between October and December of 1988, we reviewed the files of all *n*-of-1 trials done in cooperation with our *n*-of-1 service. Trials were classified as complete when three pairs of treatment periods were completed or the trial was interrupted before completing three treatment pairs because of the clinician's and patient's belief that drug effectiveness had been established or refuted. The reasons for interruption were occurrence of intolerable symptoms compatible with side effects, perceived large treatment effect of the active medication, and such a low frequency of symptoms that the medication was judged not to be needed.

Trials not in either of these categories were classified as incomplete (interrupted before completing three pairs with no clinical conclusion reached before trial termination). Among completed trials, we examined the proportion that provided a definite clinical answer. These included trials that resulted in a high level of clinicians' confidence in their management decisions after an *n*-of-1 trial (1 or 2 on a 7-point scale); and trials that were interrupted before completing three treatment pairs because of the clinician's and patient's belief that drug effectiveness had been established or refuted. To classify such trials as providing definite answers, the clinical impression of drug efficacy (or its side effect) had to be confirmed after breaking the code.

For trials in which the primary outcome measure was the symptom questionnaire that used a 7-point scale, we have developed a set of statistical criteria to classify individual *n*-of-1 trials. Categories include providing a definite answer (either confirming drug or placebo superiority or indicating no difference), showing a trend in favor of active drug or placebo, or leaving the question of intervention efficacy unanswered (indefinite). These criteria use a combination of the clinical importance cut-off (0.5 points per question mean difference [D] in symptoms score) and statistical evaluation of the difference observed (one-tailed $P \leq 0.05$, narrow CI around the difference between active drug and placebo). The complete set of criteria is presented in Appendix 1.

Examples of *n*-of-1 Trials

To show what is involved in doing an *n*-of-1 trial, we will describe a case in detail. A 23-year-old woman

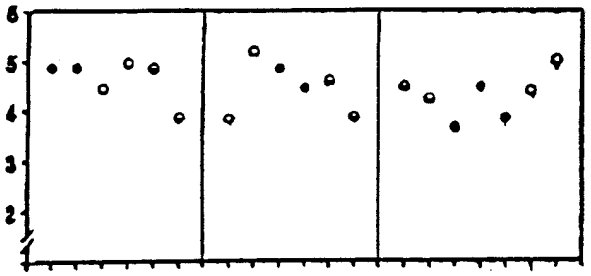

Figure 1. Results of *n*-of-1 trial, propranolol therapy for vasovagal syncope. Half-open circles represent weekly mean scores while receiving propranolol, 40 mg four times daily; open circles represent weekly mean scores while receiving propranolol, 20 mg four times daily; and closed circles represent weekly mean scores while receiving placebo.

presented in the autumn of 1987 with a history of recurrent vasovagal syncope of a year's duration. Associated symptoms included presyncope, nausea and vomiting, migrainous headaches, and flushing episodes. There was no obvious trigger to these symptoms. The syncopal episodes occurred as frequently as twice a week, the other symptoms on a more frequent basis, and the constellation of symptoms was adversely affecting the patient's quality of life. Extensive investigation showed no hormonal or autonomic nervous system abnormality. The patient was given nifedipine (for headaches) and amitriptyline as a vagolytic agent and her condition was initially judged to have improved somewhat; however, symptoms remained a major problem.

It has been hypothesized that a vasodepressor reaction (or common faint) can follow sympathetic nervous system stimulation, resulting in decreased left ventricular volume and stimulation of intracardiac receptors (11). This mechanism was thought to be playing a role in this patient's problems. A "tilt-table isoproterenol" test was abnormal; the patient developed significant bradycardia and hypotension when tilted to 60 deg and infused with 8 μg of isoproterenol (11). The patient's physician thought that propranolol might benefit (11) and contacted our *n*-of-1 service to conduct a trial.

The physician was uncertain of the optimal dosage, so the trial was set up with triplets of treatment periods instead of pairs. Each period lasted 2 weeks and, in each triplet, the patient received either placebo, 20 mg of propranolol four times daily, or 40 mg of propranolol four times daily. Treatment targets included daily rating of symptoms of lightheadedness and syncope, headaches, nausea or vomiting, feeling warm or sweating, and fatigue. Each symptom was rated on a 7-point scale. For instance, the patient was asked the following: How much trouble or distress as a result of lightheadedness or loss of consciousness have you had during the last day?

1. A very great deal of trouble or distress
2. A great deal of trouble or distress
3. A good deal of trouble or distress
4. A moderate amount of trouble or distress
5. Some trouble or distress
6. Very little trouble or distress
7. No trouble or distress

The results of the three triplets of treatment periods are summarized in Figure 1. Each data point in Figure 1 represents the mean of seven ratings of the five symptoms over a period of 1 week. The patient felt that there were no significant differences in how she felt over the 19 weeks of the trial, and this was confirmed by the symptom scores. It was concluded that propranolol was not effective.

Now uncertain about the benefit of amitriptyline in relieving symptoms, the attending physician wished to conduct a second trial before restarting the therapy. This trial was to have 4-week treatment periods, with the patient receiving placebo or 100 mg of amitriptyline at bedtime during each period. The same five symptoms were monitored, again on a daily basis. Before starting the trial, the physician replied to our questionnaire, stating that his a priori estimate of effectiveness was that the amitriptyline was of no benefit and that he was very confident of this assessment.

The patient felt much worse during the second period of the first pair than she had during the first period and, after 2 weeks of the second period, was convinced that she was receiving placebo. Without breaking the code, the period was terminated and the next pair begun. During the second period of the second pair, the patient again felt much worse and the period was terminated after the first week. After 1 week of the third pair, the patient again became convinced that she was receiving placebo and the second period of the third pair was begun early. The results are presented in Figure 2. The patient had been correct in each case about when she received placebo, and the large differences in symptom score reflect the magnitude of the differences she experienced between taking active drug and taking placebo. The mean differences in symptom score per question between active drug and placebo periods for the three pairs were 1.88, 1.81, and 2.08. A paired *t*-test with two degrees of freedom suggests that these results are very unlikely to have occurred by chance ($P < 0.001$). It was concluded that amitriptyline was effective, and the drug treatment has been continued to the present.

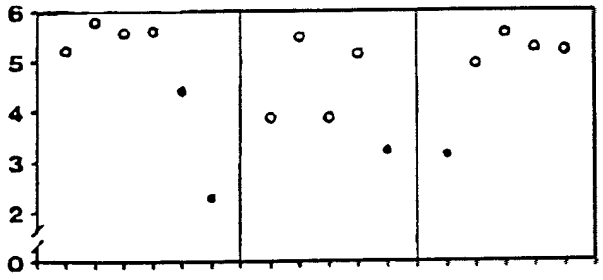

Figure 2. Results of *n*-of-1 trial, amitriptyline therapy for vasovagal syncope. Open circles represent weekly mean scores while receiving amitriptyline, and closed circles indicate weekly mean scores while receiving placebo.

Table 1. *Outcome of 73 n-of-1 Randomized Controlled Trials*

Planned *n*-of-1 trials, $n = 73$
Three *n*-of-1 trials never started (1 because of death; 1, concurrent illness; and 1, consent withdrawn)
n-of-1 trials begun, $n = 70$
Thirteen *n*-of-1 trials not completed (7 because of patients' noncompliance; 5, concurrent illness; and 1, physician noncompliance)
Completed *n*-of-1 trials, $n = 57$
Nine *n*-of-1 trials with 3 pairs completed did not provide a definite clinical answer; 2 of the 9 provided a definite statistical answer
Definite *n*-of-1 trials, $n = 50$
Forty-eight trials were clinically definite; 19, statistically definite

Results

Spectrum of Use

We have not kept systematic track of inquiries about *n*-of-1 trials that were planned but deemed infeasible after preliminary discussion. Some examples include trials with patients with inflammatory bowel disease (in whom exacerbations occur too infrequently to make a trial feasible) and major changes in prednisone in patients with obstructive airway disease (in whom functional adrenal deficiency is likely to have developed). On several occasions, an open trial resulted in obvious benefit or obvious side effects before a formal trial was begun. In several instances, we were approached about patients with many unstable medical problems that made reliable ascertainment of the effect of a single medication impossible. Finally, *n*-of-1 trials were sometimes infeasible because of reservations about the patient's ability to keep a valid symptom diary.

Overall, our service participated directly in preparing 73 *n*-of-1 trials. Some results from 5 of these trials have been reported elsewhere (4, 8, 9, 12). Most of the trials tested a specific form of therapy in patients whose underlying condition was clearly defined (for example, amitriptyline therapy for fibrositis, ipratropium or theophylline for chronic airflow limitation). In three instances, the trial was used as a diagnostic tool: In a patient with inconclusive laboratory test results, the clinician investigated the efficacy of hydrocortisone in relieving symptoms possibly caused by Addison disease; in two trials, the clinician tested the efficacy of pyridostigmine bromide in ameliorating symptoms possibly caused by myasthenia gravis. In two other cases, different dose regimens of the same medication were used to determine the balance between the drug's efficacy and its side effects (prednisone therapy for chronic airflow limitation and propranolol for syncope).

The results of the 73 *n*-of-1 trials are presented in Table 1. Three trials were planned, but never started (1 because of concurrent illness; 1, consent withdrawn; and 1, patient's death). Of the 70 *n*-of-1 trials that began, 57 were completed. The reasons for suspension of 13 trials were patients' concurrent illness (5 trials) and lack of patients' (7 trials) or physicians' (1 trial) compliance with the study protocol. Among the 57 completed *n*-of-1 trials, the number of pairs were as follows: eight pairs, 1 trial; six pairs, 1; five pairs, 2; four pairs, 9; three pairs, 31; two pairs, 11; and one pair, 2. The duration of treatment periods varied widely, from 1.5 days to 6 weeks. The majority of trials lasted 1 to 4 weeks. Appendix 2 presents the spectrum of clinical conditions in which *n*-of-1 trials were done. One physician was involved in 19 trials; another, in 8. An additional four physicians participated in more than 1 completed trial.

Results of Completed Trials

Forty-eight of 57 completed *n*-of-1 trials (84% of all completed and 66% of all planned) provided a definite clinical answer. These 48 trials included 39 that resulted in a high level of clinicians' confidence in the appropriateness of their management decisions after three pairs of treatment had been completed. An additional 9 *n*-of-1 trials were classified as complete despite trial interruption before completing three pairs. In 4 trials, differences between two treatment periods were so dramatic, the physician and patient decided to end the trial (ipratropium therapy for chronic airflow limitation on three occasions and haloperidol for psychosis on one). In each of these 4 trials, the clinical impression was confirmed after breaking the code; the clinician had guessed correctly when the patient was receiving active drug. On two additional occasions, occurrence of clinically important deleterious effects led to the termination of *n*-of-1 trials (theophylline therapy for chronic airflow limitation and clonidine for rheumatoid arthritis). Again, the clinical decision was substantiated after the code was broken. During 3 trials, the symptoms chosen as treatment targets did not occur within the first few treatment periods and the trial was terminated (propranolol therapy for syncope, dilantin for Meniere disease, and propantheline for abdominal pain). In each of the 9 *n*-of-1 trials classified as complete despite less than three pairs being done, active drug was compared with placebo.

Results of complete trials that used symptom questionnaires with responses on a 7-point scale as a primary outcome measure were reviewed according to criteria presented in Appendix 2. We had the data necessary to do this analysis in 44 *n*-of-1 trials. In 19 of 44 cases, the trial provided a definite statistical answer. In 15 trials, the beneficial role of the drug was confirmed; in 4, there was no difference between investigated therapy and placebo. None of the trials analyzed using these criteria indicated a harmful effect of a drug. All but 2 *n*-of-1 trials providing a definite statistical answer were classified as definite according to clinical criteria. In 1 of these 2 *n*-of-1 trials, the physician tested the efficacy of amitriptyline therapy for fibrositis—the impression of drug efficacy obtained during an earlier open trial was so strong that the results of the initial *n*-of-1 trial excluding drug benefit were questioned. A subsequent *n*-of-1 trial, with the same patient using a higher dosage of amitriptyline, confirmed the results of the first trial, and the physician discontinued the medication. In the second case, the physi-

cian questioned a patient's claim that pyridostigmine provided an improvement in weakness that was possibly related to myasthenia gravis. Despite a clearly positive *n*-of-1 result, failure by a neurologist to confirm the diagnosis of myasthenia led the attending physician to speculate that the patient might somehow have broken the blind, thus invalidating the results. The total number of *n*-of-1 trials providing definite clinical or statistical answer was, therefore, 50. Five *n*-of-1 trials had trends suggesting drug benefit, and, in two cases, trends favored placebo. Results of 18 completed trials were classified according to the statistical criteria as indefinite.

Management Plans and Clinicians' Confidence

In 38 trials, the data on management decisions were available both before and after the trial. In 23 cases, the original decision was unchanged after the trial result became available. In the remaining 15 trials (39%), results of the *n*-of-1 trial prompted physicians to change the original decision (in 11 cases, to stop the drug treatment completely rather than continue; in 3 cases, to continue drug therapy indefinitely rather than stop; and, in 1 case, to conduct an additional *n*-of-1 trial). The level of confidence in the new management decision, measured on a 7-point scale, was 1.82 ± 1.05 (mean ± SD). Confidence in the original decision was 4.62 ± 1.36. This change in management confidence was similar to the increase seen in the *n*-of-1 trials that supported the original decision (from 4.53 ± 1.62 to 1.82 ± 1.07). The complete spectrum of changes in physicians' confidence after the 38 *n*-of-1 trials for which data are available for both before and after the trial is depicted in Figure 3. In most cases, physicians clearly were far more confident in their management after the *n*-of-1 trial.

In 44 *n*-of-1 trials, three pairs of treatment were completed. In 39 of these trials, physicians expressed total or very high confidence in their management decision (1 or 2 on a 7-point scale). In no case was this degree of confidence present before the *n*-of-1 trial. After these 44 *n*-of-1 trials, the average score on a 7-point management confidence scale was 1.77 ± 0.99.

In most of the trials we report, the attending clinicians had already conducted their own open trials and remained uncertain about treatment efficacy. In these instances, they would have managed the patients as described in the questionnaires we administered. In a few trials, physicians preferred to have the first exposure of patients to the experimental treatment as part of an *n*-of-1 trial. Although physicians may have considered options such as continuing the medication for a period and then testing response to withdrawal or conducting open trials of withdrawal and reinstitution, such plans were made explicit on only a few occasions.

Discussion

We present our initial, 3-year experience in conducting *n*-of-1 trials and offering the *n*-of-1 service to community physicians. We tested this method of solving difficult therapeutic dilemmas in a broad spectrum of conditions and using different interventions. The clinical problem was most commonly clarification of the efficacy of a medication, generally recognized as useful, in an individual patient. In some cases, trials were used for the clarification of an optimal dosage of a medication or as an aid to diagnosis.

We were able to complete 81% of trials that were begun. The commonest reasons for not completing a trial were patients' noncompliance with the study protocol or emergence of a concurrent illness. In each trial, we tried to complete three pairs of treatments; achieving this goal was the commonest reason to categorize a trial as complete. Some trials were also categorized as complete despite the fact that three pairs of treatments had not been achieved. In all of these *n*-of-1 trials, the clinically relevant answer was reached at an earlier point. On three occasions, target end points occurred with an unexpectedly low frequency regardless of the treatment used. These *n*-of-1 trials were interrupted and classified not only as complete but also as providing a definite clinical answer: Indication for the use of a drug was refuted. These three *n*-of-1 trials dramatically show the necessity of assessing drug efficacy in a blind manner. Had the drug been tested in an open trial, the results would have been interpreted as showing the striking efficacy of the intervention.

To judge the clinical usefulness of *n*-of-1 trials, we developed a set of both clinical and statistical criteria. We felt that because the goal of an *n*-of-1 trial is to clarify a management decision, an *n*-of-1 trial can be considered definite only if this goal is achieved. A definite answer was obtained in 71% of all attempted *n*-of-1 trials. Clinicians were more liberal in their conclusions that a definite answer had been reached. When using rigorous statistical criteria for a definite answer, such an answer was attained in only 27% of trials that were begun (43% of the trials in which data required to make this assessment were present). On two occasions, physicians did not believe the statistical results; in both cases, two separate *n*-of-1 trials yielded the same results.

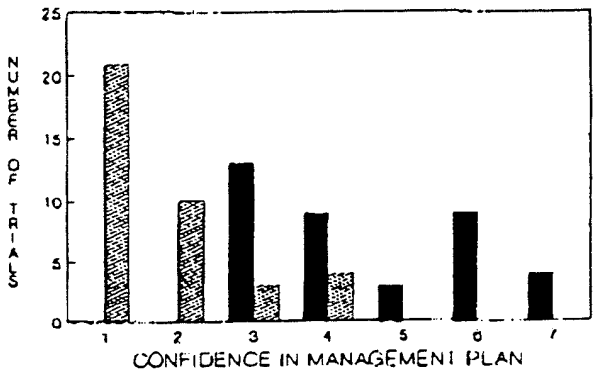

Figure 3. Impact of *n*-of-1 trials on clinicians' confidence in their management plans, data from 38 *n*-of-1 trials. The *y*-axis represents number of trials, and the *x*-axis indicates confidence in management plan. Closed bars represent confidence in management plan before trial, and open bars represent confidence in management plan after trial.

The relatively small proportion of trials in which statistical criteria for a definite result were obtained reflects to some extent the limited power of statistical tests when only three pairs have been conducted. The extent to which the clinicians were convinced of the results when statistical criteria were not met attests to the value of the method even without statistical analysis. Another limitation of statistical analysis is that the decision to continue with additional pairs can be driven by the results, potentially invalidating the nominal *P*-value obtained. Because of these limitations, we view the statistical analysis as an adjunct (but often very useful adjunct) for the interpretation of the results of *n*-of-1 trials.

The expense incurred by conducting *n*-of-1 trials will be an issue. Until now, our trials have been paid for by research funds. We have not, therefore, established a standard fee for the referral nor decided on how fees should be modified depending on the nature and length of the study. Although, in our experience, the research assistant time per trial was considerable, much of this time was spent on activities (such as administering questionnaires to physicians) that would not be part of *n*-of-1 trials once they are established in clinical practice. We believe that even without detailed information on costs, conducting *n*-of-1 trials is likely to be cost-effective. In our experience, a substantial proportion of trials result in discontinuation of medication that would otherwise have been continued for months or years. The cost savings from discontinuing medication and from reducing physician time spent in medication review and in treating adverse reactions to medication is likely to be considerable. Third-party payers may wish to consider these potential savings when developing policies on reimbursement of costs associated with *n*-of-1 trials.

We believe that our results show that *n*-of-1 trials are feasible to conduct in clinical practice and often result in clinically important changes in clinicians' confidence in their management decisions and in the management decisions themselves. We believe that most physicians try to be scientific in their approach to medication prescription and use some of the principles of the *n*-of-1 trial (such as observation of patients on and off medication) in their day-to-day practice. The methodology of the *n*-of-1 trial provides physicians with a set of tools that can further increase the scientific rigor of their clinical practice and increase the likelihood that the treatments they prescribe are indeed those that are best for the patient.

Acknowledgments: The authors thank Drs. Christopher Allen, Jennifer Blake, Dody Bienenstock, Ramona Carbotte, Clive Davis, Judah Denburg, Susan Denburg, Brian Hutchison, Jan Irvine, David Martin, Christopher Patterson, Michel Rothbone, Peter Rosenbaum, William Walsh, Robin Whyte, and the many other physicians who helped in the planning and conduct of individual trials; Professor Robin Roberts and Will Boyce and Drs. Dave Sackett, Murray Enkin, Stewart Pugsley, and John Chong, who contributed to the conceptual development of the *n*-of-1 approach; the pharmacy staff at McMaster University Medical Centre, particularly Iris Laschuck and Kathy Sussens, and at St. Joseph's Hospital, Hamilton, Ontario, particularly Doris Thompson, Betty Wong, and Nancy Giovinazzo for their help with preparation and management of medications.

Requests for Reprints: Gordon Guyatt, MD, Room 2C12, McMaster University Health Sciences Centre, 1200 Main Street West, Hamilton, Ontario, L8N 3Z5 Canada.

Current Author Addresses: Dr. Guyatt: Room 2C12, McMaster University Health Sciences Centre, 1200 Main Street West, Hamilton, Ontario, L8N 3Z5 Canada.
Ms. Keller: Room 2C3, McMaster University Health Sciences Centre, 1200 Main Street West, Hamilton, Ontario, L8N 3Z5 Canada.
Dr. Jaeschke: Fontbonne Building, St. Joseph's Hospital, 50 Charlton Avenue East, Hamilton, Ontario, L8N 4A6 Canada.
Dr. Rosenbloom: Pharmaceutical Services, McMaster University Health Sciences Centre, 1200 Main Street West, Hamilton, Ontario, L8N 3Z5 Canada.
Dr. Adachi: 25 Charlton Avenue East, Hamilton, Ontario, L8N 1Y2 Canada.
Dr. Newhouse: Firestone Chest/Allergy Services, St. Joseph's Hospital, 50 Charlton Avenue East, Hamilton, Ontario, L8N 1Y2 Canada.

Appendix 1. Criteria for Assessing the Results of an *n*-of-1 Randomized Controlled Trial

Statistical criteria

Definite answer	
Beneficial	$P \leq 0.05$ and $D \geq 0.5$
Harmful	$P \leq 0.05$ and $D \leq -0.5$
Neutral	$P > 0.05$ and $0.25 > D > -0.25$ and $\|CI\|$ not ≥ 0.5 or $P > 0.05$ and $0.25 > D > -0.25$ and $\|D\|$ for each pair ≤ 0.5
No definite answer but trend seen	
Beneficial trend	$0.3 \leq D < 0.5$ and $P \leq 0.05$ and CI includes 0.5 or $D \geq 0.5$ and $P > 0.05$
Harmful trend	$-0.3 > D > -0.5$ and $P < 0.05$ and CI includes -0.5 or $D \leq 0.5$ and $P > 0.05$
No definite answer	
Not meeting criteria for either of the above categories.	

Clinical criteria for definite trial

1. The clinician's high level of confidence in the appropriateness of the management decision after the *n*-of-1 trial (1 or 2 on a 7-point scale).
2. *n*-of-1 trial interruption before completing three treatment pairs because of the clinician's belief that drug effectiveness had been established or refuted (perceived large treatment effect or severe side effects, both confirmed after breaking the code, or low frequency of treatment endpoints).

Appendix 2. Spectrum of Clinical Conditions in Which *n*-of-1 Randomized Controlled Trials Were Used

Fifty-seven trials were completed. Twenty trials were done with 19 patients with fibrositis. In 18 of these trials, amitriptyline was tested; nitrazepam was tested in 2 trials. Sixteen trials were completed in patients with chronic airflow limitation. In 10 trials, inhaled ipratropium was tested; in 4, oral theophylline; and, in 2, inhaled salbutamol. Two other patients participated in 2 trials each. In a patient with suspected myasthenia

gravis, pyridostigmine was tested in 2 different trials. A patient with recurrent syncope participated in 1 trial testing propranolol, and 1 trial testing amitriptyline. Single trials were done in the following conditions, with the associated medication: chronic pain, maprotiline; anxiety, lorazepam; insomnia, lorazepam; suspected Addison disease, hydrocortisone; cryptosporidiosis, spiramycin; Raynaud disease, ketanserine; syncope, propranolol; coronary disease, diltiazem; familial, Mediterranean fever, colchicine; rheumatoid arthritis, clonidine; myositis, prednisone; abdominal pain, propantheline; Meniere disease, phenytoin; psychosis, haloperidol; and suspected polymyalgia rheumatica, prednisone.

Thirteen trials were begun but not completed. Eight of these trials involved patients with chronic airflow limitation. Five tested inhaled ipratropium; two, inhaled salbutamol; and one, oral theophylline. Single trials were started but not completed in the following conditions, with the associated medication: premenstrual syndrome, pyridoxine; spasticity in a paraplegic, clonidine; irritable bowel syndrome, trimebutine; idiopathic edema, captopril; and temporal lobe epilepsy, carbamazepine.

References

1. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest.* 1986;89:2S-3S.
2. How to read clinical journals: V: To distinguish useful from useless or even harmful therapy. *Can Med Assoc J.* 1981;124:1156-62.
3. Deciding on the best therapy. In: Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology. A Basic Science for Clinical Medicine.* Boston: Little, Brown and Company; 1985: 171-97.
4. Guyatt GH, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy-randomized trials in individual patients. *N Engl J Med.* 1986;314:889-92.
5. Kratochwill TR, ed. *Single Subject Research: Strategies for Evaluating Change.* New York: Academic Press; 1978: 316.
6. Hersen M, Barlow DH. *Single Case Experimental Designs: Strategies for Studying Behaviour Change.* 2d ed. New York: Pergamon Press; 1984: 419.
7. Kazdin AE. *Single-Case Research Designs: Methods for Clinical and Applied Settings.* New York: Oxford University Press; 1982: 368.
8. Guyatt GH, Sackett DL, Adachi JD, et al. A clinician's guide for conducting randomized trials in individual patients. *Can Med Assoc J.* 1988;139:497-503.
9. Keller JL, Guyatt GH, Roberts RS, Adachi JD, Rosenbloom D. An N of 1 service: applying the scientific method in clinical practice. *Scand J Gastroenterol.* 1988; 23(Suppl 147):22-9.
10. Jaeschke R, Singer J, Guyatt G. Measurement of health status: ascertaining the minimal clinically important difference. *Controlled Clinical Trials.* (In press).
11. Waxman MB, Yao L, Cameron DA, Wald RW, Roseman J. Isoproterenol induction of vasodepressor-type reaction in vasodepressor-prone persons. *Am J Cardiol.* 1989;63:58-65.
12. Woolf GM, Townsend M, Guyatt GH. Treatment of cryptosporidiosis with spiramycin in AIDS. An "N of 1" trial. *J Clin Gastroenterol.* 1987;9:632-4.