nation diet is indeed standard of care.[1] Compliance with the 6-food elimination diet in real-world settings is understandably low. Some recommend repeat EGDs at each step of the elimination diet, for up to 5 EGDs per patient,[2] at an immense burden to patients and expense to the health care system.

My review also revealed that the most common food trigger was dairy.[3] This seemed a perfect situation for an unblinded n-of-1 trial. The condition is chronic. Response to food elimination varies by individual. The symptoms are experienced near daily. And if I was going to eliminate a food from my diet for the rest of my life, I wanted to know for certain that it was the cause of my symptoms.

I eliminated all milk products from my diet for 8 weeks. My symptoms resolved. I then ate bread pudding with ice cream, sweet butter on my corn, and milk in my cereal. My symptoms recurred. I am now off dairy, symptom free, and convinced I know my food trigger. No further EGDs.

At the end of my trial, I read the article by Kravitz and colleagues[4] that demonstrated a lack of effectiveness of patients randomized to n-of-1 trials compared with usual care for treatment of chronic musculoskeletal pain. This was a laudable negative study that advanced the science of n-of-1 trials. Still, my experience suggests there may yet be a role for n-of-1 trials. My hope is that the study by Kravitz and colleagues spurs other researchers and clinicians not to abandon n-of-1 trials, but rather animates us to think creatively about scenarios in which n-of-1 trials might simplify treatment regimens, improve patient compliance, and reduce health care costs.

Alexander K. Smith, MD, MS, MPH

Author Affiliations: Division of Geriatrics, Department of Medicine, University of California, San Francisco, San Francisco (Smith); San Francisco VA Medical Center, San Francisco, California (Smith).

Corresponding Author: Alexander K. Smith, MD, MS, MPH, San Francisco VA Medical Center, 4150 Clement St, Ste 181G, San Francisco, CA 94121 (Alexander.Smith@ucsf.edu).

Conflict of Interest Disclosures: None reported.

1. Liacouras CA, Furuta GT, Hirano I, et al. Eosinophilic esophagitis: updated consensus recommendations for children and adults. *J Allergy Clin Immunol*. 2011;128(1):3-20.e6. doi:10.1016/j.jaci.2011.02.040

2. Health facts for you: the six food elimination diet for eosinophilic esophagitis. University of Wisconsin Health. https://www.uwhealth.org/healthfacts/nutrition/553.pdf. Published 2017. Accessed September 18, 2018.

3. Teoh T, Mill C, Chan E, Zimmer P, Avinashi V. Liberalized versus strict cow's milk elimination for the treatment of children with eosinophilic esophagitis [published online July 4, 2018]. *Can J Gastroenterol*. doi:10.1093/jcag/gwy030

4. Kravitz RL, Schmid CH, Marois M, et al. Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern Med*. 2018;178(10):1368-1377. doi:10.1001/jamainternmed.2018.3981

In Reply The letters from Vohra and Punja and Smith about our recent Original Investigation[1] underscore the same crucial point: patients may vary not only in their response to treatment, but also in their response to n-of-1 trials. Vohra and Punja offer an ardent counterpoint to Mirza and Guyatt's conclusion that n-of-1 trials are a "beautiful idea being vanquished by cruel and ugly evidence."[2(1379)] Smith's story offers an inspiring example of how n-of-1 trials may be applied informally in the service of better, more patient-centered care.

More generally, we believe that rumors of the demise of n-of-1 trials may be premature for 3 reasons. First, advances in mobile technology and ubiquitous home and environmental sensors will increasingly make tracking and self-experimentation much less demanding of time and effort on the part of both patients and clinicians. Second, n-of-1 trials retain promise not only for evaluating treatment benefits in individual patients, but also (as Vohra and Punja suggest) for comparing treatments of comparable efficacy that may differ in terms of costs or harms. Finally, we believe that n-of-1 trials have not yet been fully appreciated as an instrument for advancing scientific literacy and self-efficacy. Broad deployment of n-of-1 trials evaluating health behaviors such as diet, exercise, medication adherence, and stress reduction may not only show people how to be healthier, they might also teach them about the power of randomization, the importance of systematic outcomes assessment, and the need to minimize bias—concepts that are not only central to science and data literacy, but ultimately fundamental to democracy.

Richard L. Kravitz, MD, MSPH
Ida Sim, PhD, MD
Naihua Duan, PhD

Author Affiliations: Department of Internal Medicine, University of California, Davis, Sacramento (Kravitz); Department of General Internal Medicine, University of California, San Francisco, San Francisco (Sim); Department of Psychiatry, Columbia University, New York, New York (Duan).

Corresponding Author: Richard L. Kravitz, MD, MSPH, University of California, Davis, Department of Internal Medicine, Division of General Medicine, 4150 V St, Ste 2400, Sacramento, CA 95817 (rlkravitz@ucdavis.edu).

Conflict of Interest Disclosures: None reported.

1. Kravitz RL, Schmid CH, Marois M, et al. Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern Med*. 2018;178(10):1368-1377. doi:10.1001/jamainternmed.2018.3981

2. Mirza RD, Guyatt GH. A randomized clinical trial of n-of-1 trials: tribulations of a trial. *JAMA Intern Med*. 2018;178(10):1378-1379. doi:10.1001/jamainternmed.2018.3979

Finding Benefit in n-of-1 Trials

To the Editor The recent article by Kravitz and colleagues[1] reporting the results of a randomized clinical trial comparing n-of-1 trials with standard care for treatment of chronic musculoskeletal pain raises a number of interesting issues that merit further examination.

Weighed on a continuum of probability, it is certainly possible that there is a positive effect of these trials on pain interference score, which is unlikely to be because of chance alone based on the observed results; however, this effect is not demonstrated to the arbitrary level that is customary for statistical significance. This speaks directly to 2 points: (1) the *P* value standards we set are subjective and should be viewed not as dichotomous yes/no evidence, but as a continuum of probability as put forth in the American Statistical Association statement on *P* values,[2] and (2) the clinical significance of the findings of this trial are ignored in favor of identifying statistical superiority for pain-interference score. The stated goal of the authors was to establish the "benefits of participating in an n-of-1 trial, not to assess the superiority or inferiority of any particular treatment."[1(1369)] The improvement in shared decision making regarding medications is perhaps even more meaningful than a narrow statistical focus on pain interference score as an outcome owing to the mistrust of health care professionals that has become pervasive in the medical system and litigation that is a prominent feature of US medicine.

Because results were in favor of the n-of-1 arm for the primary outcome and the statistical analysis was close to our standard levels of significance ($P = .09$), the shortfall in recruitment and missing data may have important consequences in passing the accepted threshold. Owing to the inordinate focus on *P* value, it's unclear if a 1-sided statistical test would have provided the standard (arbitrary) level of statistical significance and completely reversed the overall interpretation of the results. Another problem is the usage of a 2-sample *t* test for ordinal pain scores, which by their discrete nature are not normally distributed, violating basic assumptions of this hypothesis test. A permutation or Wilcoxon signed rank test may have produced significant results at the .05 level for a difference in pain score reduction.

The study by Kravitz and colleagues[1] reported higher incidences of a 5-point pain score reduction, which may have resulted in statistical significance between groups using this as a primary outcome. The study also reported increased discussion from the patients and clinicians for the intervention group, with a *P* value of .01 for differences discussion scores at 6 months and a *P* value of .05 for 12-month discussion scores. These results actually suggest that n-of-1 trials show some promise for pain management, which warrants further investigation. Therefore, it is quite premature to conclude so thoroughly, as is done in the accompanying Invited Commentary, that "the results fail to show any benefit of n-of-1 care."[3(1378)]

Andrew Genius Chapple, PhD
James Walker Blackston, MSPH

Author Affiliations: Department of Biostatistics, School of Public Health, Louisiana State University, New Orleans (Chapple); Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana (Blackston).

Corresponding Author: James Walker Blackston, MSPH, Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, 3400 S Carrollton Ave, PO Box 13014, New Orleans, LA 70185 (jblackston@tulane.edu).

Conflict of Interest Disclosures: None reported.

1. Kravitz RL, Schmid CH, Marois M, et al. Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern Med*. 2018;178(10):1368-1377. doi:10.1001/jamainternmed.2018.3981

2. Wasserstein RL, Lazar NA. The ASA's statement on *P* values: context, process, and purpose. *Am Stat*. 2016;70(2):129-133. doi:10.1080/00031305.2016.1154108

3. Mirza RD, Guyatt GH. A randomized clinical trial of n-of-1 trials: tribulations of a trial. *JAMA Intern Med*. 2018;178(10):1378-1379. doi:10.1001/jamainternmed.2018.3979

To the Editor We read with interest the recent article by Kravitz and colleagues[1] describing a randomized clinical trial comparing n-of-1 trials with standard care for treatment of chronic musculoskeletal pain.

The goal of the study was to establish the "benefits of participating in an n-of-1 trial, not to assess the superiority or inferiority of any particular treatment."[1(1369)] However, there appears to be a disconnect between the study goal and the choice of outcomes, which were focused on pain interference scores across different treatment regimens. Therefore, the null results should be interpreted with respect to treatment efficacy, not design. The n-of-1 participants who demonstrated a better response to 1 of 2 treatments were likely to experience improved pain outcomes as a result of continuing to receive the superior treatment. However, there was a high proportion (>75%) of n-of-1 participants who had no treatment superiority, and this may explain the trial's findings.

Methodological factors, including the number and length of phases, could have influenced the degree of certainty about treatment superiority. Individual n-of-1 trials were heterogeneous, ranging from 4 to 12 weeks and with phases lasting 1 to 2 weeks. Some n-of-1 trials could have involved few crossovers, which might have affected the identification of a superior treatment and increased the risk of type-2 error. Furthermore, there was a lack of blinding, and the study did not achieve its sample size target, which may have influenced the results.

A total of 48% of n-of-1 participants incorporated nonpharmacological treatment into their n-of-1 trials. Designing phases that are long enough to show an effect (if one exists)[2] is a relatively easy task for n-of-1 pharmacological trials owing to well-known drug half-lives, but this is more difficult in nonpharmacological trials (eg, exercise, acupuncture) where one must hypothesize about the immediacy and duration of effect.[3] As such, 1 to 2 weeks is potentially too short.

Finally, the findings assume that participants adhered to the superior or recommended treatments. Lack of adherence could influence the effectiveness of interventions, and adherence rates may vary across different intervention types; adherence may be better for simple interventions (eg, daily medication) compared with more complex ones (eg, regular exercise). Nonadherence could have influenced the study findings, but this does not appear to have been explored.

Clinically meaningful results favoring the n-of-1 group were underemphasized, and between-subject variability in response to treatment was not reported. This was an ambitious but valuable study, which has illustrated a number of key issues for the field. We should avoid throwing the baby out with the bathwater[4] and instead capitalize on the contribution this study makes to optimize future n-of-1 research.

Suzanne McDonald, PhD
James McGree, PhD
Lydia Bazzano, PhD, MD

Author Affiliations: The University of Queensland, Brisbane, Queensland, Australia (McDonald); Queensland University of Technology, Brisbane, Queensland, Australia (McGree); Tulane University, New Orleans, Louisiana (Bazzano).

Corresponding Author: Suzanne McDonald, PhD, UQ Centre for Clinical Research, Bldg 71/918, Royal Brisbane & Women's Hospital Campus, The University of Queensland, Brisbane, Queensland, Australia 4029 (suzanne.mcdonald@uq.edu.au).

Conflict of Interest Disclosures: None reported.

1. Kravitz RL, Schmid CH, Marois M, et al. Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern Med*. 2018;178(10):1368-1377. doi:10.1001/jamainternmed.2018.3981

2. Kazdin A. *Single-case Research Designs: Methods for Clinical and Applied Settings*. 2nd ed. New York, NY: Oxford University Press; 2010.

3. McDonald S, Quinn F, Vieira R, et al. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychol Rev*. 2017;11(4): 307-323. doi:10.1080/17437199.2017.1316672

4. Mirza RD, Guyatt GH. A randomized clinical trial of n-of-1 trials: tribulations of a trial. *JAMA Intern Med*. 2018;178(10):1378-1379. doi:10.1001/jamainternmed.2018.3979

In Reply Many of the questions raised by Chapple and Blackston and by McDonald and colleagues about our recent Original Investigation[1] are addressed in Pocock and Stone's recent review on what to do when the primary outcome fails.[2] Certainly, a trial in which the primary outcome falls short of statistical significance can be distressing to investigators. However, as highlighted by Chapple and Blackston, the interpretation of trial results may be colored by undue attention to a single primary outcome and arbitrary *P* value cut points. These constraints make sense in confirmatory studies of new drugs and devices (where the consequences of false positives can be dire) but not necessarily in more exploratory studies (like the Personalized Research for Monitoring Pain Treatment study[3]).

Both letters raise a number of other methodological issues, including lack of statistical power, underemphasis of important secondary outcomes, problems with application of the n-of-1 intervention, and potentially poor patient adherence. As we noted in our article,[1] the study fell 12% short of enrollment goals, but it is not clear that reaching the planned sample size of 244 would have resulted in a significant *P* value. Single studies rarely provide definitive estimates of effect size, and for this reason we believe further studies (and subsequent meta-analyses) are warranted.

We agree that statistically significant between-group differences were seen in medication-related shared decision making and in the probability of achieving a 5-point pain interference score reduction. These findings are clinically important and deserving of further study. Likewise, although certain n-of-1 trial design choices (eg, offering nonpharmacologic treatments and relatively short treatment periods) may have contributed to the large proportion of inconclusive n-of-1 trials, our goal was to balance experimental rigor with patient choice and convenience.

Finally, although patients randomized to the n-of-1 arm adhered well to their assigned treatment regimens (averaging 1.4 on a 1-5 scale, with 1 indicating "always" following the directed treatment), we did not track adherence to the "winning" treatment following the trial. If the benefit of n-of-1 trial participation (if any) is mediated purely through the identification of clinically superior treatments, poor adherence to the "winner" in the aftermath of an n-of-1 trial could, as McDonald and colleagues suggest, limit the potential benefit. However, we suspect that other potential mechanisms are operative (eg, creating a more therapeutic physician-patient relationship, enhancing patients' self-efficacy as autonomous agents) and may deserve more attention than previously recognized.

Richard L. Kravitz, MD, MSPH
Christopher H. Schmid, PhD
Ida Sim, MD, PhD

Author Affiliations: University of California, Davis, Sacramento (Kravitz); Brown University, Providence, Rhode Island (Schmid); University of California, San Francisco, San Francisco (Sim).

Corresponding Author: Richard L. Kravitz, MD, MSPH, University of California, Davis, Department of Internal Medicine, Division of General Medicine, 4150 V St, Ste 2400, Sacramento, CA 95817 (rlkravitz@ucdavis.edu).

Conflict of Interest Disclosures: None reported.

1. Kravitz RL, Schmid CH, Marois M, et al. Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern Med*. 2018;178(10):1368-1377. doi:10.1001/jamainternmed.2018.3981

2. Pocock SJ, Stone GW. The primary outcome fails—what next? *N Engl J Med*. 2016;375(9):861-870. doi:10.1056/NEJMra1510064

3. Barr C, Marois M, Sim I, et al. The PREEMPT study—evaluating smartphone-assisted n-of-1 trials in patients with chronic pain: study protocol for a randomized controlled trial. *Trials*. 2015;16:67. doi:10.1186/s13063-015-0590-8

Inconsistencies in Reporting Studies of Lactic Acidosis

To the Editor In their recently published Original Investigation regarding metformin use, renal function, and acidosis, Lazarus and colleagues[1] explained why their findings were different than ours[2] and wrote that our study "was limited by sparse [estimated glomerular filtration rate] data and did not account for changes in [estimated glomerular filtration rate] over time."[1(909)] This is not true. Table 2 in our article[2] summarized that we were able to classify more than 90% of metformin exposure time to renal function. In addition, our methods section clearly stated that we determined renal function during follow-up time and ran our analysis using a time-varying Cox regression analysis in which we modeled both changes in metformin exposure and changes in renal function over calendar time.

Nevertheless, because both studies[1,2] used routine health care data, renal function recordings were probably a proxy indicator of the true renal function during the development of lactic acidosis. A more sensible explanation for the differences between the studies is that Lazarus and colleagues[1] were more likely to measure metabolic or respiratory acidosis instead of lactic acidosis. The authors used *International Classification of Diseases, Ninth Revision, Clinical Modification* codes to define their outcome. This coding system, in contrast with UK Read terminology, cannot define lactic acidosis; therefore, we feel that the words *lactic acidosis* should have been replaced by *acidosis*. In a recently published follow-up letter, Lazarus and colleagues[3] wrote that our study[2] evaluated acidosis. This is not true either; we evaluated the risk of lactic acidosis.

Frank de Vries, PharmD, PhD

Author Affiliation: Division of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands.

Corresponding Author: Frank de Vries, PharmD, PhD, Division of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Universiteitsweg 99, Utrecht 3584 CA, the Netherlands (f.devries@uu.nl).