
Response

Crowdsourcing Public Health Experiments: A Response to Jonathan Darrow's *Crowdsourcing Clinical Trials*

Ameet Sarpatwari,[†] Christopher T. Robertson,^{††}
David V. Yokum^{†††} & Keith Joiner^{††††}

INTRODUCTION

We are pleased to have this opportunity to respond to Jonathan Darrow's article, *Crowdsourcing Clinical Trials* (CCT).¹ We seek to highlight its important contributions and to commence debate over some of its arguments. In particular, we qualify the ethical arguments that characterize early clinical use of drugs as if they were research, and suggest instead that, in either domain, the ethical (and legal) analysis should remain focused on whether all material information is provided so patients may make informed decisions. We also highlight the lim-

[†] Ameet Sarpatwari, J.D., Ph.D., Research Fellow, Harvard Medical School, and Post-Doctoral Fellow, Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital.

^{††} Christopher T. Robertson, J.D., Ph.D., Visiting Professor, Harvard Law School, and Associate Professor, James E. Rogers College of Law, University of Arizona.

^{†††} David V. Yokum, J.D., M.A., M.A., Ph.D. Candidate, Department of Psychology, University of Arizona.

^{††††} Keith Joiner, M.D., M.P.H., Professor of Medicine and Economics, and Co-Director, Center for Management Innovations in Health Care, Eller College of Management, University of Arizona. The authors wish to thank Jonathan Darrow, Nabarun Dasgupta, Clark Freifeld, Joshua Gagne, and Aaron Kesselheim for their insightful comments.

[Editor's Note: This is a response to Jonathan J. Darrow, *Crowdsourcing Clinical Trials*, 98 MINN. L. REV. 805 (2014).]

Copyright © 2014 by Ameet Sarpatwari, Christopher T. Robertson, David V. Yokum, & Keith Joiner.

1. See generally Jonathan J. Darrow, *Crowdsourcing Clinical Trials*, 98 MINN. L. REV. 805 (2014) [hereinafter CCT].

its of what can be gleaned from the observational data collection efforts envisioned by CCT.

Ultimately, we exploit the core insights of CCT to expand the potential use of crowdsourcing from observational studies to truly randomized interventional trials. Randomized experiments allow causal inference because they assign subjects to a treatment and control group, and collect data from each. Furthermore, we draw attention to the fact that much of public health is driven not by pharmaceuticals, but by lifestyle factors. We suggest that CCT's envisioned platform for crowdsourcing also has great potential to engage the public in producing new and trustworthy knowledge in the domains of diet, exercise, nutritional supplements, and integrative medicine, which are primary drivers of health outcomes and spending.

I. THE IMPORTANCE OF CROWDSOURCING CLINICAL TRIALS

CCT offers several key insights of importance to patients, clinicians, academics, and policymakers. In particular, it aptly highlights the amorphous divide between pre-approved and approved drugs, a problem that exemplifies the line-drawing problem inherent in many legal contexts.² Although our understanding of the safety and efficacy of a drug grows incrementally from ignorance towards greater knowledge, at some point the Food and Drug Administration (FDA) is called upon to make a binary decision about whether the product appears safe and efficacious enough to be marketed.

CCT wisely avoids the simplistic suggestion that the FDA should move the epistemic line to mandate more (or less) scientific testing prior to marketing.³ After all, it is costly, and po-

2. See, e.g., Dale Hattis, *Drawing the Line: Quantitative Criteria for Risk Management*, 38 ENV'T 10, 13 (1996) (noting that the requirements of the Federal Food Drug and Cosmetic Act "lead[] to a highly dichotomized view of food safety, in which foods are either perfectly safe or completely tainted"); David A. Weisbach, *An Efficiency Analysis of Line Drawing in the Tax Law*, 29 J. LEGAL STUD. 71, 72 (2000) (stating that for many issues in tax law, "the policy maker may change only the boundaries between differently treated items given that the distinctions must exist").

3. Compare Andrew von Eschenbach, *Medical Innovation: How the U.S. Can Retain Its Lead*, WALL ST. J., Feb. 14, 2012, <http://online.wsj.com/news/articles/SB10001424052970203646004577215403399350874> (arguing for a drug approval process based solely on safety, not efficacy), with Alec B. O'Connor, *Building Comparative Efficacy and Tolerability into the FDA Approval Process*, 303 J. AM. MED. ASS'N 979, 979-80 (2010) (arguing for mandatory active-comparator trials).

tentially a zero-sum game, to move the line in either direction. A longer period of testing precludes potential lifesaving opportunities for using the drugs clinically, and those foregone sales may reduce the manufacturers' incentive to innovate *ex ante*. A shorter period of testing prior to marketing, on the other hand, increases the risk that unsafe and ineffective products will make it onto the market, where they may displace standard-of-care drugs. From the perspective of policymaking, it is hard to say what the optimal amount of pre-market testing is as a general matter. That is why the FDA's expert regulators make those decisions on a case-by-case basis.

Instead of tinkering with the placement of the line, CCT insightfully suggests that we make it somewhat less consequential and more diffuse where possible. While there must be some minimal threshold of proven safety and efficacy to reach the market, the crossing of that line need not be the end of our epistemic investments. CCT argues that the FDA should require rigorous scientific study of drugs even after they reach the market, which may be conducted "by a government affiliated non-profit or the government itself."⁴ This point has been long recognized.⁵ CCT's contribution, however, is to show how newly-emerging techniques and technologies of crowdsourcing can power such an emerging reform. The article effectively addresses concerns regarding data quality, scientific complexity, and ethics. It further explains why crowdsourcing is particularly useful to detect rare and meaningful adverse events, which even well-powered traditional trials may be unable to capture.⁶ With sufficiently large populations, it may even be possible to identify markers that predict which patients are more likely to suffer adverse events or more likely to benefit from drugs.⁷ CCT is thus on the cutting edge of personalized medicine. Ultimately, in this domain, the choice is between lighting a candle and cursing the darkness, and CCT wisely chooses the former.

4. CCT, *supra* note 1, at 851.

5. See, e.g., Rose Lee Bell & E. O'Brian Smith, *Clinical Trials in Post-Marketing Surveillance of Drugs*, 3 CONTROLLED CLINICAL TRIALS 61 (1982) (discussing the realistic advantages of post-marketing surveillance of drugs).

6. See CCT, *supra* note 1, at 842–44 (synthesizing comparisons with existing post-market surveillance and crowdsourcing—where patients can input their adverse events personally).

7. See *id.* at 846.

II. SOME QUALIFICATIONS OF THE ARGUMENTS

A number of CCT's arguments, however, merit extended discussion and, in some cases, qualification. They include the relevance of the research-treatment divide on informed consent, differences between CCT's proposal and existing FDA initiatives, and the promise of observational crowdsourcing.

A. THE RESEARCH-TREATMENT DIVIDE & DISCLOSURE OF UNCERTAINTY

Acknowledging that current law and ethical doctrine have been interpreted in a contrary fashion,⁸ CCT cites the Nuremberg Code,⁹ the Federal Common Rule,¹⁰ and various other strictures on human subjects research to argue that "patients are being used as de facto test subjects following drug approval without their knowledge or informed consent."¹¹ Here, however, we suggest that the line between research and treatment is inconsequential. In the literature on informed consent, it is well settled that patients and research subjects should alike be informed of all material facts.¹² CCT can be taken as arguing that our uncertainty about a new drug's safety is itself a material fact for the patient's decision, such that omission of that information makes the patient's choice uninformed and consent thereby defective. If newness on the market is material to the patient's decision, then the FDA, drug manufacturers, and physicians have a duty to communicate such information to patients.

On the other hand, simply disclosing more information to patients is not a panacea. There is a lot of information that one

8. CCT, *supra* note 1, at 816 ("Under current law and practice, however, the 'human testing' (which term is rarely used) that follows FDA approval has traditionally escaped the confining strictures of human subject research on the apparent basis that the drug has been formally tested as much as is practicable, and that the discovery of adverse events via legally-required monitoring does not constitute human subject research at all.")

9. 2 TRIALS OF WAR CRIMINALS BEFORE THE NUREMBERG MILITARY TRIBUNALS UNDER CONTROL COUNCIL LAW NO. 10 (1949).

10. 45 C.F.R. § 46.116 (2013).

11. CCT, *supra* note 1, at 811.

12. *Canterbury v. Spence*, 464 F.2d. 772, 786 (D.C. Cir. 1972) ("The scope of the physician's communications to the patient, then, must be measured by the patient's need, and that need is the information material to the decision."); see David S. Shimm & Roy G. Spece, Jr., *Conflict of Interest and Informed Consent in Industry-Sponsored Clinical Trials*, 12 J. LEGAL MED. 477, 510-13 (1991) (noting that informed consent requires full disclosure of risks and benefits to experimental subjects).

could consider material, yet patients may rationally prefer to defer to their physicians to synthesize this information and to provide a bottom-line recommendation. One way to conceive of materiality is to ask whether the information would change the decisions of a substantial number of patients.¹³ On this conception, materiality raises an empirical question, and we are unsure of whether the newness of a drug would be material. Whether additional information will be consequential—and perhaps more importantly whether those consequences will be salutary on net—are open empirical questions, which are beyond the scope of CCT.

Regardless of whether CCT persuades on these patient-focused ethical arguments, we remain persuaded by CCT's population-level arguments, suggesting that we can promote aggregate social welfare by collecting this information. The costs of collecting and analyzing this information are clearly outweighed by the benefits, as they will inform FDA policy decisions and inform the larger scientific consensus, which should then drive physician prescribing behavior.

B. FDA & MEDWATCHER

CCT additionally distinguishes its proposal from existing FDA efforts to harness crowdsourcing, notably MedWatch.¹⁴ Notwithstanding CCT's justified critique of the program, it should be noted that the FDA has taken initial steps to adopt a more flexible and functional reporting framework. On April 22, 2013, the Agency launched MedWatcher, a supplement to MedWatch, an online system for the general public to report adverse events.¹⁵ MedWatcher includes both Web and mobile apps,¹⁶ and is run by Epidemico, a spin-off company of the Computational Epidemiology Group at Boston Children's Hos-

13. In a forthcoming article, some of the authors here have argued for such a behavioral test of materiality and show how it can be operationalized in laboratory tests of human subjects in vignette-based experiments. See Roy Spece, David Yokum, Andrea-Gale Okoro & Christopher Robertson, *An Empirical Method for Materiality: Would Conflict of Interest Disclosures Change Patient Outcomes?*, 40 AM. J.L. & MED. (forthcoming 2014), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2378524.

14. CCT, *supra* note 1, at 836–39.

15. See *About MedWatcher*, EPIDEMICO, <https://medwatcher.org/about.php> (last visited Apr. 18, 2014).

16. *Get the App*, EPIDEMICO, <https://medwatcher.org/getapp.php> (last visited Apr. 18, 2014).

pital.¹⁷ In addition to submitting data, users may also search the system, sorting reports by product, gender, and age. To date, there have been approximately 35,000 downloads of apps, and there are on average over 1000 active MedWatcher users on any given day.¹⁸ While this new initiative does not yet offer product-tailored reporting forms and is not part of a conditional approval process emphasizing disclosure, which CCT champions, it does “aggregate[] patient feedback and present[] it in a user-friendly manner,” enabling patients to “become aware of actual drug risks.”¹⁹ In this respect, MedWatcher constitutes a potentially meaningful embrace of the principle of crowdsourcing.

C. THE LIMITS OF OBSERVATIONAL CROWDSOURCING

Finally, CCT sensibly states that under its proposal, the FDA would not rely solely upon crowdsourced data to approve new indications of drugs.²⁰ Along these lines, it is imperative to detail more fully why the form of crowdsourcing CCT advocates should not replace existing post-marketing surveillance efforts (e.g., risk evaluation and mitigation strategies [REMS] and Mini-Sentinel) but rather supplement them. As CCT notes, “[O]nly a fraction of patients are likely to thoroughly embrace and participate in any voluntary crowdsourcing platform.”²¹ Observational investigations utilizing the platform will therefore be subject to selection bias, “[a] distortion in the estimate of the effect due to the manner in which subjects are selected for the study.”²²

On this point, one might quibble with CCT’s titular characterization of its proposed investigations as “clinical trials.” Typically, the biomedical literature uses the term differently. Thandani, for instance, explains, “Clinical studies can be divided into two broad categories: trials, in which the investigator intervenes to prevent or treat a disease, and observational studies, in which the investigator makes no intervention and

17. *About MedWatcher*, *supra* note 15.

18. E-mail from Nabarun Dasgupta, Principal, Epidemico, to Ameet Sarpatwari, Research Fellow, Harvard Med. Sch. (Oct. 24, 2013, 11:33 EST) (on file with author).

19. CCT, *supra* note 1, at 830.

20. *Id.* at 866 (explaining that crowdsourcing “attempts to capture the data that is already being generated post-approval” and ensures that the information is recorded and analyzed).

21. *Id.* at 843.

22. A DICTIONARY OF EPIDEMIOLOGY 225 (Miquel Porta ed., 5th ed. 2008).

patients are allocated treatment based on clinical decisions.”²³ As we understand the proposal, the studies conceived in CCT generally fall in the latter category of observational research.

Much of our current scientific knowledge is based on such observational research, and some fields of knowledge (such as econometrics) rely almost exclusively on such methods. However, the limitations of such non-randomized scientific research deserve emphasis. In particular, the problem of self-selection is already apparent in the nascent biomedical literature using crowdsourcing. In 2012, Armstrong et al., analyzed data from CureTogether, a website in which patients voluntarily report data on the effectiveness of the drugs they are taking. Focusing on acne treatments, they compared these crowdsourced data to study findings reported in the peer-reviewed scientific literature, revealing a striking disparity:

While approximately 80% of tretinoin users observed clinical improvement after a 12-week treatment period in clinical trials, 46% of online users reported improvement in an unspecified time period. For most topical treatments, medication with high efficacy in clinical trials did not produce high effectiveness ratings based on the crowdsourced online data.²⁴

The smaller effects seen in the crowdsourced data may be due to unsatisfied patients seeking out the CureTogether site and reporting their frustration in greater numbers than satisfied pimple-free patients, who have other priorities for their time.²⁵ The foregoing example deals with the effectiveness of drugs, but similar dynamics apply to adverse events in observational studies. If one assumes that patients suffering adverse events are more likely to submit a report, then the data are likely to overestimate the rate of adverse events.

23. Ravi Thadhani, *Chapter 14: Formal Trials Versus Observational Studies*, in *FABRY DISEASE: PERSPECTIVES FROM 5 YEARS OF FOS* (Atul Mehta et al., eds., 2006), available at <http://www.ncbi.nlm.nih.gov/books/NBK11586/>.

24. April W. Armstrong et al., *Harnessing the Power of Crowds: Crowdsourcing as a Novel Research Method for Evaluation of Acne Treatments*, 13 AM. J. CLINICAL DERMATOLOGY 405, 405 (2012).

25. Admittedly, the smaller effects may also be due, in part, to well-known problems with clinical research trials. See generally *Unreliable Research: Trouble at the Lab*, ECONOMIST, Oct. 19, 2013, <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble> (discussing the problem of irreproducibility of scientific experiments across disciplines). For these reasons, many clinical trials are themselves unable to be successfully replicated. See C. Glenn Begley & Lee M. Ellis, *Raise Standards for Preclinical Cancer Research*, 483 NATURE 531, 532 (2012) (reporting being able to replicate scientific findings in only six of fifty-three landmark preclinical research studies).

Even though the observational studies CCT envisions will not provide accurate assessments of the rates of adverse events, they will nonetheless prove invaluable for generating information on potential causal associations (i.e., signals). To provide sufficient drug safety information from which to base regulatory decisions, however, these signals must be refined and evaluated,²⁶ steps that can utilize Mini-Sentinel and REMS, respectively. The former initiative, though complex,²⁷ has assembled the infrastructure necessary to test for the existence of associations between drugs and adverse events having already accumulated “over 300 million person-years of observation time, 2.4 billion unique encounters including 38 million acute inpatient stays, and 2.9 billion dispensing of prescriptions.”²⁸ Through REMS, meanwhile, the FDA can compel manufacturers to conduct more systematic, registry-based investigations to assess causality.²⁹ While each is insufficient alone at this time, collectively, crowdsourcing, Mini-Sentinel, and REMS provide the FDA with the tools necessary to craft a powerful post-marketing surveillance system.

III. RANDOMIZED CROWDSOURCED TRIALS OF LIFESTYLE INTERVENTIONS

CCT concludes by noting that while “[t]he instant proposal is a relatively modest one that involves crowdsourcing of only the post-approval phase of drug evaluation, . . . it is easy to imagine a more expanded version.”³⁰ We would like to take this opportunity to pick up the mantle to extend these ideas in new directions.

Going beyond CCT’s domain of observational studies of FDA-approved pharmaceuticals, there is considerable potential in using crowdsourcing to perform truly randomized trials (i.e.,

26. See Richard Platt et al., *The U.S. Food and Drug Administration’s Mini-Sentinel Program: Status and Direction*, 21 PHARMACOEPIDEMIOLOGY & DRUG SAFETY 1, 2 (2012) (defining signal refinement as “the assessment of predefined exposure-outcome pairs to determine whether there is evidence of association” and signal evaluation as “assessing whether an association is likely to be causal, and addressing questions such as dose-response, duration-response, and inter-individual variability in risk”).

27. See CCT, *supra* note 1, at 840–41 (describing Sentinel as “layered and cumbersome” and noting that the FDA acknowledges that it is a “complex endeavor”).

28. Platt et al., *supra* note 26, at 4.

29. See 21 U.S.C. § 355-1(f)(3) (2012) (describing elements to assure safe use).

30. CCT, *supra* note 1, at 866–67.

experiments), and to do so in the domain lifestyle interventions that could combat such diverse and pressing public health problems as obesity, migraines, smoking, and back pain. This expansion of CCT's platform could also allow investigation of the many health-related interventions that fall outside of the FDA's current legal jurisdiction to mandate scientific research. While most of these interventions are not readily amenable to analysis using post-marketing surveillance of the sorts discussed above, it is possible that the same technological platform and recruitment mechanisms proposed by CCT could power these broader investigations.

The need is pressing. Lifestyle factors are primary drivers of health and health spending in the United States.³¹ A third of all deaths are attributed to tobacco use, poor diet, and physical inactivity.³² Over 30 million Americans, meanwhile, suffer from migraines each year,³³ and one in four reports having experienced lower back pain lasting the whole day over the last three-months.³⁴ Collectively, the economic impact of these lifestyle-related diseases is staggering; Finkelstein et al. estimate that the costs of treating obesity-related conditions alone could total \$147 billion annually.³⁵

These examples remind health policy scholars that individuals do not intrinsically value the prescription drugs that are the focus of CCT. What people seek is health, and there are many paths to that outcome. One way to conceptualize this dynamic is to say that individuals make decisions to maximize their "stock" of health capital.³⁶ People make time-based trade-

31. Louis Goodman & Timothy Norbeck, *Who's to Blame for Our Rising Healthcare Costs?*, FORBES, Apr. 3, 2013, <http://www.forbes.com/sites/realspin/2013/04/03/whos-to-blame-for-our-rising-healthcare-costs/> (categorizing "life style and chronic conditions" as vital drivers of healthcare costs, as well as "the most preventable" kind).

32. Ali H. Mokdad et al., *Actual Causes of Death in the United States*, 2000, 291 J. AM. MED. ASS'N 1238, 1240 (2004).

33. R.B. Lipton et al., *Migraine Prevalence, Disease Burden, and the Need for Preventive Therapy*, 68 NEUROLOGY 343, 346-47 (2007) (reporting that 11.7% of Americans suffer from migraines annually and that, of those, 25.7% should be offered prevention).

34. Richard A. Deyo et al., *Back Pain Prevalence and Visit Rates: Estimates from U.S. National Surveys, 2002*, 31 SPINE 2724, 2725 (2006).

35. Eric A. Finkelstein et al., *Annual Medical Spending Attributable to Obesity: Payer and Service-Specific Estimates*, 28 HEALTH AFF. 822, 822 (2009).

36. See generally Michael Grossman, *On the Concept of Health Capital and the Demand for Health*, 80 J. POL. ECON. 224 (1972) (conceptualizing

offs between work, leisure, medical care, and home goods production. While illness, aging, and unhealthy behaviors (e.g., smoking) lead to consumption of health capital, other actions (e.g., education, income generation, and healthy eating) represent investments that produce health capital. For example, individuals commonly choose to substitute unhealthy processed foods for home-cooked goods, not primarily because the benefits of the latter are unclear, but because the time allotted to prepare the latter cannot be justified in comparison to other activities that also impact health capital.³⁷ Crowdsourcing can elicit choices around health capital maximization and engage the public in producing knowledge about how to best produce the health outcomes that they experience and value.

A. CANDIDATES FOR INVESTIGATION

As candidates for crowdsourced testing beyond that proposed by CCT, we proffer three categories of activities as interventions. They include activities for which good biological evidence exists but behavioral testing is lacking to activities, which though widely practiced, have uncertain biological justification.

To start, there are activities for which there is an “accepted wisdom” and clear biological plausibility but which require behavioral change. Some of these activities are believed to be beneficial—exercise, healthy eating, and weight control. Others are known to be harmful—smoking and excessive drinking. In this domain, crowdsourced randomized experiments would generate useful knowledge about which behavioral interventions are most effective in achieving better health outcomes. For example, it has long been understood that effective weight control depends on roughly balancing calories consumed with those expended, but this goal proves difficult for many people to achieve.³⁸ Recent scientific research, popularized by several

health as human capital and exploring various effects that individual choice has on one’s stock throughout the course of one’s life).

37. See Arleen A. Leibowitz, *The Demand for Health and Health Concerns After 30 Years*, 23 J. HEALTH ECON. 663, 669 (2004) (suggesting that working mothers may be inclined to “substitute prepared foods for their own time in producing meals for their children” despite the fact that prepared meals are generally less healthy).

38. See, e.g., *Healthy Weight, Balancing Calories*, CDC, <http://www.cdc.gov/healthyweight/calories/> (last updated Jan. 15, 2004) (explaining that the key to maintaining a healthy weight is “balancing the number of calories you consume with the number of calories your body uses or ‘burns off’”).

books, has suggested that a diet based on fasting once per week, may be more effective than trying to manage caloric intake all week.³⁹ Nonetheless, this method has not been tested in large-scale randomized experiments, which could identify the benefits and risks of this particular approach to behavior-modification.

Next, there are activities that are intended to eliminate or diminish some undesirable symptom—back pain, migraines, and insomnia—but which are supported by only anecdotal evidence and may not be effective despite being biological plausible. Lifestyle modifications to manage migraines, for example, include sleep hygiene, stress management, regular exercise, and a variety of diets.⁴⁰ Some of these diets involve the avoidance of potential food triggers (e.g., alcohol, chocolate, caffeine, and monosodium glutamate).⁴¹ Conversely, others recommend the addition of vitamins and supplements, including magnesium, riboflavin, CoQ10, feverfew, and butterbur. Studies, many of which are randomized, have investigated these individual agents relative to placebo, showing benefit.⁴² Small sample sizes and lack of comparison between agents, however, limit the generalizability of the conclusions.⁴³ The crowdsourced platform could be used to test combinations of therapies, often in factorial design, with statistical power deriving from large sample sizes. In such instances, it would often be more attractive to use a “knockout” study, which asks participants to remove a poten-

39. See, e.g., JAMES B. JOHNSON & DONALD R. LAUB, *THE ALTERNATE-DAY DIET: TURN ON YOUR “SKINNY GENE,” SHED THE POUNDS, AND LIVE A LONGER AND HEALTHIER LIFE* (2009); MICHAEL MOSLEY & MIMI SPENCER, *THE FASTDIET: LOSE WEIGHT, STAY HEALTHY, AND LIVE LONGER WITH THE SIMPLE SECRET OF INTERMITTENT FASTING* (2013).

40. *Migraine: Non-Drug Treatments and Lifestyle Changes*, N.Y. TIMES, <http://www.nytimes.com/health/guides/disease/migraine/non-drug-treatments-and-lifestyle-changes.html> (last updated Dec. 23, 2013) (recommending the following lifestyle changes in order to relieve migraines: avoid food triggers such as monosodium glutamate, chocolate, and caffeine; eat regularly; stay physically active; limit estrogen-containing medications).

41. See *id.*

42. See, e.g., D.A. Marcus et al., *A Double-Blind Provocative Study of Chocolate as a Trigger of Headache*, 17 CEPHALALGIA 855 (1997) (finding that, contrary to popular belief, chocolate does not appear to be a trigger for migraines); A. Verotti et al., *Impact of a Weight Loss Program on Migraine in Obese Adolescents*, 20 EUR. J. NEUROLOGY 394 (2013) (finding that intervention programs of health eating and cognitive stimulation, which yield lower BMIs, were linked to fewer migraines).

43. Cf. Verotti et al., *supra* note 42, at 397 (recognizing that their study is limited by its lack of a control group).

tial cause of migraines or remove a potential homeopathic treatment, rather than asking them to start a new one.

Last, there are activities that are purported to have benefit but for which the evidence is lacking or contradictory, and biological plausibility is often tenuous at best. In general, the lack of evidence for these activities stems from the inability or unwillingness to conduct large-scale trials, given the complexity, cost, and, for advocates, the unstated but real fear that they will be ineffective at best and harmful at worst. Two prominent candidates for scientific testing within this domain are the practice of integrative medicine and consumption of many dietary supplements. Integrative medicine aspires to bridge the gap between traditional western medicine, behavioral change, and health promoting activities from other cultures.⁴⁴ While its less “standard” interventions, which include massage, reiki, and aroma therapy practiced in conjunction with traditional medicine, are widely popular in the United States, evidence is typically lacking that they have benefit. Almost half of all Americans and two-thirds of all cancer patients, meanwhile, consume dietary supplements, whose sales surpass \$30 billion annually.⁴⁵ These products, however, are only lightly regulated by the FDA such that their efficacy remains largely unclear.⁴⁶ Illustrative of the issue, the Children’s Hospital of Philadelphia recently announced that most dietary supplements would be removed from the hospital formulary, the first hospital in the nation to do so.⁴⁷ Crowdsourcing provides an opportunity to test these interventions in a comparative fashion.

44. See generally Ralph Snyderman & Andrew T. Weil, *Integrative Medicine: Bringing Medicine Back to its Roots*, 162 ARCHIVES INTERNAL MED. 395 (2002).

45. See U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-13-244, DIETARY SUPPLEMENTS: FDA MAY HAVE OPPORTUNITIES TO EXPAND ITS USE OF REPORTED HEALTH PROBLEMS TO OVERSEE PRODUCTS 1 (2013), available at <http://www.gao.gov/assets/660/653113.pdf>; Mary L. Hardy, *Dietary Supplement Use in Cancer Care: Help or Harm*, 22 HEMATOLOGY/ONCOLOGY CLIN. N. AM. 581, 581 (2008).

46. James D. Lewis & Brian L. Strom, *Balancing Safety of Dietary Supplements with the Free Market*, 136 ANNALS INTERNAL MED. 616, 617 (2002) (“[T]here are no data for most supplements, merely because they do not claim to prevent or treat specific medical conditions.”).

47. Press Release, Children’s Hosp. of Phila., Children’s Hosp. of Phila. Becomes First in Nation to Disallow Use of Dietary Supplements (Oct. 8, 2013), available at http://www.eurekalert.org/pub_releases/2013-10/chop-cho100813.php.

B. ADVANTAGES OF RANDOMIZED CROWDSOURCED TRIALS

Randomized crowdsourced trials for lifestyle interventions present several advantages. First, given the FDA's limited jurisdiction to mandate scientific research and the impracticability of the widespread use of traditional clinical trials in this setting, the vast majority of lifestyle interventions would not otherwise be rigorously tested. These lifestyle choices are primary drivers of health outcomes and health spending, and our current ignorance about their risk and benefits is unacceptable.

Second, although it would not be possible to blind patients and investigators to many lifestyle interventions, randomization would nevertheless permit the distribution of both known and unknown confounders equally between the intervention and active-comparator (of placebo) arms of the trial.⁴⁸ Randomized crowdsourced trials would, in this respect, resemble open-label pharmaceutical trials, which the FDA considers robust.⁴⁹ Randomization may also be attractive to participants, since it will function like a game or lottery. Individuals pre-commit to participate, and then are randomly assigned to do something in their own lives and to report back the results to the crowd.

Third, as CCT emphasizes, crowdsourcing would enable the enrollment of potentially thousands of participants, increasing the ability of trials to detect associations. Indeed, increased statistical power resulting from crowdsourcing will permit investigators to explore the best combination of interventions for tackling complex public health epidemics.⁵⁰

Finally, such trials would help fuel patient-driven research. Crowdsourcing can be used not only to secure trial participants but also to generate research hypothesis, to identify

48. Chris Roberts & David Torgerson, *Understanding Controlled Trials: Randomisation Methods in Controlled Trials*, 317 BRIT. MED. J. 1301, 1301 (1998) ("The main purpose of randomisation is to avoid bias by distributing the characteristics of patients that may influence outcome randomly between treatment groups so that any difference in outcome can be explained only by treatment.").

49. See U.S. FDA, GUIDANCE FOR INDUSTRY: E9 STATISTICAL PRINCIPLES FOR CLINICAL TRIALS 11 (1998) (noting that while double-blind trials are optimal, open-labeled trials may prove the only option owing to issues of practicability and ethics).

50. See, e.g., Guangwei Li et al., *The Long-Term Effect of Lifestyle Interventions to Prevent Diabetes in the China Da Qing Diabetes Prevention Study: A 20-Year Follow-Up Study*, 371 LANCET 1783 (2008); Martha L. Skender et al., *Comparison of 2-Year Weight Loss Trends in Behavioral Treatments of Obesity: Diet, Exercise, and Combination Interventions*, 96 J. AM. DIETETIC ASS'N 342 (1996).

meaningful clinical outcomes, and to disseminate trial findings. This potential for empowerment is consistent with the broader movement to increase stakeholder participation in all stages of public health research and policymaking, most notably the recently created Patient-Centered Outcomes Research Institute.⁵¹ We hope to democratize health-related research, as long as we can do so without sacrificing rigor.

C. CHALLENGES OF RANDOMIZED CROWDSOURCED TRIALS:
USING BEHAVIORAL SCIENCE AND SELF-QUANTIFICATION TO
ENHANCE ADHERENCE AND DATA COLLECTION

Using crowdsourcing as a vehicle to conduct randomized trials of lifestyle interventions, however, will not be free of challenges. In particular, two methodological concerns stand out: participant adherence to the trial protocol and the quality of data collected, given that the platform will largely depend on self-reported data. As CCT notes, “Laypersons may be perceived as lacking the appropriate education, experience, or motivation to reliably obtain or enter clinical data.”⁵²

CCT successfully dispels some of this worry. We build upon this defense by arguing that the concerns are not simply obstacles but also opportunities, both for the crowdsourcing platform and underlying public health studies, which will also be concerned with adherence and self-perceived outcomes. On one hand, existing behavioral research could be used to design the platform so as to maximize compliance with data reporting responsibilities; on the other hand, the platform could also be used to field new experiments on what compliance interventions are most effective. Several behavioral tactics are relevant here.

The crowdsourced platform should prompt people to develop a concrete plan for how they will actually implement an intention. In prior lab and field experiments, such planning has been repeatedly shown to increase attainment of desired

51. See Patient Protection and Affordable Care Act § 6301, 42 U.S.C. § 1320e (2012) (“The purpose of the Institute is to assist patients, clinicians, purchasers, and policy-makers in making informed health decisions by advancing the quality and relevance of evidence concerning the manner in which diseases, disorders, and other health conditions can effectively and appropriately be prevented, diagnosed, treated, monitored, and managed through research and evidence synthesis . . .”).

52. CCT, *supra* note 1, at 845.

goals.⁵³ Milkman et al., for example, conducted a field experiment wherein employees were randomly assigned to receive one of two mailings about workplace vaccination clinics.⁵⁴ Each mailing included the same information about clinic availability, but one letter also included optional blank spaces for writing in the day and time of when the recipient planned to visit the clinic.⁵⁵ This simple intervention significantly increased the rate of vaccination.⁵⁶ Implementation intention effects have been found in other health domains—colorectal screening,⁵⁷ cervical screening,⁵⁸ physical exercise,⁵⁹ and breast self-examination⁶⁰—as well as non-health domains, such as voting.⁶¹ Crowdsourced compliance rates could thus be boosted by having participants form an implementation intention. When registering into the online system, users could be prompted to specify when and upon which computer they will update their profiles each week, and exactly what routine they will use for the intervention. Such testing of implementation intentions will also enhance our understanding of behavioral science.

53. The mechanism of effect is a current area of study. *See generally* Thomas L. Webb & Paschal Sheeran, *Mechanisms of Implementation Intention Effects: The Role of Goal Intentions, Self-Efficacy, and Accessibility of Plan Components*, 47 BRIT. J. SOC. PSYCHOL. 373 (2008).

54. Katherine L. Milkman et al., *Using Implementation Intentions Prompts to Enhance Influenza Vaccination Rates*, 108 PROC. NAT. ACAD. SCI. 10,415 (2011). In fact, there was a third condition. *Id.* However, only two are discussed here to simplify the example.

55. *Id.* at 10,416.

56. *Id.*

57. Katherine L. Milkman et al., *Planning Prompts as a Means of Increasing Preventive Screening Rates*, 56 PREVENTIVE MED. 92, 92 (2013).

58. Paschal Sheeran & Sheina Orbell, *Using Implementation Intentions to Increase Attendance for Cervical Cancer Screening*, 19 HEALTH PSYCHOL. 283, 286 (2000).

59. Sarah Milne, Sheina Orbell & Paschal Sheeran, *Combining Motivational and Volitional Interventions to Promote Exercise Participation: Protection Motivation Theory and Implementation Intentions*, 7 BRIT. J. HEALTH PSYCHOL. 163, 172 (2002).

60. Sheina Orbell, Sarah Hodgkins & Paschal Sheeran, *Implementation Intentions and the Theory of Planned Behavior*, 23 PERSONALITY SOC. PSYCHOL. BULL. 945, 950 (1997).

61. David W. Nickerson & Todd Rogers, *Do You Have a Voting Plan? Implementation Intentions, Voter Turnout, and Organic Plan Making*, 21 PSYCHOL. SCI. 194, 196 (2010); *see also* Utpal M. Dholakia & Richard P. Bagozzi, *As Time Goes By: How Goal and Implementation Intentions Influence Enactment of Short-Fuse Behaviors*, 33 J. APPLIED SOC. PSYCHOL. 889, 901 (2003) (showing the effect with errands).

The crowdsourced platform should reinforce salutary social norms. CCT emphasizes that there may be a mutually-reinforcing dynamic between providers and consumers of information, a loop that crowdsourcing can close.⁶² Indeed, information about what others are actually doing has also been found to have an effect on behavior.⁶³ In particular, people are more likely to mimic the behavior believed to be most common, regardless of whether that behavior is socially approved.⁶⁴ Early work showed that this insight has important and counterintuitive implications. A public service campaign warning that underage college drinking is an epidemic to be combated, for example, might inadvertently *increase* underage drinking, because the implicit descriptive norm—many college students binge drink—drives behavior more than the normative argument against binge drinking.⁶⁵ In contrast, when students who are heavy drinkers are truthfully educated that most students actually drink relatively little, their alcohol consumption declines.⁶⁶ Similar results have been found in studies aiming to reduce littering,⁶⁷ park vandalism,⁶⁸ and household energy consumption.⁶⁹ Crowdsourcer users could likewise be informed that most of their peers were consistently reporting data and adhering with the intervention under investigation, which would increase their own compliance. Here, too, crowdsourced trials

62. CCT, *supra* note 1, at 849 (describing a “better-aligned mix of motivations” and having “a dog in the race”).

63. See Robert B. Cialdini, *Crafting Normative Messages to Protect the Environment*, 12 CURRENT DIRECTIONS PSYCHOL. SCI. 105, 106 (2003). Descriptive norms are in contrast to injunctive norms, which pertain to beliefs about what *should* be done. Craig Michael Mc Nees, *Two Kinds of Norms*, APPLIED SOC. PSYCHOL. (Nov. 29, 2012, 8:54 PM), <http://www.personal.psu.edu/bfr3/blogs/asp/2012/11/two-kinds-of-norms.html>.

64. Cialdini, *supra* note 63, at 108–09.

65. See generally Brian Borsari & Kate B. Carey, *Descriptive and Injunctive Norms in College Drinking: A Meta-Analytic Integration*, 64 J. STUD. ALCOHOL 331 (2003).

66. See Clayton Neighbors, Mary E. Larimer & Melissa A. Lewis, *Targeting Misperceptions of Descriptive Drinking Norms: Efficacy of a Computer-Delivered Personalized Normative Feedback Intervention*, 72 J. CONSULTING & CLINICAL PSYCHOL. 434, 438 (2004).

67. See generally Robert B. Cialdini, Raymond R. Reno & Carl A. Kallgren, *A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places*, 58 J. PERSONALITY & SOC. PSYCHOL. 1015 (1990).

68. Cialdini, *supra* note 63, at 106.

69. Hunt Allcott, *Social Norms and Energy Conservation*, 95 J. PUB. ECON. 1082, 1087 (2011).

could also be conducted to help resolve some of the controversy behind when and how social norms exert their effects.

Finally, the crowdsourced platform should take advantage of self-identities: the “salient and enduring aspects of one’s self-perception.”⁷⁰ Depending on the circumstances, one might identify oneself as a student, a political advocate, or here, a crowdsourcing health experimenter. As Rise et al. elaborate: “self-identities (or “me” identifications) are the perspective one takes toward oneself when taking the role of specific or generalized others, implying that one incorporates the meanings and expectations associated with a relevant categorization into the self, thus forming a set of identity standards that guide identity-relevant behaviors.”⁷¹ Behavior can be a function of which self-identity is salient at the moment of action.⁷² In one experiment, for example, residents were contacted just before an election and asked to complete a survey, which was experimentally manipulated to prime voter or baseline identities.⁷³ After the election, public voting records showed that those primed with a voter identity were significantly more likely to vote.⁷⁴ Active participation in the crowdsourcing might shift users from a passive patient identity into a more active health-experimenter identity, a shift that would likely enhance adherence and data reporting.

Even so, it will be critical for all crowdsourced trials to collect a mix of objective (e.g., blood pressure and weight) and subjective (e.g., quality of life surveys) measures that are easily ascertainable using standardized, validated techniques. The

70. Jostein Rise, Paschal Sheeran & Silje Hukkelberg, *The Role of Self-Identity in the Theory of Planned Behavior: A Meta-Analysis*, 40 J. APPLIED SOC. PSYCHOL. 1085, 1087 (2010).

71. *Id.* (internal citations omitted).

72. LeBoeuf, Shafir, and Bayuk, for example, had undergraduates respond to questions about either geopolitical or campus-related gender issues, priming scholarly and socialite identities, respectively. Robyn A. LeBoeuf, Eldar Shafir & Julia Belyavsky Bayuk, *The Conflicting Choices of Alternating Selves*, 111 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 48, 50 (2010). They found that the former preferred periodicals such as *The Economist* and *The Wall Street Journal* while the latter chose *Cosmopolitan* and *USA Today*. *Id.*

73. See Christopher J. Bryan et al., *Motivating Voter Turnout by Invoking the Self*, 108 PROC. NAT’L ACAD. SCI. 12,653 (2011). The precise manipulation, based on prior research, was to refer to voting as a self-relevant noun (e.g., “How important is it to you to be a voter in the upcoming election?”) as opposed to voting as a verb (e.g., “How important is it for you to vote in the upcoming election?”). *Id.*

74. *Id.* at 12,654–65.

quantified self movement is an important form of self-identity that is already developing to exploit a plethora of new self-tracking devices along the lines we suggest.⁷⁵ As one of its founders, Gary Wolf, notes,

Numbers are making their way into the smallest crevices of our lives. We have pedometers in the soles of our shoes and phones that can post our location as we move around town. We can tweet what we eat into a database . . . There are sites and programs for monitoring mood, pain, blood sugar, blood pressure, heart rate, cognitive alacrity, menstruation, and prayers. Even sleep . . . is yielding to the skill of the widget maker. With an accelerometer and some decent algorithms, you will soon be able to record your sleep patterns with technology that costs less than \$100.⁷⁶

Some of these apps have already garnered FDA approval as alternatives to more traditional medical devices.⁷⁷ Many others devices—such as the new iPhone, which includes a dedicated processing unit to track the owner’s physical activities—are not regulated by the FDA at all, because they do not make health claims. Whether it is a scale that automatically uploads weight measurements to the internet, a phone GPS that shows the number of miles jogged, or a phone camera that allows other crowd-participants to code the caloric content of photographed food, these devices will allow collection of real-time objective data from the crowd. In these ways, the quantified self-movement allowing assessment of both compliance with the experimental interventions (independent variables) and health outcomes (dependent variables).

D. FOSTERING INTEGRITY AND TRUST

It will be imperative for developers of the crowdsourced platform to foster public trust, as both a safe place for participating in experimentation, and also as a destination for reliable health information. Talk is often cheap in this domain, and many contested lifestyle health choices have related industries and self-proclaimed experts. Conflicts of interest currently plague the healthcare system, casting doubt even on clinical guidelines in peer-reviewed journals.⁷⁸ Thus, this new

75. See Gary Wolf, *The Data-Driven Life*, N.Y. TIMES, Apr. 28, 2010, <http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html>.

76. Gary Wolf, *Know Thyself: Tracking Every Facet of Life, from Sleep to Mood to Pain*, 24/7/365, WIRED MAG., June 22, 2009, http://www.wired.com/medtech/health/magazine/17-07/lbnp_knowthyself.

77. Sabrina Tavernise, *F.D.A. Will Apply Its Rules to Only Some Health Apps*, N.Y. TIMES, Sept. 24, 2013, at A12.

78. See, e.g., Lisa Cosgrove et al., *Conflicts of Interest and the Quality of*

crowdsourcing institution should be designed from the ground up to engender integrity and credibility.

Accordingly, the crowdsourced platform should operate on a strictly not-for-profit basis. Whether it should be closely tied to the FDA as CCT suggests,⁷⁹ however, is debatable. Concerns over political influence and bureaucratic constraints suggest that an independent nonprofit but entrepreneurial organization may be more nimble and engender greater public participation than a government-run platform would.

A useful model here is Wikipedia, which permits public submissions, editing, and initiation of “Articles for Deletion” reviews.⁸⁰ Although not without weaknesses, Wikipedia science entries were found to have comparable accuracy to Encyclopedia Britannica articles in a 2005 *Nature* investigation.⁸¹ The crowdsourcing platform may be able to thoughtfully bridge the gap between health scientists at elite academic institutions and members of the general public.

As we move into true experiments, the crowdsourced platform will further require its own Institutional Review Board (IRB), which engages the public in setting standards and weighing risks; transparent informed consent forms that do not bury risks in verbiage; and concrete plans for handling adverse events. We believe that these needs create a great opportunity to reengage the public in thinking about and setting standards for human subjects research, rather than merely delegating these tasks to the academics, who currently dominate IRBs. Of course, as a bare minimum, the platform must strictly comply with federal privacy standards in the Health Insurance Portability and Accounting Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH) and human subject protections codified in the Common Rule, to the extent that those statutes and rules are applicable.

Recommendations in Clinical Guidelines, 19 J. EVALUATION CLINICAL PRAC. 674 (2012).

79. See CCT, *supra* note 1, at 851 (“For this reason, the post-approval phase of drug evaluation should probably be administered either by a government affiliated non-profit or the government itself.”).

80. *HBS Cases: How Wikipedia Works (or Doesn't)*, WORKING KNOWLEDGE: THE THINKING THAT LEADS (Harvard Bus. Sch., Boston, Mass), July 23, 2007, at 2, available at <http://hbswk.hbs.edu/item/5605.html>.

81. Jim Giles, *Internet Encyclopaedias Go Head to Head*, 438 NATURE 900, 900–01 (2005).

CONCLUSION

In sum, we share CCT's enthusiasm for crowdsourcing, and believe that citizens and scientists should critically understand the strengths and weaknesses of this new tool for generating and communicating knowledge. Beyond the collection of observational data about pharmaceutical drugs proposed in CCT, we believe that such a crowdsourcing platform holds great promise for generating trustworthy knowledge and behavior change about some of the largest drivers of mortality and morbidity. So much of health outcomes and spending depend on lifestyle choices, and in this domain we desperately need new models for engaging the public to create greater understanding and improved behaviors. Thus, the envisioned crowdsourcing platform could provide a robust and citizen-driven process of inquiry and a destination for robust and reliable information, so that individuals can achieve their own health goals.