

**Title:** Waiting for  $p < 0.05$

James Heywood<sup>1</sup>, Timothy E Vaughan<sup>1</sup>, Paul Wicks<sup>1</sup>

<sup>1</sup>PatientsLikeMe, 155 Second Street, Cambridge MA, 02141, USA

\*Corresponding author:

James Heywood

155 Second Street

Cambridge

MA, 02141

USA

**Email:** [research@patientslikeme.com](mailto:research@patientslikeme.com)

**Figures:** 1

**Tables:** 2

**Disclosures:** JH, TV, and PW are employees of PatientsLikeMe and own stock / stock options in the company. The PatientsLikeMe R&D team has received research funding from Abbott, Acorda, the AKU Society, Avanir, Biogen, Boehringer Ingelheim, Genzyme, Johnson & Johnson, Merck, the NIH, Novartis, Sanofi, and UCB.

Evidence-based medicine is built on methods designed to test interventions, with the goal of providing sufficient evidence of value to guide regulatory approval, reimbursement and practice. The most definitive of these methods is the double blind, placebo-controlled randomized clinical trial (RCT). In an idealized setting<sup>1</sup> when both positive and negative trials are published equally, an RCT should eliminate the potential for bias and provide the closest we can get to definitive proof.

Increasingly, however, patients see rapid technological advances occurring outside of medicine and are frustrated that the pace of drug development is not on par. Today's "empowered patients" increasingly demand their data and seek methods that rapidly generate evidence derived from their experiences<sup>1,2</sup>. While there are obvious limitations<sup>3</sup>, many of these can be addressed with larger samples, clinical validation, and the use of objective measures to supplement patient-reported outcomes.

This parallel track of real world experience of patients living with disease can complement the traditional research and development processes. New institutions, such as the Patient-Centered Outcomes Research Institution are requiring that researchers include patient priorities and patient concerns in the development of new research processes. The Food and Drug Administration has proposed new methods for incorporating patient feedback into drug development processes and regulatory priorities, and novel Patient-Reported Outcome Measures are being considered for inclusion of post-market surveillance on the efficacy and safety of

drugs once they pass through their final regulatory hurdles and reach market approval.

Amyotrophic lateral sclerosis (ALS) is a rapidly fatal neurodegenerative disease<sup>4</sup>. Many patients with ALS (PALS) are desperate for effective treatments and have little tolerance for delays inherent in obtaining rigorous results from clinical trials. Self-experimentation with supplements or off-label drugs is common<sup>5</sup> and those diagnosed with the disease consume whatever information is available in order to determine whether there is a reasonable (or sometimes unreasonable<sup>6</sup>) way to slow the progression of their relentless disease.

This history of self-experimentation is so common in ALS that a group of over 80 ALS clinicians and researchers has been assembled by Dr. Richard Bedlack to form “ALS Untangled”<sup>7</sup>. The mission is to investigate non-traditional therapies from a number of different perspectives, suggested to them by patients. Claims of efficacy for treatments from cannabis<sup>8</sup> to coconut oil<sup>9</sup> have been examined from the perspective of basic science, animal models, published trials, and compared to individual reports from clinicians and patients on PatientsLikeMe. To date 18 investigations of treatments that have been completed and published in the field journal *Amyotrophic Lateral Sclerosis* (and at [www.alsuntangled.com](http://www.alsuntangled.com)) to allow patients to read them.

PatientsLikeMe is an online patient network that facilitates the sharing and collection of longitudinal data on members' treatments and outcomes using patient-reported outcome measures such as the ALS Functional Rating Scale (ALSFRS-R<sup>10</sup>) which agrees well with clinical impression<sup>11</sup>. In response to a RCT published in PNAS that suggested the drug lithium carbonate slowed the progression of ALS<sup>12</sup>, we used this platform to see if we could replicate the putative effects of the drug on ALS progression. We developed a novel model that compares patients who self-reported using the compound off-label with a population of similar patients who did not report use of the drug. In our preliminary analysis shared just nine months after the original study was published we were unable to replicate the finding<sup>13</sup>. We subsequently published our methods and our data set using a sample some ten times larger than the patient population in the original RCT<sup>14</sup>. These findings, that lithium carbonate did not impede the progression of ALS, were later replicated by four RCTs which either found no efficacy or failed due to futility<sup>15</sup>.

We had been interested in developing predictive models even before patients began experimenting with lithium. From a clinical perspective, giving patients a clearer sense of when they might meet certain disease milestones such as a wheelchair or feeding tube might help those patients who wanted to know such things to plan for the future.

Our population of over 6,000 patients with ALS shares a robust source of observations on which to evaluate the potential efficacy of interventions. We have a

self-imposed obligation to honor that contribution of our members by deriving value from their data, facilitating high quality research and sharing the results with our patients as promptly as possible. When it comes to decisions about treatments, PALS are making decisions on both relatively safe and potentially dangerous treatments with limited information, and in some cases without clinical guidance.

This raises some questions. Should our analysis of this data on unproven therapies be provided to the patients and clinicians treating them or should it remain in our servers until it reaches the significance or scientific rigor required for submission to a peer-reviewed journal? What questions should guide our decision about sharing our conclusions? Given that our predictive ALS model and its methods are new, and still contain unexamined biases, is it appropriate to attempt to share information on efficacy and risks? How does one evaluate the risk and benefit of sharing a preliminary non-definitive finding with the clinical community against the status quo where use of unsupported therapies is already widespread?

It is possible our predictive model could show a preliminary benefit and therefore encourage the use of an intervention that is ultimately proven ineffective, subjecting patients to additional risk. An example of this happening before PatientsLikeMe was even launched concerns the early enthusiasm for minocycline (obtained off-label by some patients when they couldn't get in to trials), which later turned out to have a worsening effect on ALS progression<sup>16</sup>. Perhaps more damning, we could inappropriately refute what would eventually be an efficacious therapy and

discourage future study. When we evaluated and disclosed initial findings on lithium carbonate, clinicians voiced this fair concern as well as their critique that we might hurt enrollment in traditional studies.

There is an additional dilemma: our tool is not limited to evaluation of off-label or alternative therapies. Many patients (N=150) currently participating in formal RCTs are also members of PatientsLikeMe, tracking the same ALSFRS-R outcomes online that they are being used in the trial. This means we can follow an ongoing phase II or phase III study and compare the entire trial group (without knowing who is on treatment or placebo) to their predicted outcome using our previously described matching algorithm<sup>14</sup>. Because our ALS model uses the full disease history of each patient prior to starting the trial and we have such a large control group, we could potentially have greater signal than some of the smaller studies have themselves. We are still subject to a range of biases, which are likely to include known issues such as selection bias, the placebo effect, the Hawthorne effect, and other unknown biases too. Individual patient data is shared openly on PatientsLikeMe [e.g. the KNS-760704 (dexpramipexole) phase III study <http://www.patientslikeme.com/treatments/show/21026>] and when trial patients identify certain side effects (e.g., neutropenia) this may suggest they are on the treatment, rather than placebo arm. Separately, patients are collating such data themselves and using it to decide whether to stay involved in the trial.

Real-time evaluation of efficacy by patients in a trial represents a direct challenge to the status quo. If it is possible for patients to know (and share) whether a study drug

is working or not then the potential bias this introduces to the study itself might invalidate it and jeopardize approval. One cannot imagine a more tragic result than delaying the approval of an effective therapy for ALS because a study was compromised. Such information about the outcome of a pivotal trial also has enormous financial implications. Large-scale ALS trials can cost hundreds of millions of dollars, and billions of dollars of investor capital may be at stake.

These are not just theoretical challenges. We are striving to make these decisions effectively as we grow our platform and add additional disease models.

PatientsLikeMe is also not the only system that will challenge our existing blinded research process and others will develop approaches<sup>17</sup> that will ask similar questions.

Although our tool has more sophisticated data collection and analysis, it is worth noting that even before we built our lithium tool, patients traded Excel files, shared blog posts, and uploaded their experience to shared Google Docs. The ALS community may be at the fore, but these vacua exist in other spaces burgeoning across the e-patient movement<sup>2,18</sup>.

To initiate what we hope will be a patient-centered and spirited debate, we present the following results from a mixed effects model of ALSFRS progression.

Table 1 shows the parameter estimates for base rate of decline, and change in that rate (effect size) for several potential treatments, along with confidence levels that

each treatment has a clinically meaningful effect. For the two clinical trials, the patients are blinded to which arm of the study they are in, so we do not know that information either. Therefore, we present results under two different scenarios:

- Assume that treatment effect is integrated over all patients in trial. This will “dilute” the treatment effect, as it is almost certainly true that not all PatientsLikeMe members are in the treatment arm. This is, however, the only effect we can measure without further assumptions.
- Assume that the balance of patients across treatment and placebo arms is the same for PatientsLikeMe members as it is for the trial as a whole. This likely leads to a closer estimate of the true treatment effect, but at the cost of making an unfounded assumption.

Table 2 show estimates of a model in which the in-trial group is decomposed into two subgroups: those who reported side effects and those who did not. The additional effect in NP001 may be a signal that those patients are on treatment (as patients claimed in our forums). The absence of an additional effect in dexpramipexole provides no such information beyond the grouped analysis. It is included for completeness only.

Finally, Figure 1 shows graphically the estimated probability ranges for the treatment effect sizes (based on the “equal balance” assumption stated above). The



curves peak at what we would estimate to be the most likely treatment effect, with a narrow peak indicating tighter confidence bounds on that estimate.

	Lithium carbonate	NP001	KNS-760704 (dexpramipexole)	Sodium chlorite
Total N of reported trial from ClinicalTrials.gov	Various	105	943	N/A
N of patients/matched controls meeting our data criteria on PatientsLikeMe	78 / 390	28 / 280	29 / 319	17 / 85
Duration of PatientsLikeMe observation window	12 months	6 months	12 months	2.5 months
Baseline rate of decline for all patients	0.91 (0.04)	0.93 (0.07)	0.96 (0.05)	0.87 (0.30)
Change in rate of decline for patients reporting intervention or study participation	-0.02 (0.10)	-0.21 (0.21)	-0.13 (0.16)	0.69 (0.67)
Placebo effect [fixed offset for any report <=2 months]	-0.45 (0.17)	-0.54 (0.17)	-0.44 (0.26)	-0.38 (0.29)
Model confidence that treatment slows progression to a clinically significant degree across all study participants [combination of treatment/placebo arms]	5%	55%	36%	12%
Estimated effect size if PatientsLikeMe has same balance between treatment/placebo arms as actual trial	-0.02 (0.10)	-0.32 (0.26)	-0.27 (0.23)	0.69 (0.67)
Model confidence that treatment slows progression to a clinically significant degree, for patients in treatment arm(s) using estimate from above	5%	69%	64%	12%

Table 1: ALS model estimates for four potential treatments. Negative change in rate of decline indicates slowing of progression. Units of progression rate estimates are ALSFRS-R points per month. ALSFRS-R = ALS Functional Rating Scale (Revised) <sup>10</sup>; SEM = Standard error of the mean; “clinically significant degree” = slowing of rate of ALSFRS-R decline by 20% or more<sup>19</sup>.

	NP001	KNS-760704 (dexpramipexole)
Number of patients reporting side effects	18	10
Change in rate of decline for all study participants	-0.13 (0.34)	-0.16 (0.20)
<i>Additional</i> change in rate for those reporting side effects	-0.13 (0.42)	0.07 (0.32)

Table 2: In-trial treatment effect broken out into two separate effects.

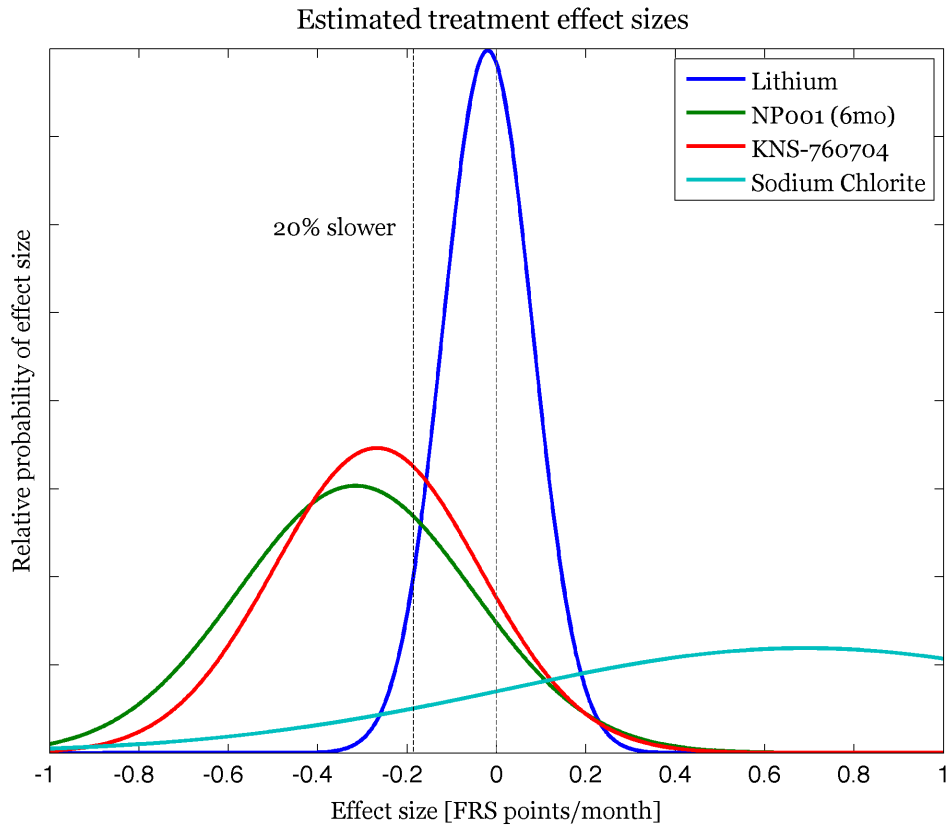


Figure 1: Graphical representation of parameter estimates of effect size for selected ALS treatments. Each line represents the probability distribution of the effect size; a high, narrow peak indicates that the effect size is more precisely estimated (generally because of larger sample size). More effective treatments will be centered toward the left. A measure of clinical effectiveness is indicated by the dashed line at about -0.2 points/month<sup>19</sup>. For purposes of these results, we assume that the balance between treated and placebo arms of the studies are the same for PatientsLikeMe members as they are for the trials as a whole.

**Discussion:**

Open data architectures, the Internet, education, empowerment of patients, and what appears to be a frustratingly slow pace of clinical innovation have all combined to bring us to where we are today. It is important to determine whether models such as ours can hasten this process or be deployed to investigate the “real world” efficacy of treatments that have already been approved. Or, as may be the case, whether such contributions merely add confusion to the mix.

We have provided data and analysis to help drive debate. ALS Untangled showed that, in the absence of other evidence, clinicians were willing to listen to patients’ experiences of side effects, efficacy, and tolerability to help dispel the notion that off-label supplements such as coconut oil might be effective. For the most part, patient reported data agrees with what clinicians would recommend for symptom management in ALS<sup>20</sup>. But what about when there is more rigorous evidence forthcoming, or the trials have already been conducted, what then? Do these data lose value? Does the publication of a positive pivotal RCT mean we can abdicate responsibility to learn anything new about a treatment from how it performs in the real world?

At the time of submission, neither the results of the NP001 phase II trial nor the KNS-760704 phase III trial have been released, and we have waited for both RCTs to conclude before sharing our data. We have conducted some unrelated research for Biogen, but we have had no insight or inside information about these studies from either sponsor, nor do we have a financial position in the outcome of these results.

While independently measuring and predicting the outcome of a trial could be legitimately critiqued for being disruptive or sensationalist, we also present alongside it data that resulted entirely from patient frustration with the trial system and could not have been analyzed any other way. Patients started taking sodium chlorite after searching through the patents of the sponsor of NP001, and began ingesting this industrial cleaner in the hopes that it was equivalent to NP001 and that the trial would work. Some were enrolled in the trial and then started using sodium chlorite when the study ended. As is sometimes the way with dynamic situations, the lay media have better coverage of the situation than the scientific literature (<http://abcnews.go.com/Health/Wellness/lou-gehrigs-disease-patients-research-hands/story?id=16150103#.UIIQMGmEuDo>). Our model suggests that sodium chlorite has a potentially negative effect; we have more than 80% confidence that it is worsening patients' progression rate. We hope our findings will help to serve patients and clinicians that are currently using this treatment in ALS and perhaps reduce harm. Like any new technology, this type of real world surveillance might sometimes do good and sometimes ill but we feel it is preferable to the absence of information about many putative treatments.

There are many situations where null hypothesis testing and  $p < 0.05$  has its uses, but we all recognize it is an arbitrary threshold and that many important decisions are taken within and outside of medicine and research on the basis of much lower levels of confidence. It will be important to balance quality against timeliness when using

emerging data, and having better language to describe our shared confidence in conclusions can only help clinical practice and patient understanding.

In conclusion we have chosen on this occasion not to wait for  $p < 0.05$  to share our work on how models and systems that provide the ability to measure benefit in real time could be used to advance knowledge and benefit patients. There are many issues still to cover including the interesting ethical dimensions such work raises and we hope to encourage enthusiastic dialogue.

### **Acknowledgements**

We dedicate this work to Rob Tison, whose screen name across several online communities was “Persevering”. Rob was an ALS patient and advocate who was the definition of the citizen scientist. He lobbied drug companies to let as many patients as possible take part in trials, was a participant in the NP001 study, and helped educate other patients in scientific methods and advocacy. We started this work with the aim of being able to tell him whether or not his efforts were in vain and in this we failed to meet our primary endpoint; sadly, Rob did not live to see the analysis we present here. We are grateful to all our patient members for sharing their data openly and for supporting our efforts, and to Lori Scanlon and David Clifford for assistance in preparing the manuscript.

## References

1. Rothwell, P. M. Treating Individuals 1: External Validity of randomised controlled trials: 'To whom do the results of this trial apply?'. *Lancet* (2005).
2. Swan, M. Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. *J. Med. Internet Res.* 14, e46 (2012).
3. Janssens, A. C. J. W. & Kraft, P. Research Conducted Using Data Obtained through Online Communities: Ethical Implications of Methodological Limitations. *PLoS Med* 9, e1001328 EP – (2012).
4. Wijesekera, L. C. & Leigh, P. N. Amyotrophic lateral sclerosis. *Orphanet journal of rare diseases* (2009).
5. Miller, R. G. et al. Outcomes research in amyotrophic lateral sclerosis: lessons learned from the amyotrophic lateral sclerosis clinical assessment, research, and education database. *Ann. Neurol.* 65 Suppl 1, S24–8 (2009).
6. Regenberg, A. C., Hutchinson, L. A. & Schanker, B. Medicine on the Fringe: Stem Cell-Based Interventions in Advance of Evidence. *Stem Cells* (2009).
7. Bedlack, R. & Hardiman, O. ALSUntangled (ALSU): a scientific approach to off-label treatment options for people with ALS using tweets and twitters. *Amyotroph Lateral Scler* 10, 129–130 (2009).
8. ALSUntangled Group ALSUntangled No. 16: cannabis. *Amyotroph Lateral Scler* 13, 400–404 (2012).
9. ALSUntangled Group ALSUntangled 15: coconut Oil. *Amyotroph Lateral Scler* 13, 328–330 (2012).

10. Cedarbaum, J. M., Stambler, N., Malta, E. & Fuller, C. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences* (1999).
11. Maier, A. et al. Online assessment of ALS functional rating scale compares well to in-clinic evaluation: a prospective trial. *Amyotroph Lateral Scler* 13, 210–216 (2012).
12. Fornai, F., Longone, P. & Cafaro, L. Lithium delays progression of amyotrophic lateral sclerosis. *Proc Natl Acad Sci USA*. 12;105(6):2052-7. (2008).
13. Wicks, P., Massagli, M., Frost, J., Macedo, H. & Felzer, K. A patient-led trial of lithium in ALS using the internet. (ALS: 2008).
14. Wicks, P., Vaughan, T. E., Massagli, M. P. & Heywood, J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat. Biotechnol.* 29, 411–414 (2011).
15. Armon, C. Is the lithium-for-ALS genie back in the bottle? *Neurology* 75(7):586-7 (2010).
16. Gordon, P. H. et al. Efficacy of minocycline in patients with amyotrophic lateral sclerosis: a phase III randomised trial. *Lancet Neurol* 6, 1045–1053 (2007).
17. Staa, T. P., Goldacre, B., Gulliford, M. & Cassell, J. Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ* 344:e55 (2012).
18. Swan, M., Hathaway, K., Hogg, C. & McCauley, R. Citizen science genomics as a model for crowdsourced preventive medicine research. *J Participat Med*



(2010).

19. Castrillo-Viguera, C., Grasso, D. L., Simpson, E., Shefner, J. & Cudkowicz, M. E. Clinical significance in the change of decline in ALSFRS-R. *Amyotroph Lateral Scler* 11, 178–180 (2010).
20. Nakamura, C., Bromberg, M. & Bhargava, S. Mining Online Social Network Data for Biomedical Research: A Comparison of Clinicians' and Patients' Perceptions About Amyotrophic Lateral Sclerosis. *J Med Internet Res* 14(3):e90 (2012).