

Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm

Paul Wicks, Timothy E Vaughan, Michael P Massagli & James Heywood

Patients with serious diseases may experiment with drugs that have not received regulatory approval. Online patient communities structured around quantitative outcome data have the potential to provide an observational environment to monitor such drug usage and its consequences. Here we describe an analysis of data reported on the website PatientsLikeMe by patients with amyotrophic lateral sclerosis (ALS) who experimented with lithium carbonate treatment. To reduce potential bias owing to lack of randomization, we developed an algorithm to match 149 treated patients to multiple controls (447 total) based on the progression of their disease course. At 12 months after treatment, we found no effect of lithium on disease progression. Although observational studies using unblinded data are not a substitute for double-blind randomized control trials, this study reached the same conclusion as subsequent randomized trials, suggesting that data reported by patients over the internet may be useful for accelerating clinical discovery and evaluating the effectiveness of drugs already in use.

Online communities such as PatientsLikeMe that provide robust methods for patients to record and share data may have the potential to be used to conduct observational studies to assess the effectiveness of treatments. Although observational studies inherently cannot meet the gold standard of randomized clinical trials, they provide an opportunity to collect possibly useful early-phase data by capturing patients' self-experimentation. Empowering observational studies of patients' self-experimentation carries some risks. Nevertheless, an increasing level of self-experimentation is already happening¹. In this context, it is possible that patient-reported outcome data collected over the Internet could be integrated into academic and/or industry-led cycles of product development and evaluation².

Approximately half of ALS patients take vitamins and unproven supplements³, whereas a smaller number go to extraordinary lengths to experiment with unproven treatments such as stem cell transplants in the developing world⁴. Recently, a consortium of 75 ALS physicians, scientists and experts (ALSUntangled.com) has been formed

to investigate the use of self-experimentation, complementary and alternative medicine, and off-label drug usage⁵. There are a number of benefits to systematically studying patients' self-experimentation. First, it is important to respect patients' autonomy and their decisions; helping them participate in systematic evaluations may increase scientific literacy. Second, there is an obligation to collect data on the safety of self-experimentation. Unproven treatments might have substantial safety concerns, and risks to patients may be increased without a way to report safety issues. Finally, there is the chance that something (i.e., off-label usage, a change in dosage, delivery route or combination with other treatments) might actually be shown to be efficacious, leading to further study.

ALS is a condition where both randomized trials and nonrandomized clinical studies have yet to provide an effective therapy. It is a cruel and rapidly fatal neurodegenerative disease causing progressive weakness and muscle atrophy; median survival from symptom onset is 2–5 years⁶. In 2008, a study described the potential efficacy of lithium carbonate to slow the progression of ALS in a small, single-blind trial of 16 treated patients and 28 controls⁷. Despite skepticism from the medical community^{8–10}, some ALS patients were enthusiastic about the treatment¹¹ and by their own initiative used an online spreadsheet to gather data. PatientsLikeMe built a lithium-specific data collection tool (see **Supplementary Fig. 1**) to capture information about the 348 ALS patients registered with the PatientsLikeMe website who began taking the drug off-label via their physician. To investigate whether the major effect of lithium carbonate reported in the original study was corroborated in these 348 patients, we undertook an observational analysis of self-reported outcomes. ALS disease progression is evaluated using the Revised ALS Functional Rating Scale (ALSFRS-R¹², henceforth referred to as FRS), which measures patient-reported functional impairment in domains such as speech, swallowing, walking, arm function and respiratory function. This metric is one of the standard outcome measures used in ALS clinical trials. In the absence of randomization, blinding or a placebo group, a technique was needed to overcome potential biases, such as the inherent self-selection of self-experimentation in an online sample, the placebo effect and attrition.

RESULTS Participants

As of the date that our data set was finalized (28 February 2010), there were 4,318 ALS patients on PatientsLikeMe, all of whom were invited

Research and Development, PatientsLikeMe Inc., Cambridge, Massachusetts, USA. Correspondence should be addressed to P.W. (pwicks@patientslikeme.com).

Received 2 April 2010; accepted 10 March 2011; published online 24 April 2011; doi:10.1038/nbt.1837

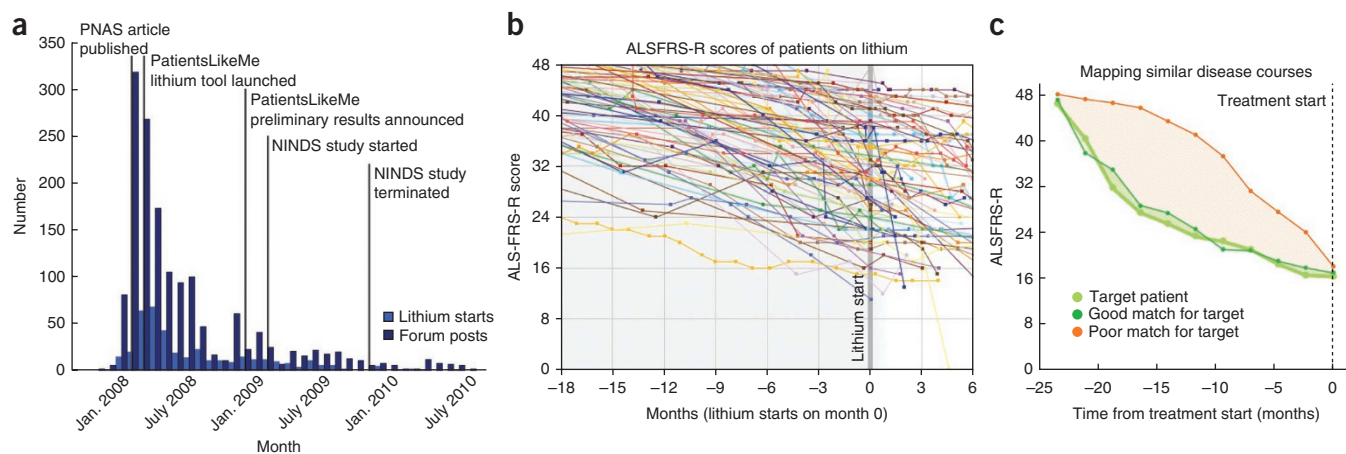


Figure 1 Conceptual overview of the online study environment and matching algorithm. **(a)** The number of patients choosing to experiment with lithium carbonate peaked in the months after publication of a small clinical trial in Italy. Preliminary negative results from this patient-led study were announced before the first randomized control trial had started recruitment. **(b)** Aggregate view of FRS scores for all 348 patients analyzed in this study. These data were publicly available online during the study. **(c)** Illustration of disease progression curves of control individuals that are good and poor matches for a particular patient. Both control individuals would be considered comparable by traditional matching criteria. The PatientsLikeMe matching algorithm minimizes the area between the disease progression curves for a target patient and a control.

to report their FRS scores, symptoms, treatments (with start and stop dates), site of ALS onset and demographic data using online tools provided on the PatientsLikeMe website¹³ (Fig. 1a,b and Supplementary Fig. 1). Of these, 3,674 (85%) provided at least basic demographic and diagnosis data; 348 of those (9%) reported taking lithium. After applying inclusion and exclusion criteria (see Online Methods), 149 patients remained eligible for subsequent analysis in an ‘intent to treat’ group (that is, they took lithium but may have discontinued within 12 months) and 78 patients were eligible in a ‘full course’ group (that is, a subset of the intent-to-treat group who continued to take lithium for the entire 12 months). De-identified data on all patients described in this study are provided in Supplementary Data File.

Matching algorithm

ALS trials have used a variety of methods to match patients receiving a treatment with appropriate control patients, with trial designs including futility design, multistage adaptive design, lead-in periods, selection design and historical controls¹⁴. Our design most closely resembles a combination of historical controls with a lead-in period; this has the advantage of being able to test drugs that may have the potential to have a very large clinical effect, but has the disadvantage that participants enrolling in an unblinded study may differ from historical controls or be biased in the data that they report¹⁴. To help minimize such biases, we developed an algorithm, the PatientsLikeMe matching algorithm, to match lithium-treated and control patients based on their entire disease progression, as measured by the FRS, before treatment was initiated (Figs. 1c and 2a). This technique is reliant upon having a large historical database of prospectively captured data over several years. In contrast, lead-in studies in ALS typically have only had brief (3–6 month) lead-in periods.

For each patient taking lithium, the algorithm matched multiple controls from our database that had as similar an FRS trajectory as possible to the treated patient’s disease trajectory from onset to start of lithium. To aid in matching the period soon after onset, we assigned all patients a score of 48 (the scale maximum) on the day of onset, unless they reported a lower score for that date (true for 33 treated patients and 8 controls reporting lower scores at onset, mean FRS: 46). The algorithm tends to match patients to controls who have similar

time since onset (otherwise the early trajectories diverge), study-start severity (otherwise recent trajectories diverge) or slope change, meaning mild decline followed by plummet or vice versa (otherwise middle of trajectories diverge). Treated and control patients were not required to be progressing contemporaneously along the disease course because we translated the controls’ progression backward or forward in time to obtain the optimal alignment with the treated patient. Full mathematical details of the algorithm are in Online Methods.

Analysis of treatment efficacy

We performed two analyses, an intent-to-treat analysis of 149 patients who reported taking lithium for at least 2 months (but may have discontinued taking the drug or died within 12 months of commencing treatment), and an analysis of the subset of 78 patients who stayed on lithium for a full 12 months or died within that period. For all treated patients, the PatientsLikeMe matching algorithm was used to select a control group matched on pretreatment FRS progression (Fig. 2a). Although other factors were not explicitly used to match, we did not observe significant group differences for age ($D_{149,447} = 0.07$, $P = 0.60$, treated: 51.3 years (s.d. = 10), control: 52.3 years (s.d. = 11)), site of onset ($\chi^2(2) = 5.5$, $P = 0.07$, treated: 46% arms/32% legs/21% bulbar, control: 36% arms/39% legs/26% bulbar) or FRS score at treatment start ($D_{149,447} = 0.04$, $P = 0.999$, treated FRS: 34.1 (s.d. = 7.9), control FRS: 34.1 (s.d. = 8.0)). However, the distribution by sex was significantly different across treatment and control groups, with males accounting for 72% of the treated and 59% of the control group ($\chi^2(1) = 7.7$, $P < 0.01$).

We did not observe a statistically significant difference in FRS score at 12 months ($D_{149,447} = 0.10$, $P = 0.22$) in the intent-to-treat group. All other monthly checkpoints were consistent with this result, with $P > 0.05$ at each checkpoint (Fig. 2b). Based on a Kaplan-Meier plot (Supplementary Fig. 2), we did not observe a significant difference in survival between treated patients and controls ($P = 0.93$). This analysis had 80% power to detect an absolute difference of 2.2 FRS points at 12 months, 22% of the average decline of the control group. This power is in line with clinician perceptions of a meaningful improvement in progression¹⁵. The observed absolute difference was 0.74 FRS points, in contrast to the previously reported difference of 8 FRS points at 12 months⁷.

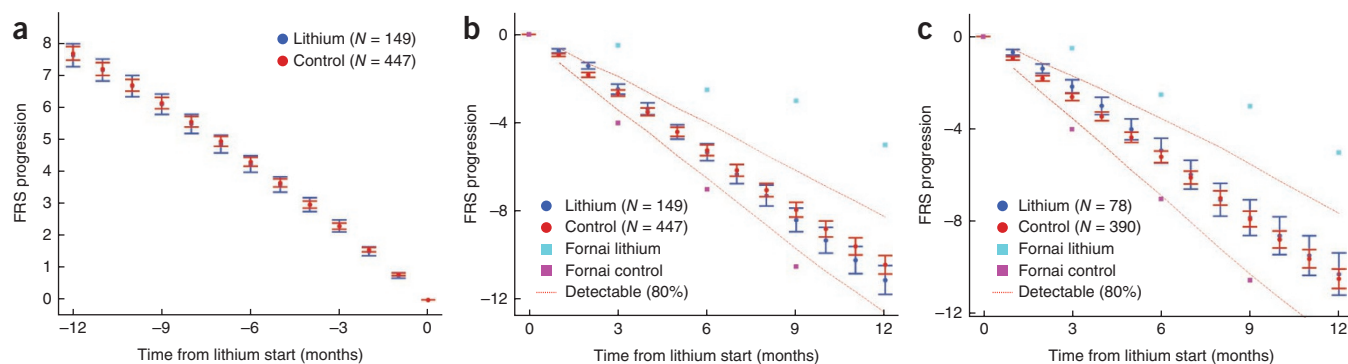


Figure 2 Results of analyses show no significant effect of lithium carbonate on rate of ALS progression. (a) Summary of pretreatment disease progression curves for 149 intent-to-treat patients matched by the PatientsLikeMe matching algorithm. Error bars are 1 s.e.m. in each direction. (b) Intent-to-treat analysis of 149 patients treated with lithium carbonate compared with controls fails to find any significant differences in progression ($P > 0.05$ at 12 months). Squares represent data from a previous trial⁷. Error bars are 1 s.e.m. in each direction. Dashed lines indicate the smallest detectable effect ($\alpha = 0.05$, 80% power). (c) Full-course analysis of 78 patients treated with lithium carbonate compared with controls fails to find any significant differences in progression ($P > 0.05$ at 12 months). Dashed lines as in b.

We also analyzed the subset of 78 patients who either completed the entire 12-month course of treatment or died within that period. We did not observe statistically significant differences in FRS scores between treated patients and matched controls at 12 months ($D_{78, 390} = 0.05$, $P = 0.42$) or any time point ($P > 0.05$) (Fig. 2c). This analysis had 80% power to detect an absolute difference of 2.8 FRS points at 12 months, 28% of the average decline of the control group.

Our conclusion is robust against varying the parameters of the study (listed in the “Inclusion and exclusion criteria” section of Online Methods), finding no choice of parameter sets that led to a statistically significant difference between our treated and control groups at the 12-month endpoint. However, it is still possible that parameters not included in our analyses or unknown to us could confound these results.

DISCUSSION

Using our analytic approach and data-capture methodology, we have been unable to replicate previously described promising results⁷ of the efficacy of lithium to slow the progression of ALS. Randomized clinical trials in the United States¹⁶ and Europe¹⁷, have been halted early for futility owing to lack of detected efficacy and safety concerns, which may have missed a small but real effect such as that found for riluzole¹⁸. In the UK and the Netherlands, two studies designed to detect smaller effects than the halted trials are ongoing, and will continue until the planned study end at 18 months and 2 years, respectively; results are still pending¹⁹.

Positive results from phase 1 and phase 2 trials can lead to changes in patient behavior, particularly when a drug is readily available. For instance, ALS patients also rushed to use the widely available broad-spectrum antibiotic minocycline when animal model results showed it delayed progression^{20,21} and phase 1 and 2 studies showed it to be safe²². However, minocycline was subsequently found to accelerate disease progression when tested in a multicenter randomized placebo-controlled phase 3 trial²³. The ongoing availability of a surveillance mechanism such as ours might help provide evidence to support or refute self-experimentation. Indeed, during this study it was reported that clinicians were citing preliminary results available on the PatientsLikeMe website as a way of dissuading self-experimentation¹. Had our findings been positive, further randomized trials would still have been necessary to replicate the finding and establish dosing, safety, side effects and combination therapy effects.

There are several potential advantages of collecting patient-reported outcome data online. The first is speed. It took only 9 months from initiation of the tool (March 2008) to the first public sharing of preliminary results (December 2008)¹⁹. The second is patient access. There is a potential to rapidly recruit widely dispersed patients with rare conditions and to overcome selection bias favoring patients living near specialist centers. The third is availability of control participants. Clinical outcome data were passively collected from thousands of patients who served as potential matched controls. The fourth is cost. Online studies have lower marginal costs per patient as compared with thousands of dollars per patient in traditional trials. The fifth is patient engagement. Patients who submitted data using the PatientsLikeMe website were connected with other patients, which may have a range of benefits¹³.

There are several limitations inherent in using self-reported data and historical controls¹⁴. Unlike randomized trials, which match the comparison groups on all possible confounding factors, subject only to chance variation, observational studies cannot control for unmeasured covariates. The FRS is typically a clinician-led interview; nevertheless, it has been validated for self-reporting and found to have good reliability²⁴. In addition, patients were not ‘blinded’ as to whether they were taking lithium. Patients who decided to take lithium may have been overly optimistic in their self-assessments, which could have led to a placebo effect; nevertheless, we did not observe a statistically significant difference in the current study, which might have been attributable to the placebo effect had it arisen. There may be biases in the psychological makeup of patients who managed to persuade their physicians to prescribe them lithium; anecdotally, patients reported switching physicians until they found one who would prescribe the drug for them. Therefore these patients may be biased toward a high degree of health literacy and have other health behavior attributes that we are unable to evaluate. Compared to the typical sex ratio of the ALS population⁶, there was a higher proportion of men taking lithium, perhaps reflective of a greater propensity for men to take an experimental treatment. This group may not compare well with patients in other trials of lithium.

As patients decided for themselves to take lithium, our treatment cohort was not selected using the same set of inclusion and exclusion criteria typically used in ALS trials, because we were observing a naturally occurring phenomena rather than assigning treatment arms. These criteria typically exclude patients with young-onset ALS, disease duration > 2 years, forced vital capacity (FVC) $< 50\%$ or familial ALS. However,

our sample may be considered more representative of the broader population. For instance, the use of El Escorial Criteria as inclusion criteria for standard clinical trials may exclude as many as 40% of patients²⁵.

Our methodology does not permit us to know with full confidence whether any given patient may have died during the study window; patient survival is thought to be the optimal outcome measure for clinical trials in ALS²⁶. Moreover, at the time of study, the PatientsLikeMe website did not provide a mechanism for patients to electronically report serious adverse events to manufacturers and regulators. This feature is now available and means that future studies of this nature could include reporting of adverse events to either a product manufacturer or to a regulator such as the Food and Drug Administration. Finally, there is a possibility that incorrect or falsified data may be entered on the PatientsLikeMe website. Although not definitive proof, several lines of evidence suggest that much of the data are valid. For instance, the name of the diagnosing physician is provided for 76% of controls and 69% of lithium takers, though we were unable to independently validate that these physicians saw these patients. Also, the prospective data would be extremely time consuming to falsify. Lastly, there is no direct benefit to participants, financial or otherwise, to contributing data to the platform, and therefore little incentive to falsify data.

In regards to the broader applicability of our approach, ALS may be a somewhat special case in that disease progression is relatively predictable, there are no effective treatments and patients are highly motivated to submit data. However, the same is true for many other rare and life-changing illnesses in need of effective treatments²⁷. Preliminary attempts to replicate the matching techniques used in this study with multiple sclerosis patients suggest some benefit from using a matching algorithm to increase the accuracy of predictions in this episodic, treatable and slower-to-progress condition, although the benefits have been less substantial than for ALS. In testing and evaluating the validity of research conducted in a self-reported, online data environment, there is a rich opportunity for future work. Attempting to establish the efficacy of a treatment in a prospective manner inevitably draws comparisons with methodologies that have the highest standards of rigor in medicine, and by comparison this discipline is in its infancy. Other applications of internet-based observational studies might include measuring disease variability or disease burden or identifying unmet needs for treatment strategies for other patients with life-changing illnesses seeking to improve their outcomes.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nbt/index.html>.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We are grateful to K. Felzer and H. Macedo (deceased) for their initial work in capturing patient-reported ALSFRS-R data on ALS patients. N. Daniels assisted in earlier designs of the study.

AUTHOR CONTRIBUTIONS

P.W. designed the study, oversaw the project and drafted and revised the manuscript. T.E.V. developed the matching algorithm, analyzed data and revised the manuscript. M.P.M. designed the lithium data capture tool, developed statistical methods and revised the manuscript. J.H. designed the study, designed the ALS web site, developed the matching algorithm and revised the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/nbt/index.html>.

Published online at <http://www.nature.com/nbt/index.html>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Ross, E. Unapproved drug use: compassionate or cause for concern? *Lancet Neurol.* **8**, 136–137 (2009).
- Boguski, M.S., Mandl, K.D. & Sukhatme, V.P. Drug discovery. Repurposing with a difference. *Science* **324**, 1394–1395 (2009).
- 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–S28 (2009).
- Regenberg, A.C., Hutchinson, L.A., Schanker, B. & Mathews, D.J. Medicine on the fringe: stem cell-based interventions in advance of evidence. *Stem Cells* **27**, 2312–2319 (2009).
- ALSUntangled Group. ALSUntangled No. 8: Low dose naltrexone for ALS. *Amyotroph. Lateral Scler.* **12**, 76–78 (2011).
- Wijesekera, L.C. & Leigh, P.N. Amyotrophic lateral sclerosis. *Orphanet J. Rare Dis.* **4**, 3 (2009).
- Fornai, F. *et al.* Lithium delays progression of amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. USA* **105**, 2052–2057 (2008).
- Figlewicz, D.A. Comments on Fornai *et al.* (PNAS/ Feb 2008). *Amyotroph. Lateral Scler.* **9**, 125–126 (2008).
- Meininger, V., Shefner, J. & Cudkowicz, M. Lithium therapy in ALS. *Amyotroph. Lateral Scler.* **9**, 122 (2008).
- Bedlack, R.S., Maragakis, N. & Heiman-Patterson, T. Lithium may slow progression of amyotrophic lateral sclerosis, but further study is needed. *Proc. Natl. Acad. Sci. USA* **105**, E17 (2008).
- Frost, J.H., Massagli, M.P., Wicks, P. & Heywood, J. How the Social Web Supports patient experimentation with a new therapy: the demand for patient-controlled and patient-centered informatics. *AMIA Annu. Symp. Proc.* **6**, 217–221 (2008).
- Cedarbaum, J.M. *et al.* The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). *J. Neurol. Sci.* **169**, 13–21 (1999).
- Wicks, P. *et al.* Sharing health data for better outcomes on PatientsLikeMe. *J. Med. Internet Res.* **12**, e19 (2010).
- Cudkowicz, M.E. *et al.* Toward more efficient clinical trials for amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.* **11**, 259–265 (2010).
- 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).
- Aggarwal, S.P. *et al.* Safety and efficacy of lithium in combination with riluzole for treatment of amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* **9**, 481–488 (2010).
- Chio, A. *et al.* Lithium carbonate in amyotrophic lateral sclerosis: lack of efficacy in a dose-finding trial. *Neurology* **75**, 619–625 (2010).
- Verstraete, E., Veldink, J.H. & van den Berg, L.H. Would riluzole be efficacious in the new ALS trial design? *Lancet Neurol.* **9**, 949–950, author reply 950–951 (2010).
- Armon, C. Is the lithium-for-ALS genie back in the bottle? *Neurology* **75**, 586–587 (2010).
- Kriz, J., Nguyen, M.D. & Julien, J.P. Minocycline slows disease progression in a mouse model of amyotrophic lateral sclerosis. *Neurobiol. Dis.* **10**, 268–278 (2002).
- Van Den Bosch, L., Tilkin, P., Lemmens, G. & Robberecht, W. Minocycline delays disease onset and mortality in a transgenic model of ALS. *Neuroreport* **13**, 1067 (2002).
- Gordon, P.H. *et al.* Placebo-controlled phase I/II studies of minocycline in amyotrophic lateral sclerosis. *Neurology* **62**, 1845 (2004).
- 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).
- Montes, J. *et al.* Development and evaluation of a self-administered version of the ALSFRS-R. *Neurology* **67**, 1294–1296 (2006).
- Logroscino, G. *et al.* Descriptive epidemiology of amyotrophic lateral sclerosis: new evidence and unsolved issues. *J. Neurol. Neurosurg. Psychiatry* **79**, 6–11 (2008).
- Gordon, P.H. *et al.* Defining survival as an outcome measure in amyotrophic lateral sclerosis. *Arch. Neurol.* **66**, 758–761 (2009).
- Call, J., Scherzer, N.J., Joseph, P.D. & Walentas, C. Evaluation of self-reported progression and correlation of imatinib dose to survival in patients with metastatic gastrointestinal stromal tumors: an open cohort study. *J. Gastrointest. Cancer.* **41**, 60–70 (2010).

ONLINE METHODS

Statistical techniques. Statistical analysis was conducted using MATLAB version R2009b. Categorical differences were tested with χ^2 tests of homogeneity. To test for differences in age and FRS scores, we applied both Welch's *t* test and the Kolmogorov-Smirnov test (two sided, two sample); only K-S results are reported, but Welch's *t* test had the same findings. In the event of a significant finding, Bonferonni-corrected comparisons of progression rates month by month would be applied to control for multiple comparisons. Power calculations used MATLAB "sampsizepwr" function. Minimum detectable differences assumed 80% power and $\alpha = 0.05$. The Welch-Satterthwaite approximation was used to estimate the degrees of freedom. Because these power calculations assume normality, we also estimated power by means of simulation; power presented in the paper is conservative (that is, detectable effect sizes were even smaller when simulated). Kaplan-Meier curves were compared using a log-rank test with Yates correction.

Inclusion and exclusion criteria. Patients were required to meet certain requirements to be included in the study. To facilitate this, a "study start date" was defined for each treated and control patient. For treated patients, this was the date they began taking lithium. For control patients, we randomly selected one of the dates for which they reported an FRS score. Once that date was established, all patients were required to meet the following inclusion criteria:

1. Reported diagnosis of ALS, but not primary lateral sclerosis or progressive muscular atrophy.
2. Reported date of birth, date of first ALS symptom and site of symptom onset.
3. At least 12 months have passed since the patient's study start date, such that they had had sufficient 'data opportunity' (time from their study start date to the finalization of our data set) to reach the study endpoint. (Note that requiring data, rather than just data opportunity, could introduce a bias toward longer-lived patients.)
4. Reported at least one FRS point within 65 d of study start date (that is, within the 'start window').
5. Reported at least one FRS point before, and one after, the start window.
6. Had an average rate of decline since first symptom within the range of 0.1–2.0 FRS points per month (reflecting typical trial design to ensure patients had reported some progression, but not overly severe).
7. Had their first symptom within 6–60 months of the start of the trial (reflecting typical trial design).
8. For those who took lithium, they must have reported taking it for at least 2 months (intent to treat) or 12 months (full course of treatment) or have died (as reported by their caregiver) while on lithium.

Our results were robust against variation in each of these parameters.

In addition to the FRS scores reported by each patient, a score of 48 (maximum of scale) was assigned on the patient's reported date of first symptom, unless they reported a lower score for that date (true for 33 treated patients and 8 controls reporting lower scores at onset; mean FRS, 46). Linear interpolation was used from known data points to create monthly checkpoints.

Treatment and/or control matching. As is common to observational studies in general, we lacked the key feature of randomized control trials (RCTs) that makes them the gold standard of clinical trials: the ability to randomize and/or balance patients into the treated and control groups. In this way, RCTs minimize factors (both known and unknown) other than the treatment itself that would cause the outcomes of the two groups to diverge.

We found that the factors thought to affect progression were balanced in our data set, or where they differed we performed separate analyses for the subgroups, demonstrating no association of these additional factors with outcome.

However, given our inability to randomize patients into the treatment group, we could not eliminate unknown factors that might cause the post-treatment progression of the two groups to diverge. Our approach to this issue was to match participants on the pretreatment disease progression as the best proxy to 'balancing' these potential confounding factors. We describe alternate approaches and our algorithm below.

"PatientsLikeMe Matching" algorithm. We found that using random treatment-control matching with one control often produced a noticeable systematic bias, with lithium patients having declined as much as 1 FRS point more than controls (on average) for the 12 months preceding treatment start.

Our initial method to match controls with treated patients was to minimize the difference in FRS at study start date, and difference in time since onset. We realized that we could match treated and control more effectively by comparing them along their entire FRS profiles before the study start date.

Mathematical description of algorithm. We developed an algorithm to minimize the area between the FRS progression curves of patients and controls over the entire course of the disease (before lithium start). The area is illustrated in **Figure 1c**.

1. Define t_0 as the lithium start date of the patient who took lithium.
2. Determine the patient's FRS at twice-monthly intervals t_i before their study start date (linearly interpolating if necessary), back to time of onset. These are FRS_i .
3. For each control, define their t_0 as one of their reported FRS dates, and determine FRS_i as above.
4. For each patient-control pair, calculate the area:

$$i.\text{Area} = \sum_i \text{abs}(FRS_i^{\text{treated}} - FRS_i^{\text{control}}) \times (t_i - t_{i-1})$$

5. For each patient, choose the control that minimizes this area.

The matching was done sequentially by patient, rather than searching for a global optimum for all patients. A given control patient would not be matched with more than one treated patient. We tested whether matching the patients in a different ordering affected the results; there was no significant impact.

This algorithm led to improved pretreated matching (**Fig. 2a**).

The improved matching is expected to have a concomitant reduction in the bias due to propensity to use lithium based on prior disease course. Because of our very large control pool ($N = 637$), we were able to match multiple controls to each treated patient. When multiple controls were used per patient, it was possible to reduce bias even further, by using the second and later controls to offset any bias introduced by the first control. For the intent-to-treat patient pool, the bias was minimized for three controls per patient. For the full-treatment group, the bias was minimized with five controls per patient. In any given study, we always chose to match the same number of controls per treated patient, to avoid the possibility that, for example, a "rapid decliner" had more controls than a "slow decliner," which would introduce a bias into the outcome.

Accounting for missing data. Given the high dropout rate relative to a clinical trial, it is important to consider the management of missing data. There is no consensus on the best method to analyze informative dropout, but it is agreed that it is important to test the sensitivity of the results to the dropout assumption.

We tested four assumptions on data dropout:

1. Include only patients who provided complete data.
2. Include all patients, but analyze only provided data.
3. Include all patients, and set FRS to 0 (that is, presume worst outcome) after last provided point.
4. Include all patients, and extrapolate each patient's FRS beyond last provided point, assuming linear progression along that patient's previous average slope since the study start date.

Also, there are two fundamentally different reasons we could be missing data at 12 months from a patient: the patient may have died (which may or may not have been reported to us), or the patient could be alive but simply stopped reporting data to the site. Therefore, we also tested different assumptions based on whether the patient's death was reported or not.

Whatever method is selected will, of course, affect the resulting statistics of the two groups. However, all of these choices lead to the same conclusion, which is that there is no statistically significant difference between the average FRS of the treated and control groups. For the results presented

in this paper, we have used the extrapolation method, believing that this most closely approximates the true course of the patient (relative to the other methods).

Alternative methods of analysis. One alternative method of analysis is to use a linear mixed-effects model to test whether the slope of the ALSFRS was different in the two groups²⁸. The impact of matching could also be accommodated in such an analysis. A deidentified copy of the data set is provided as a **Supplementary Data File** to allow readers to make their own analysis of the data.

Analysis by sex. Because of significant differences in the proportion of male and female between the groups, we performed our analysis separately on women and men, and found no difference between treated and control for either subset at 12 months post-treatment (males only: $N = 107$, $D_{107,321} = 0.10$, $P = 0.42$; females only: $N = 42$, $D_{42,126} = 0.18$, $P = 0.22$).

Analysis by site of onset. Because there was a trend approaching significance for a difference in site of onset, with lower representation of bulbar-onset

patients in the treated group than in the control group, we performed our analysis on just the subset of bulbar-onset patients, and the results were consistent with the larger group at 12 months post-treatment ($N = 32$, $D_{32,96} = 0.20$, $P = 0.27$).

Analysis by riluzole (Rilutek) intake. The patients in a previous study⁷ were all receiving riluzole. When our sample was restricted only to those patients on riluzole, there continued to be no significant difference between treated and control at 12 months ($N = 95$, $D_{95,190} = 0.12$, $P = 0.29$).

Analysis by lithium carbonate blood level. When we replicated a previous study⁷ as closely as possible, requiring patients to both be on riluzole and report a blood level between 0.4 and 0.8 mmol/l, there continued to be no significant difference at 12 months ($N = 28$, $D_{28,56} = 0.13$, $P = 0.91$).

28. Lindstrom, M.J. & Bates, D.M. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Am. Stat. Assoc.* **83**, 1014–1022 (1988).

