Defining Survival as an Outcome Measure in Amyotrophic Lateral Sclerosis

Paul H. Gordon, MD; Philippe Corcia, MD, PhD; Lucette Lacomblez, MD; Ksenia Pochigaeva, MD; Jean-Louis Abitbol, MD; Merit Cudkowicz, MD; P. Nigel Leigh, FMedSci; Vincent Meininger, MD, PhD

**Objectives:** To examine how respiratory interventions affect survival as an outcome measure and to define survival rate for trials in amyotrophic lateral sclerosis.

**Design and Setting:** We reviewed the data of 3 phase 3 clinical trials and examined differences in times to death, tracheostomy, and permanent assisted ventilation. We assessed the outcomes with χ² and Fisher exact tests for categorical variables and unpaired, 2-tailed t tests for continuous variables. We used Kaplan-Meier methods to estimate the differences in survival times between interventions. A power analysis generated sample size estimates for different end points.

**Patients:** In all, 2077 patients in 2 phase 3 trials of xaliproden and 400 patients in a phase 3 trial of pentoxifylline.

**Main Outcome Measures:** Death or combined death, tracheostomy, or permanent assisted ventilation.

**Results:** Of 743 deaths, 611 (82.0%) were owing to respiratory failure and 134 (18.0%) to other causes. The use of respiratory interventions across centers ranged from 0% to 6.6% (P=.001) of patients for tracheostomy and 11.1% to 23.1% (P=.05) of patients for noninvasive ventilation. Twelve of 55 patients (21.8%) undergoing tracheostomy had a vital capacity of 50% or more. Mean (SD) survival time was 457.9 (3.1) days using a combined end point and 467.2 (2.9) days with death alone as the outcome (P=.02). An estimated sample size to detect a 10% difference at 18 months between groups was 490 patients per arm for the combined end point and 410 patients for death alone.

**Conclusions:** Tracheostomy and permanent assisted ventilation are not equivalent to death in amyotrophic lateral sclerosis. The use of respiratory interventions differs between centers, leading to variability in combined outcome assessments. The time to the end point can differ significantly depending on its definition, and combining outcomes does not reduce the estimated sample size of a trial. The death rate alone is the least variable and most easily identifiable measure of survival rate in amyotrophic lateral sclerosis.

Arch Neurol. 2009;66(6):758-761

**Author Affiliations:**
- Department of Neurology, Columbia University, New York, New York (Dr Gordon);
- Fédération des Maladies du Système Nerveux (Drs Gordon, Corcia, Lacomblez, Pochigaeva, and Meininger) and Department of Pharmacology (Dr Lacomblez), Assistance Publique–Hôpitaux de Paris, Hôpital de la Salpêtrière, Paris, France; Center SLA, Department of Neurology, CHRU, Tours, France (Dr Corcia); Trophos S. A., Marseille, France (Dr Abitbol);
- Department of Neurology, Harvard Medical School, Boston, Massachusetts (Dr Cudkowicz); and Department of Clinical Neuroscience, King's College London, and Medical Research Centre for Neurodegeneration Research, Institute of Psychiatry, London, England (Dr Leigh).

**METHODS**

We analyzed the occurrence of death and different respiratory interventions, where the data were available, from 3 phase 3 randomized controlled trials (eTable; http://www.archneurol.com). We examined the data of 2 phase 3 trials of xaliproden; study 1 enrolled 867 patients at 40 centers in 5 countries, and study 2 enrolled 1210 patients at 39 centers in 8 countries. In each trial, participants were randomized to 1 or 2 mg of xaliproden or to placebo; in study 2, participants also received riluzole. There were no sig-
significant treatment effects in terms of time to death or combined death, tracheostomy, or PAV (combined end point) in either trial. We also examined the database of a trial of pentoxifylline conducted in 2004 in 12 centers in 4 European countries. Four hundred patients were enrolled, and the primary outcome measure was survival, regardless of the use of ventilatory support. At the end of the trial, 103 participants were alive in the treatment group was survival, regardless of the use of ventilatory support. At the end of the trial, 103 participants were alive in the treatment group and 120 participants were alive in the placebo group (P = .053 in unadjusted analyses).

For this study, the timing of interventions and their relation to vital capacity (VC) were analyzed using descriptive statistics, χ² and Fisher exact tests for categorical variables, and unpaired, 2-tailed t tests for continuous variables. We used Kaplan-Meier methods to estimate the differences in survival time between interventions with commercially available statistical software (SPSS, version 10; SPSS Inc, Chicago, Illinois). A power analysis–generated sample size estimates by commercial statistical software (SPSS, version 10). A power analysis–generated sample size estimates by commercial statistical software (SPSS, version 10).

Kaplan-Meier methods to estimate the differences in survival time between interventions with commercially available statistical software (SPSS, version 10). A power analysis–generated sample size estimates by commercial statistical software (SPSS, version 10). A power analysis–generated sample size estimates by commercial statistical software (SPSS, version 10).

CAUSES OF DEATH AND USE OF RESPIRATORY INTERVENTIONS

In the xaliproden trials, 745 patients died; approximately 82% of deaths were owing to respiratory failure, and 18.0% of the patients died of causes other than respiratory insufficiency, most commonly sudden death of unknown cause. Tracheostomy was administered to 93 of 2077 participants (4.5%), of whom 29 died (31.2%; mean time to death, 126 [124] days). Forty-one patients (2.0%) used PAV, defined as longer than 23 hours per day of NIV; 24 of these died (58.5%; mean time to death, 363.8 [129] days). The cause of death in most patients with tracheostomy or PAV was unknown. The VC was not significantly different at the time of intervention for patients who received tracheostomy or PAV (43.4% [21.9%] vs 37.2% [18.4%]; P = .12 for the t test).

In the pentoxifylline trial, 177 participants died, but the causes of death were not available. Thirteen of 400 patients underwent a tracheostomy (3.3%); of these, 6 died (46.2%; mean time to death, 128.2 [186.4] days). Seventy-seven patients used NIV (19.3%); of these, 42 died (54.5%; mean time to death, 142.5 [131.7] days).

VARIABILITY IN ADMINISTRATION OF INTERVENTIONS

Among the 2077 patients enrolled in the xaliproden trials, the percentage of patients in different countries who underwent tracheostomy varied from 0% to 6.6% (P = .001) (Table 1) and from a mean of 2.4% in Europe as a whole to 6.6% in the United States (P < .001). Variability also occurred across centers in the same country; the proportion of patients undergoing tracheostomy in Germany, for example, ranged from 3.2% to 5.3%.

In the xaliproden trials, the value of the last VC before tracheostomy could be analyzed in 50 of 93 patients (ie, the VC was performed during the 30 days before the tracheostomy) (Table 2). For 14% of the patients, the VC was 60% or more of the predicted value, and for 10% of the patients, the VC was 70% or more of the predicted value. These proportions were not equivalent in Europe and the United States, but in both regions approximately 22% of the patients who underwent a tracheostomy had a VC of 50% or more of the predicted value.

In the pentoxifylline trial, variability in the administration of tracheostomy between countries occurred with a range similar to that of the xaliproden trials (Table 1). According to data available only in this trial, variability among countries was also observed for NIV; the proportion of participants who used NIV was different between countries (P = .05; Table 3). Information is unavailable about the VC value at the time of initiation of NIV or tracheostomy.
IMPACT OF RESPIRATORY INTERVENTIONS ON TRIAL DESIGN AND DEFINITION OF END POINT

Considering the overall population of the 2 xaliproden trials, the rate of death was 39.5% (mean survival time, 457.9 days) at 18 months as defined by the combined end point and 35.1% (mean survival time, 467.2 days) with death alone as the outcome (P = .02, log-rank test; Figure). In the pentoxifylline trial (using different inclusion criteria), the rate of death was 46.0% at 18 months using the combined end point and 44.3% using death alone as the outcome (P = .69). According to the data from all 3 trials, the estimated sample size is not lower for the combined end point (Table 4; α = .05; β = .90).

The primary outcome measure of a trial is defined during the design phase and used to calculate the sample size to ensure that the study has sufficient power. The primary end point should be clinically relevant, valid, reliable, and sensitive to change; the least variable outcomes are the most sensitive to change. Functional end points have been used in place of survival rate in some trials in the hope that they could improve efficiency by acting as a surrogate for survival or better reflect quality of life. However, changes in functional measures, such as the ALS Functional Rating Scale, are not always predictive of survival, may show nonlinear decline that renders them less reliable for phase 3 trials, and do not correlate strongly with quality of life.

Survival analyses were used for the first time in the initial trials of riluzole. In both trials, the definition of the survival rate included death from any cause, tracheostomy, and intubation with artificial ventilation leading to tracheostomy. The reason for assimilating invasive ventilation with death was the idea that, at the terminal stage of the disease, respiratory failure could lead to death or tracheostomy (with or without a previous intubation). Respiratory failure was considered the unique reason for death in patients with ALS.

However, accumulating evidence began to show that death is not always related directly to respiratory muscle dysfunction, that ventilation is not always performed for severe dysfunction in respiratory muscles, and that ventilation does not always prevent death. In the xaliproden trials, among the 93 participants with tracheostomy, 31.2% died, and among the 41 participants with PAV, 58.5% died. The cause of death in these patients was not related to the VC value when the participant had the tracheostomy or began PAV. The cause of death in most patients undergoing ventilation is not clear. Many patients who undergo PAV die when they decide to stop the ventilation, but patients with ALS also die of cardiac infarction, pulmonary embolism, or other events. Overall, approximately 18% of the participants died of causes other than a clear respiratory insufficiency, usually termed “sudden death.” These data indicate that death is not necessarily related to the function of the respiratory muscles and that the time to administration of respiratory life support cannot be assumed to be equivalent to the time to death.

The introduction of NIV as a therapeutic intervention in ALS led to the inclusion of time undergoing ventilation as equivalent to death. The primary outcome measure in the xaliproden trials was the time to death, tracheostomy, or PAV, whichever occurred first. Permanent assisted ventilation was taken as the date at which the patient reported using noninvasive positive pressure ventilation for more than 23 hours per day. A similar definition has been used in US-based trials. However, incorporating tracheostomy and PAV into the survival outcome appears to have introduced variability in outcome measurement through several mechanisms. First, it was difficult to record the time of NIV, which was dependent on the quality of patient logs. These data were often poor or missing, and the time to PAV was less precise than the time to death or tracheostomy. Second,
the lack of standardized criteria for tracheostomy, as demonstrated by the widely variable VC value at the time of tracheostomy, led to significant variability in outcome assessments between centers. The variability among countries was also observed for NIV, although clinical guidelines exist (at least in the United States), reflecting differences in practice among investigators. This variability has also been seen in US-based trials and in epidemiological studies outside trials. In France, where there are guidelines for initiating NIV, the percentage of patients per center who undergo NIV ranges from 8.5% to 35% (V.M., unpublished data, July 2007).

The current lack of consensus for the timing of tracheostomy and noninvasive interventions has prevented the imposition of criteria for their use in international clinical trials. The reason for the absence of widely accepted guidelines is 2-fold. First, philosophies vary about end-of-life care in ALS and, second, there is a lack of data guiding best practice. Investigators have not accepted, for example, imposed criteria for applying tracheostomy in a trial because the use of invasive ventilation in clinical care varies. The incorporation of tracheostomy into a combined outcome measure was considered acceptable because, although its use might differ between centers, investigators would apply their approach equally to participants in the active and placebo groups in a blinded randomized trial. When the results of ongoing research examining the timing of various noninvasive therapies become available, existing guidelines will surely be modified and new ones will appear. We recommend that these guidelines be applied to the design of future trials, but it is unlikely that accepted criteria will be issued for the use of invasive ventilation in ALS.

The last consideration is the impact of introducing tracheostomy or PAV as a primary end point equivalent to death with regard to estimated sample size and trial results. Theoretically, the definition of death could affect the number of patients to be enrolled. If PAV and tracheostomy are included in the definition of survival, they might increase the number of events and therefore decrease the required sample size of a trial. However, our data show that not only does the combined outcome lead to an increase in endpoint variability, it also fails to meaningfully reduce the number of patients enrolled. Furthermore, our data show significant differences in the time to outcome, depending on how survival is defined. The time to the combined end point or to death alone differed in the xaliproden trials, suggesting that the results of a trial could depend on how the primary survival end point is defined. Although the disparity between survival end points is not large, it is significant, and the use of the most reliable measure will improve the validity and sensitivity of trial designs.

In summary, our data show that death and respiratory interventions are not equivalent in ALS, that there is variability in the use of respiratory interventions in clinical trials, and that, although there is no saving in the required sample size of a trial by using respiratory interventions with survival, the definition of the survival outcome could affect trial results. For these reasons, we recommend a return to the use of death rate alone—the simplest and most robust measure of survival rate—as the primary outcome measure for phase 3 trials. We suggest that various respiratory interventions be analyzed as covariates in secondary survival analyses and that standardized approaches to the timing of respiratory interventions be defined in a trial a priori according to current clinical standards of care.

Accepted for Publication: October 7, 2008.

Correspondence: Paul H. Gordon, MD, Eleanor and Lou Gehrig MDA/ALS Research Center, Neurological Institute, Box 107, 710 W 168th St, New York, NY 10032 (paul.gordon@psl.aphp.fr).

Author Contributions: Study concept and design: Gordon, Corcia, Lacomblez, Pochigaeva, Leigh, and Meininger. Acquisition of data: Lacomblez, Pochigaeva, Abitbol, Cudkowicz, Leigh, and Meininger. Analysis and interpretation of data: Gordon, Corcia, Lacomblez, Pochigaeva, Abitbol, Cudkowicz, Leigh, and Meininger. Drafting of the manuscript: Gordon, Corcia, Lacomblez, Pochigaeva, Abitbol, Leigh, and Meininger. Critical revision of the manuscript for important intellectual content: Gordon, Lacomblez, Pochigaeva, Abitbol, Cudkowicz, Leigh, and Meininger. Statistical analysis: Gordon, Pochigaeva, Abitbol, and Meininger. Administrative, technical, and material support: Cudkowicz and Meininger. Study supervision: Gordon, Corcia, and Meininger.

Financial Disclosure: None reported.

Additional Information: An eTable is available at http://www.archneurol.com.

REFERENCES