High plasma concentrations of organic pollutants negatively impact survival in amyotrophic lateral sclerosis

Stephen A. Goutman, MD, MS[*,1,2], **Jonathan Boss, MS**[*,3], **Adam Patterson**[1,2], **Bhramar Mukherjee, PhD**[#,3], **Stuart Batterman, PhD**[#,4], and **Eva L. Feldman, MD, PhD**[#,1,2]

[1]Department of Neurology, University of Michigan, Ann Arbor, MI

[2]Program for Neurology Research and Discovery, University of Michigan, Ann Arbor, MI

[3]Department of Biostatistics, University of Michigan, Ann Arbor, MI

[4]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI

Abstract

Objective: To determine whether persistent organic pollutants (POPs) affect amyotrophic lateral sclerosis (ALS) survival.

Methods: ALS participants seen at the University of Michigan (Ann Arbor, MI, USA) provided plasma samples for measurement of POPs. ALS Disease and clinical features were collected prospectively from the medical records. Survival models used a composite summary measure of exposure due to multiple POPs (environmental risk score or ERS).

Results: 167 participants (40.7% female, n=68) with ALS were recruited, of which 119 died during the study period. Median diagnostic age was 60.9 years (Interquartile Range (IQR) 52.7–68.2), median time from symptom onset to diagnosis was 1.01 years (IQR 0.67–1.67), bulbar onset 28.7%, cervical onset 33.5%, and lumbar onset 37.7%. Participants in the highest quartile of ERS (representing highest composite exposure), adjusting for age at diagnosis, sex, and other covariates, had a 2.07 times greater hazard rate of mortality (p-value 0.018, 95%CI 1.13–3.80) compared to those in the lowest quartile. Pollutants with the largest contribution to the ERS were PBDE 154 (HR 1.53, 95%CI 0.90–2.61), PCB 118 (HR 1.50, 95%CI 0.95–2.39), PCB 138 (HR

Corresponding Author: Stephen A. Goutman, MD, MS, Department of Neurology, 1500 E Medical Center Dr, Ann Arbor, MI 48109-5223, Phone: 734-936-8586, Fax: 734-936-5185, sgoutman@med.umich.edu.
*These authors contributed equally to the manuscript as co-first authors
#These authors contributed equally to the manuscript as co-senior authors

1.69, 95%CI 0.99–2.90), PCB 151 (HR 1.46, 95%CI 1.01–2.10), PCB 175 (HR 1.53, 95%CI 0.98–2.40), and p,p'-DDE (HR 1.39, 95%CI 1.07–1.81).

Conclusions: Higher concentrations of POPs in plasma are associated with reduced ALS survival, independent of age, gender, segment of onset, and other covariates. This study helps characterize and quantify the combined effects of persistent organic pollutants on ALS and supports the concept that environmental exposures play a role in disease pathogenesis.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease.[1] The ALS gene-time-environment hypothesis suggests that underlying genetic susceptibility combined with environmental risk factors initiate disease.[2] We recently reported higher concentrations of persistent organic pollutants (POPs), especially organochlorine pesticides (OCPs), in individuals with ALS compared to healthy controls,[3] as well an increased exposure risk to pesticides.[34]

Analysis of environmental exposures is complicated by multicollinearity--persons are exposed to mixtures of correlated toxins over time.[3] Therefore, methods deciphering causative toxins, or understanding the joint effects, in the multipollutant exposure setting are necessary.[5] A combined single summary measure, the environmental risk score (ERS), accounts for the combined health effects of multiple pollutants.[6]

Given the neurotoxic properties of the POPs previously identified, we evaluated the impact of POPs, via ERS, on ALS survival to yield insight into disease mechanisms and identify modifiable ALS risk factors.

METHODS

Study Design and Participants

All patients greater than 18 years old, able to provide consent and communicate in English, and with definite, probable, probable with laboratory support, possible, or suspected ALS by El Escorial criteria[7] were enrolled during clinical visits at the University of Michigan (UM) ALS Clinic and informed of the study during clinical visits and/or via National ALS Registry announcements between July 25, 2012 and November 16, 2016. Clinical data were collected through February 14, 2018. The study received Institutional Review Board approval; participants provided written informed consent.

Data Sources and Collection

Data collection and pollutant measurement methods are described elsewhere.[3] Briefly, demographic details included age, sex, onset segment, date of symptom onset, date of diagnosis, ALS functional rating scale revised (ALSFRS-R),[8] El Escorial criteria, family history of ALS, use of non-invasive ventilation (NIV), date of death, and social history. Participants completed a self-administered written survey that also included education, military, and smoking history, and average weight 5 years prior to study entry. Plasma samples provided by participants were analyzed for 122 POPs, including OCPs,

polychlorinated biphenyls (PCBs), and polybrominated diphenyl ethers (PBDEs), by gas chromatography/mass spectrometry and individual POPs were selected if they exceeded detection limits as previously described.[3]

Missing Data and Multiple Imputation

The counts and percentage of missing data were calculated. All participants had complete POP measurements. Covariates eligible for imputation are in Supplemental Table 1, of which body mass index (BMI) and BMI slope (defined as the yearly rate of change in BMI over the five years prior to study entry) had the largest fraction of missingness. Military service, family history of ALS, smoking history, NIV use, time between diagnosis and blood draw, education, and El Escorial criteria also had some missingness. To handle missing data, twenty imputed datasets were generated using predictive mean matching via multiple imputation by chained equations (MICE) to augment missing covariates.[9] The imputation model included all pollutant concentrations, sex, age at plasma collection, BMI, BMI slope, education, smoking status, military service, family history of ALS, NIV use, onset segment, El Escorial criteria at diagnosis, age at diagnosis, time between onset and diagnosis (log-transformed), time between diagnosis and blood collection (log-transformed), and the Nelson Aalen cumulative hazard function as recommended by White and Royston.[10] Standard imputation diagnostics compared the imputed and observed distributions.[11] Kernel density estimates for imputed BMI and BMI slope were compared with kernel density estimates for the observed values to verify that imputed quantities generated by MICE were consistent with the observed data. Unless explicitly noted, multiply imputed datasets were used for all descriptive analyses and statistical modeling.

Descriptive Analysis

Descriptive analyses of demographic characteristics, survival, onset segment, ALS prognostic factors, and other covariates were conducted. Summary statistics were tabulated for each pollutant, collinearity among exposure concentrations was investigated using pairwise Spearman correlations, and the non-detect rate was evaluated by the percent below the limit of detection (LOD). Substitution with $LOD/\sqrt{2}$ was used to fill-in POP concentrations below their respective LODs.[12] Univariable and multivariable Cox proportional hazards (PH) models determined if significant associations between adjustment covariates and post-diagnosis survival were present.

Single-Pollutant Models

Unadjusted and adjusted single-pollutant Cox PH models were constructed to understand univariable associations between POP exposure and survival by modeling the covariate adjusted increased risk of an earlier death for ALS participant conditional on exposure. POP concentrations were log transformed and subsequently scaled by their interquartile range (IQR) to remove the right skewness from specific POP distributions, improve interpretability of resultant hazard ratio (HR) estimates by comparing the 75th percentile to the 25th percentile of exposure, and facilitate comparability between POPs. All POPs were measured in ng/L.

ERS Construction

To examine the risk attributable to mixtures of POPs, an ERS[6] was computed for each subject. An ERS is a weighted sum of pollutant concentrations: $ERS_i = \widehat{\beta_1} E_{1,i} + \widehat{\beta_2} E_{2,i} + \cdots + \widehat{\beta_p} E_{1p,i}$ where ERS_i denotes the ERS for participant i, $E_{1,i} \ldots E_{p,i}$ are the p log-transformed IQR-scaled exposure measurements, and $\widehat{\beta_1} \ldots \widehat{\beta_p}$ are the regression coefficients corresponding to each POP obtained from a regression model for time to event data. ERS was constructed using beta-coefficients obtained from a regularized Cox regression, which has better performance in the presence of collinear exposures than standard Cox regression.[13][14] Specifically, we chose to introduce a ridge-type penalty term to handle collinearity among the POPs.[15] The implicit physiological assumption in constructing ERS with ridge regression is that exposure of many individual toxins each contribute in small amounts to post-diagnosis survival.

Survival Model with ERS as the Predictor

We selected the ridge tuning parameter corresponding to the regularized Cox PH model from a grid of fixed values that minimized the cross-validated partial likelihood deviance (5 folds), a measure of model fit in Cox regression. The resulting regression coefficients were used as weights to construct ERS. One common ERS was obtained by averaging weights from the 20 imputed datasets.

For ALS participants, unadjusted and adjusted Cox PH survival models were fit for continuous ERS and ERS quartiles, where the start time was defined as the diagnosis date. ERS was IQR standardized such that the HR corresponding to ERS is the mortality risk for a participant in the 75th versus 25th exposure percentile. All models were adjusted for sex, age at diagnosis, BMI at study entry, BMI slope, education, smoking status, military service, family history of ALS, time between diagnosis and symptom onset (log-transformed), time between diagnosis and blood draw (log-transformed), NIV use, onset segment, and El Escorial criteria at diagnosis. Scaled Schoenfeld residuals were used to assess the proportionality of hazards assumptions in the Cox PH models.[16] For each imputed dataset and ERS quartile, adjusted survival curves corresponding to the adjusted Cox PH model were determined using inverse probability weighting, and adjusted median survival was subsequently estimated.[17] Adjusted median survival estimates and their respective bootstrapped variance estimates were pooled across imputed datasets to obtain confidence intervals.[18][19]

Statistical Software

All analyses were performed using R statistical software, version 3.4.4 (www.R-project.org). Multiple imputation was implemented using the mice package, version 2.46.0 and regularized Cox regression was fit using the glmnet package, version 2.0–13. Adjusted survival curves and median survival estimates by ERS quartile were determined using the survminer package, version 0.4.2.

RESULTS

Study Population

Demographic and clinical characteristics of the study population, which included 167 ALS participants and 119 observed deaths, are summarized in Table 1 prior to multiple imputation. One-hundred-twelve participants declined participation or were ineligible during the enrollment period. Mean ALSFRS-R was 35.9±6.3 (n=142) at study entry and 31.3±7.2 (n=138) at plasma collection. Prognostic risk factor summaries (Table 1) include: (1) age at diagnosis (median, IQR = 60.9, 52.7–68.2 years); (2) sex (40.7% female, 59.3% male); (3) time between symptom onset to diagnosis (median, IQR = 1.01, 0.67–1.67 years); (4) time between diagnosis and blood collection (median, IQR =0.57, 0.36–0.83 years); (5) onset segment (28.7% bulbar, 33.5% cervical, 37.7% lumbar); (6) El Escorial criteria (definite 26.9%, probable 30.5%, probable laboratory supported 25.7%, possible 16.2%); (7) NIV users 58.7%; (8) non-smokers 55.1%; and (9) family history of ALS 10.8%. Participants generally had a decreasing BMI prior to enrollment (median annual decrease in BMI over five years prior to enrollment: 0.2 kg/m^2, IQR: –0.1–0.6). Riluzole use history was known for 125 of the 167 cases, where 116 used riluzole; therefore, riluzole use was not included as an adjustment covariate in the survival models. The study cohort predated the availability of edaravone. Although every ALS case had complete POP measurements, only 97 of the 167 ALS cases had complete covariate information. Specifically, 33.8% of ALS cases had missing self-reported BMI information prior to study entry, while other covariates had at most 5% missingness (Supplemental Table 1).

POPs

Pollutant concentrations are listed in Supplemental Table 2, and pairwise Spearman correlations among POPs are summarized in Supplemental Figure 1. PBDEs and OCPs had relatively high detect rates, with the lowest detect rate in the two POPs classes being PBDE 154 (28% below LOD) and cis-chlordane (35% below LOD), respectively. With the exception of PCB 151 (0.6% below LOD) and PCB 202 (16% below LOD), non-detect rates for PCBs exceeded 25% below LOD. Overall, ALS participants had moderate negative interclass correlations between PBDEs and PCBs, moderate positive intraclass correlations among PBDEs, and strong positive intraclass correlations among PCBs. OCPs were weakly correlated with PBDEs, PCBs, and amongst themselves.

Hazard Models

Univariable HR for demographic and ALS-related covariates prior to multiple imputation are listed in Table 1, while univariable and multivariable HRs for these covariates after multiple imputation are in Supplemental Table 3. Irrespective of covariate adjustment or imputation, longer times between diagnosis and blood draw (unadjusted HR: 0.49, 95% CI: 0.38, 0.63) and longer time between symptom onset and diagnosis (unadjusted HR: 0.76, 95% CI: 0.60, 0.96) are significantly associated with lower rate of death (Supplemental Table 3). Moreover, a subject with definite ALS at their initial visit has a 2.35 times higher unadjusted mortality rate (95% CI: 1.20, 4.58) compared to a subject with possible ALS based on El Escorial criteria (Supplemental Table 3). Bulbar onset is significantly associated with shorter survival in the unadjusted Cox model (HR: 2.31, 95% CI: 1.47, 3.61); however,

it becomes non-significant in a model adjusted for other covariates after imputation (HR: 1.36, 95% CI: 0.76, 2.44) (Supplemental Table 3). Conversely, females have a nearly significant association with shorter survival in the unadjusted model (HR: 1.40, 95% CI: 0.97, 2.02), but a non-significant association upon adjustment for other covariates (HR: 0.95, 95% CI: 0.57, 1.60) (Supplemental Table 3).

Unadjusted and adjusted single-pollutant Cox PH models are shown in Figure 1 (see Supplemental Table 4 for exact numerical results). A subject in the 75th percentile of p,p'-DDE concentration has a 1.39 times higher mortality rate (95% CI: 1.07, 1.81) compared to a subject in the 25th percentile of p,p'-DDE exposure when adjusted for sex, age at ALS diagnosis, BMI at study entry, BMI slope, education, smoking status, military service, family history of ALS, time between diagnosis and symptom onset (log-transformed), time between diagnosis and blood draw (log-transformed), NIV use, onset segment, and baseline El Escorial criteria. All other unadjusted and adjusted single-pollutant associations are either marginally significant or non-significant. Supplemental Table 5 compares the adjusted single-pollutant model results between the imputed and complete case datasets. Results from complete case analysis are different from the unadjusted single-pollutant models and adjusted single-pollutant models after imputation, indicating that the ALS cases with complete covariate information may systematically differ from ALS cases without complete covariate information.

ERS Models

ERS weights derived for each imputed dataset are included in Supplemental Table 6a (imputed datasets 1–10) and Supplemental Table 6b (imputed datasets 11–20). The "Average" column in Supplemental Table 6b refers to the final weights used to construct the overall ERS. The largest contributors to ERS were PBDE 154, PCB 118, PCB 138, PCB 151, PCB 175, and p,p'-DDE, all of which had notably larger weights than the other POPs. From Figure 1 and Supplemental Table 4, the rate of death is 1.22 times higher for a subject in the 75th percentile of exposure, as determined via ERS, compared to a subject in the 25th percentile of exposure (95% CI: 0.95, 1.56). After adjusting for all covariates in Table 1, with the exception of age at plasma collection (since age at diagnosis and time between diagnosis and plasma collection are already included), the association is amplified to a 1.59 times higher rate of dying for a subject in the 75th percentile of exposure compared to a subject in the 25th percentile of exposure (95% CI: 1.16, 2.19).

ALS participants were categorized into quartiles based on their ERSs to account for potential non-linearity in the association between mixtures of POPs and post-diagnosis survival. The unadjusted and adjusted models after imputation are presented in Table 2. After adjusting for covariates, a subject in ERS quartile 2 had a 1.50 times higher mortality rate (95% CI: 0.82, 2.76), a subject in ERS quartile 3 had a 1.84 times higher mortality rate (95% CI: 0.96, 3.52), and a subject in ERS quartile 4 had a 2.07 times higher mortality rate (95% CI: 1.13, 3.80) compared to a subject in ERS quartile 1. Figure 2 shows the adjusted survival curves corresponding to the Cox PH model with ERS quartiles. The estimated adjusted median survival time is 2.47 years (95% CI: 1.84, 3.11) in ERS quartile 1, 2.12 years (95% CI: 1.68, 2.56) in ERS quartile 2, 1.99 years (95% CI: 1.61, 2.37) in ERS

quartile 3, and 1.89 years (95% CI: 1.55, 2.22) in ERS quartile 4. This corresponds to an approximately seven-month survival advantage (95% CI: −1.4, 15.4 months) for ALS subjects in ERS quartile 1 compared to ALS subjects in ERS quartile 4, though the result is not statistically significant.

DISCUSSION

ALS is a heterogeneous disease, and identifying factors that contribute to disease susceptibility and progression may enhance mechanistic insight and support therapeutic developments.[20] Several ALS clinical and genetic prognostic factors are known: age, bulbar onset, frontotemporal dementia, and poor nutritional status are negatively associated with disease survival;[21] and genetic factors, such as *ATXN2* polyQ repeats[22] and *SOD1,*[23] *CAMTA1,*[24] and *KIF5A*[25] genotype, modify ALS survival. Mounting evidence suggests non-genetic ALS risk factors, namely the ALS exposome, may be a third factor underlying disease risk, and interest in elucidating components of the exposome[26] is increasing. Several environmental exposures are associated with ALS risk including smoking, heavy metal exposure, pesticide exposure, military service.[2] POPs are synthetically produced toxic chemicals that bioaccumlate and have long half-lives in the environment and can spread via food or air.[27] POP uses included pest control (OCPs), heat exchange or additives in paper or plastics (PCBs), or flame retardants (PBDEs).

We are interested in POPs as these toxins bioaccumulate in the Great Lakes, and thus are of particular concern in Michigan. While certain air concentrations of PCBs and OCPs are declining, their halving times are more than a decade[28] and PBDEs concentrations are still increasing in humans worldwide.[29] We show that higher concentrations of mixtures of organochlorine pesticides, PBDEs, and PCBs are associated with reduced ALS survival, independent of known ALS prognostic variables. Our adjusted Cox PH model (Table 2) also demonstrated significant survival covariates of time between symptom onset and diagnosis and a higher El Escorial diagnostic certainty; while these prognostic factors have been supported in other studies,[21] to our knowledge, exposure to POPs has not been linked previously to ALS survival. This finding is important because environmental toxins are hypothesized to play a role in ALS disease susceptibility,[2] yet the identification of such factors is hindered by difficulties in quantifying the exposome. Based on our prior work,[3] we believe that toxic environmental exposures are linked to ALS risk. Further validation of our findings in the current study could provide insight into underlying disease mechanisms and highlight the possibility that POPs not only contribute to disease susceptibility, but also progression. These associations in a case-only analysis also diminish the likelihood of disease progression bias (release of lipophilic environmental toxins as fat reserves diminish)[30] in case-control studies since all participants herein had ALS, lessening differential effects.

The ERS is a novel approach to estimate effects of exposure to multiple pollutants for epidemiological research, and has successfully analyzed National Health and Nutrition Examination Survey (NHANES) biomonitoring data.[6] The observation that small effects corresponding to individual pollutants are cumulative, leading to a much stronger effect corresponding to the combined summary risk score, is consistent with prior literature.[6] The

strength of ERS lies in characterizing small effects from multiple pollutants that may be correlated, with the key advantage that multiple exposures are reduced to a simple summary score. This allows effects of multiple pollutants to be taken into account, even in studies with small sample sizes. Limitations in using ERS generated from regularized regression include an inability to account for potential non-linearity of exposure effects or exposure interactions.

The study has limitations. We are unable to account for a lifetime of exposures to all environmental toxins that may have influenced survival, but instead restricted our analysis to POPs. We selected a subset of 23 out of 122 organic compounds in our suite of routinely analyzed compounds; these compounds were selected as they are known and persistent toxicants. POP where adjusted by BMI and BMI slope, and not lipids. The one-time collected exposures may not represent true lifetime exposures due to sampling variation, finite chemical half-lives, and temporal variation. Many newer environmental toxins have shorter half-lives, and retrospective exposures cannot be accurately measured. Results may have been affected by disease progression bias (see above),[30] although the models adjust for BMI and time from diagnosis to blood collection to minimize this potential bias. The study sample size is small, and outcomes could vary in a larger population or by region. The single-pollutant adjusted Cox PH models only show statistically significant results for PCB 151 and p,p'-DDE, although these could be false-positives due to the construction of multiple confidence intervals. The fact that 18 of 23 pollutants have adjusted single-pollutant HRs greater than 1 supports the approach of multiple pollutants causing a small effect on survival, which is detected in a combined (ERS) model. Missing data are common in clinical research and the present study is no different. To address this, imputation was used, and in a sensitivity analysis it was demonstrated that single-pollutant HRs for complete cases are overall greater than the HRs in the imputed datasets, indicating that, at a minimum, the association between exposures and survival in ALS subjects is underestimated. Almost all participants had some use of riluzole, and the generalizability outside of riluzole users is not known. While our results suggest an association between POPs and ALS survival, they do not prove causation. Despite these limitations, we contend that the higher concentrations of POPs found in participants with reduced survival may be an important factor in disease progression. Validation of these findings with other cohorts is needed, as there may be geographical variation in POP exposure. Future integration of these findings with genomics will help to better clarify gene-time-environment interactions.

Overall, we demonstrate that higher POP concentrations, as summarized by the dimension-reducing ERS in the multipollutant context, are associated with reduced ALS survival. This study demonstrates the advantage of summarizing the simultaneous effects of multiple pollutants in one model, thereby providing a framework for future studies. Finally, these results have important implications for our understanding of ALS pathogenesis, heterogeneity, and progression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

1. Goutman SA. Diagnosis and Clinical Management of Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders. Continuum (Minneap Minn) 2017;23(5, Peripheral Nerve and Motor Neuron Disorders):1332–59. doi: 10.1212/con.0000000000000535 [published Online First: 2017/10/03] [PubMed: 28968365]

2. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. Nature reviews Neurology 2013;9(11):617–28. doi: 10.1038/nrneurol.2013.203 [PubMed: 24126629]

3. Su FC, Goutman SA, Chernyak S, et al. Association of Environmental Toxins With Amyotrophic Lateral Sclerosis. JAMA neurology 2016;73(7):803–11. doi: 10.1001/jamaneurol.2016.0594 [PubMed: 27159543]

4. Yu Y, Su FC, Callaghan BC, et al. Environmental risk factors and amyotrophic lateral sclerosis (ALS): a case-control study of ALS in Michigan. PloS one 2014;9(6):e101186. doi: 10.1371/journal.pone.0101186 [PubMed: 24979055]

5. Sun Z, Tao Y, Li S, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. Environmental health : a global access science source 2013;12(1):85. doi: 10.1186/1476-069x-12-85 [published Online First: 2013/10/08] [PubMed: 24093917]

6. Park SK, Tao Y, Meeker JD, et al. Environmental risk score as a new tool to examine multi-pollutants in epidemiologic research: an example from the NHANES study using serum lipid levels. PloS one 2014;9(6):e98632. doi: 10.1371/journal.pone.0098632 [PubMed: 24901996]

7. Brooks BR, Miller RG, Swash M, et al. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis 2000;1(5):293–99. doi: doi: 10.1080/146608200300079536 [PubMed: 11464847]

8. Cedarbaum JM, Stambler N, Malta E, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). Journal of the neurological sciences 1999;169(1–2):13–21. [published Online First: 1999/12/14] [PubMed: 10540002]

9. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations inR. Journal of Statistical Software 2011;45(3):67. doi: 10.18637/jss.v045.i03 [published Online First: 2011–12-12]

10. White IR, Royston P. Imputing missing covariate values for the Cox model. Statistics in medicine 2009;28(15):1982–98. doi: 10.1002/sim.3618 [published Online First: 2009/05/20] [PubMed: 19452569]

11. Raghunathan T, Bondarenko I. Diagnostics for Multiple Imputations SSRN 2007

12. Hornung RW, Reed LD. Estimation of Average Concentration in the Presence of Nondetectable Values. Applied Occupational and Environmental Hygiene 1990;5(1):46–51. doi: 10.1080/1047322X.1990.10389587

13. Simon N, Friedman JH, Hastie T, et al. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent 2011 2011;39(5):13. doi: 10.18637/jss.v039.i05 [published Online First: 2011–03-01]

14. Xue X, Kim MY, Shore RE. Cox regression analysis in presence of collinearity: an application to assessment of health risks associated with occupational radiation exposure. Lifetime Data Anal 2007;13(3):333–50. doi: 10.1007/s10985-007-9045-1 [PubMed: 17661175]

15. Hui Z, Trevor H. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005;67(2):301–20. doi: doi:10.1111/j. 1467-9868.2005.00503.x

16. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81(3):515–26. doi: 10.1093/biomet/81.3.515

17. Therneau TM, Crowson CS, Atkinson EJ. Adjusted Survival Curves 2015

18. Little RJA, Rubin DB. Statistical Analysis with Missing Data 2 ed: Wiley Series in Probability and Statistics 2002.

19. Dong Y, Peng CY. Principled missing data methods for researchers. SpringerPlus 2013;2(1):222. doi: 10.1186/2193-1801-2-222 [published Online First: 2013/07/16] [PubMed: 23853744]

20. Goutman SA, Feldman EL. Clinical Trials of Therapies for Amyotrophic Lateral Sclerosis: One Size Does Not Fit All. JAMA neurology 2015;72(7):743–4. doi: 10.1001/jamaneurol.2014.4275 [PubMed: 25961436]

21. Chio A, Logroscino G, Hardiman O, et al. Prognostic factors in ALS: A critical review. Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases 2009;10(5–6):310–23. doi: 10.3109/17482960802566824

22. Chio A, Calvo A, Moglia C, et al. ATXN2 polyQ intermediate repeats are a modifier of ALS survival. Neurology 2015;84(3):251–8. doi: 10.1212/wnl.0000000000001159 [published Online First: 2014/12/21] [PubMed: 25527265]

23. Cudkowicz ME, McKenna-Yasek D, Sapp PE, et al. Epidemiology of mutations in superoxide dismutase in amyotrophic lateral sclerosis. Annals of neurology 1997;41(2):210–21. doi: 10.1002/ana.410410212 [PubMed: 9029070]

24. Fogh I, Lin K, Tiloca C, et al. Association of a Locus in the CAMTA1 Gene With Survival in Patients With Sporadic Amyotrophic Lateral Sclerosis. JAMA neurology 2016;73(7):812–20. doi: 10.1001/jamaneurol.2016.1114 [PubMed: 27244217]

25. Nicolas A, Kenna KP, Renton AE, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron 2018;97(6):1268–83.e6. doi: 10.1016/j.neuron.2018.02.027 [PubMed: 29566793]

26. Al-Chalabi A, Pearce N. Commentary: Mapping the Human Exposome: Without It, How Can We Find Environmental Risk Factors for ALS? Epidemiology 2015;26(6):821–3. doi: 10.1097/EDE. 0000000000000381 [PubMed: 26414852]

27. (EPA) EPA. Persistent Organic Pollutants: A Global Issue, A Global Response [Available from: https://www.epa.gov/international-cooperation/persistent-organic-pollutants-global-issue-global-response accessed December 21, 2018 2018.

28. Salamova A, Venier M, Hites RA. Revised Temporal Trends of Persistent Organic Pollutant Concentrations in Air around the Great Lakes. Environmental Science & Technology Letters 2015;2(2):20–25. doi: 10.1021/acs.estlett.5b00003

29. Venier M, Salamova A, Hites RA. Halogenated flame retardants in the Great Lakes environment. Acc Chem Res 2015;48(7):1853–61. doi: 10.1021/acs.accounts.5b00180 [published Online First: 2015/06/09] [PubMed: 26050713]

30. Cragg JJ, Cudkowicz ME, Weisskopf MG. The Role of Environmental Toxins in Amyotrophic Lateral Sclerosis Risk. JAMA neurology 2016;73(7):779–80. doi: 10.1001/jamaneurol.2016.1038 [published Online First: 2016/05/10] [PubMed: 27158795]
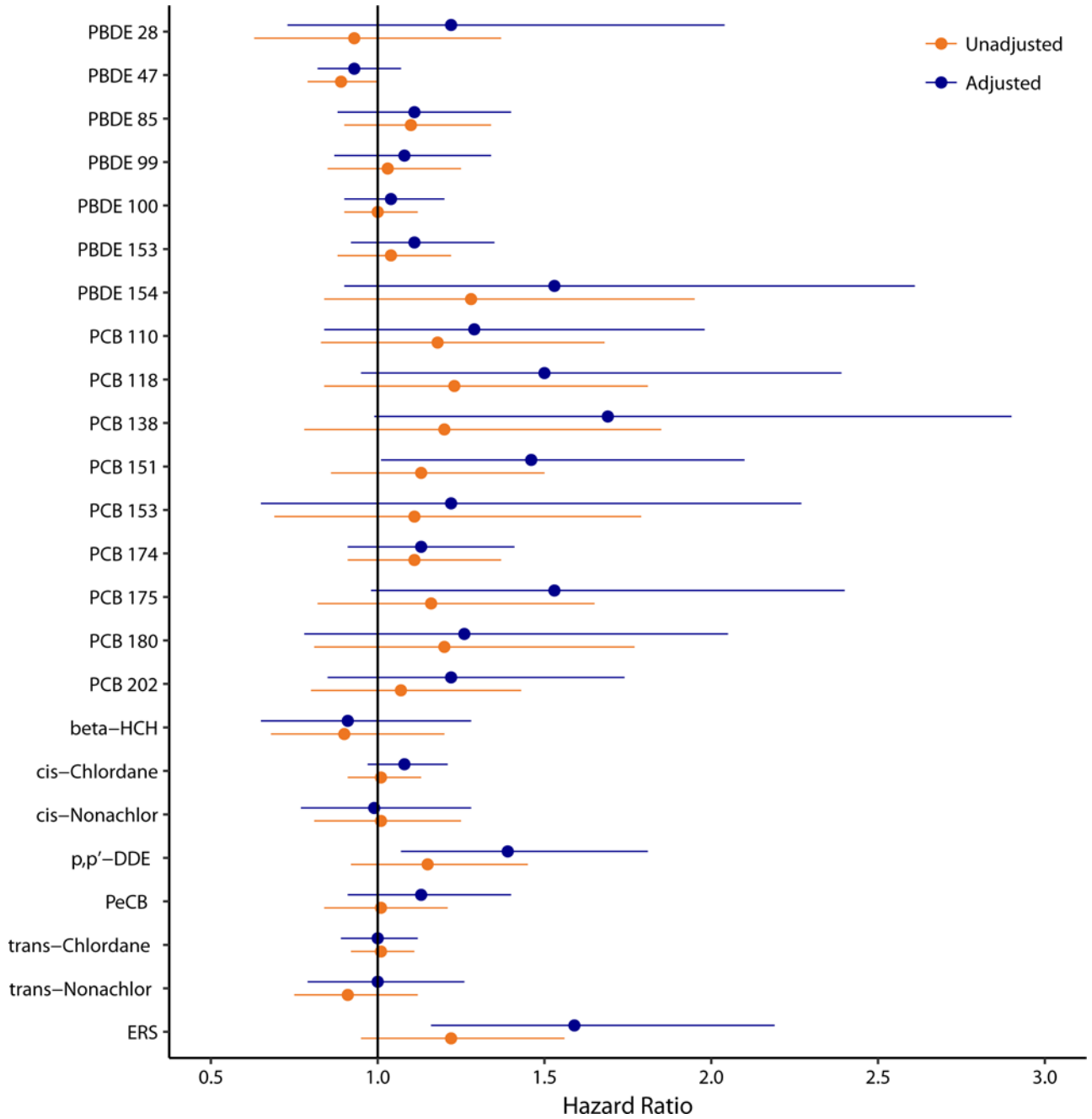
Figure 1. Adjusted Single Pollutant Cox Regression Models.
Hazard ratios from imputed dataset represent the times higher risk of dying comparing ALS cases in the 25th percentile of exposure and the 75th percentile of exposure. Outcome is number of days between diagnosis date and death/censoring time. Models are adjusted for sex, age at diagnosis, time between onset and diagnosis (log-transformed), time between diagnosis and blood draw (log-transformed), BMI at enrollment, BMI slope (defined as the rate of change in BMI over the 5 years prior to study entry), education, smoking status, military service, family history of ALS, NIV use, onset segment, and El Escorial criteria at diagnosis. Abbreviations: DDE, dichlorodiphenyldichloroethylene; ERS, environmental risk

score; HCH, hexachlorocyclohexane; PBDE, polybrominated diphenyl ethers; PCB, polychlorinated biphenyl; PeCB, pentachlorobenzene
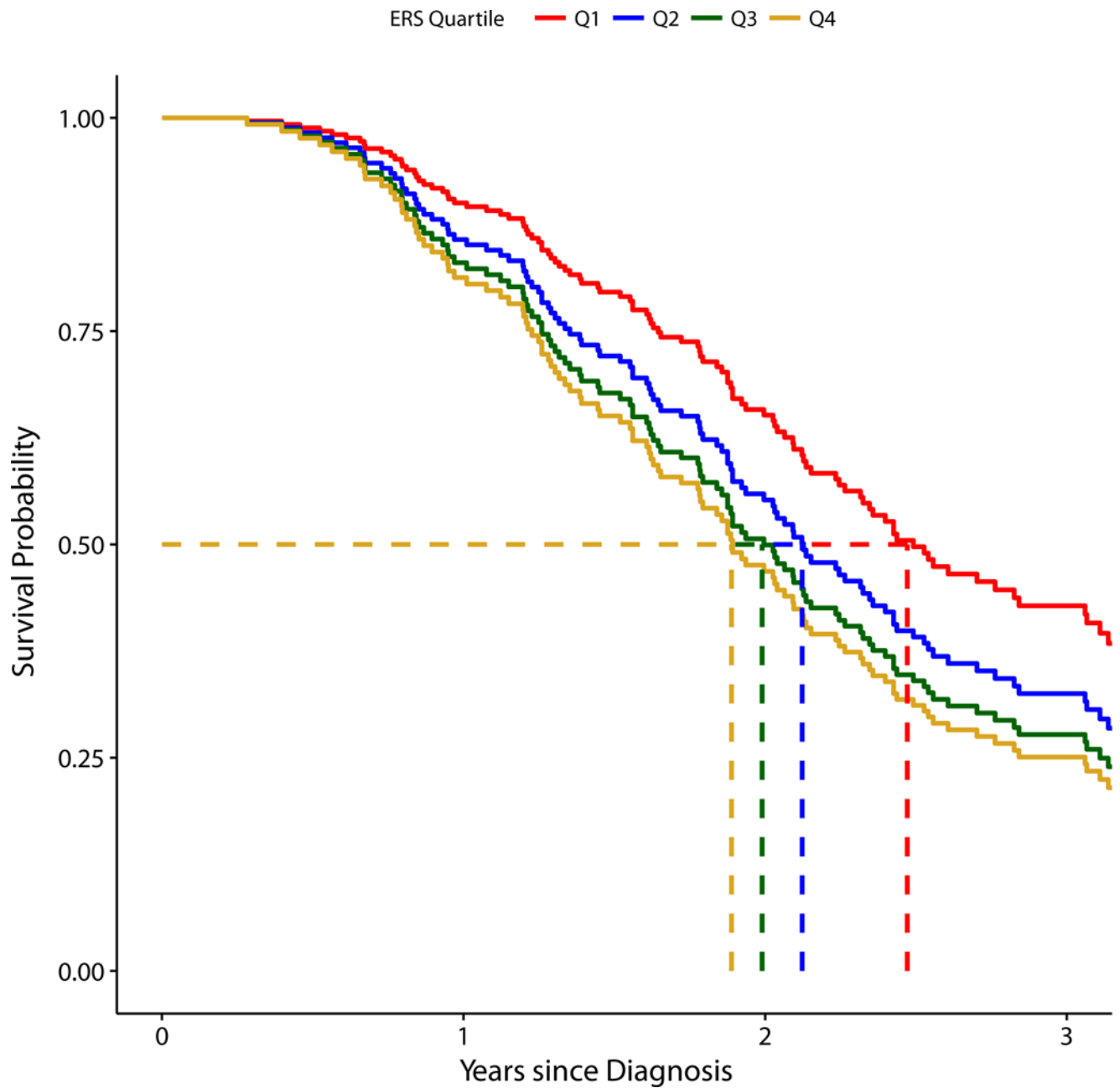
Figure 2. Adjusted Survival Curves Stratified by ERS Quartile.
Dashed lines indicate the median survival in each environmental risk score (ERS) strata. The estimated adjusted median survival time is 2.47 years (95% CI: 1.84, 3.11) for Quartile 1, 2.12 years (95% CI: 1.68, 2.56) for Quartile 2, 1.99 years (95% CI: 1.61, 2.37) for Quartile 3, and 1.89 years (95% CI: 1.55, 2.22) for Quartile 4. The adjusted survival curves and adjusted median survival time estimates are pooled across all 20 imputed datasets.

Table 1.

ALS Participant Demographics Prior to Multiple Imputation

Covariate	ALS Cases (N = 167)	HR	95% CI	P-Value
Age at plasma collection (years) *	**61.6 (53.3–68.8)**	**1.02**	**(1.01, 1.04)**	**0.007**
BMI at study entry (kg/m²) **	25.7 (22.3–28.4)	1.02	(0.97, 1.07)	0.507
BMI slope **	−0.2 (−0.6–0.1)	0.91	(0.65, 1.28)	0.598
Education 1		0.80	(0.55, 1.17)	0.245
HS/GED or less	54 (32.3)			
Some College / Associate's	44 (26.3)			
Bachelor's	42 (25.1)			
Master's / Professional	26 (15.6)			
Missing	1 (0.6)			
Family history of ALS				
No	142 (85.0)	Ref	Ref	Ref
Yes	18 (10.8)	1.26	(0.73, 2.18)	0.402
Missing	7 (4.2)			
Military service				
No	133 (79.6)	Ref	Ref	Ref
Yes	26 (15.6)	0.72	(0.42, 1.25)	0.242
Missing	8 (4.8)			
Sex				
Female	68 (40.7)	1.40	(0.97, 2.02)	0.071
Male	99 (59.3)	Ref	Ref	Ref
Smoking				
Non-smoker	92 (55.1)	Ref	Ref	Ref
Former smoker	52 (31.1)	0.84	(0.56, 1.26)	0.407
Current smoker	18 (10.8)	0.80	(0.42, 1.51)	0.487
Missing	5 (3.0)			
Age at diagnosis (years)	**60.9 (52.7–68.2)**	**1.03**	**(1.01, 1.04)**	**0.001**
El Escorial criteria				
Possible	27 (16.2)	Ref	Ref	Ref
Probable, LS	43 (25.7)	1.86	(0.94, 3.68)	0.074
Probable	51 (30.5)	1.65	(0.84, 3.27)	0.148
Definite	45 (26.9)	**2.36**	**(1.20, 4.61)**	**0.012**
Missing	1 (0.6)			
NIV user				
No	65 (38.9)	Ref	Ref	Ref
Yes	98 (58.7)	0.89	(0.61, 1.30)	0.548
Missing	4 (2.4)			

Covariate	ALS Cases (N = 167)	HR	95% CI	P-Value
Onset segment				
Bulbar	48 (28.7)	**2.31**	**(1.47, 3.61)**	**<0.001**
Cervical	56 (33.5)	Ref	Ref	Ref
Lumbar	63 (37.7)	1.07	(0.69, 1.66)	0.760
Time between diagnosis and blood draw (years) [***2]	0.57 (0.36–0.83)	**0.39**	**(0.30, 0.49)**	**<0.001**
Time between symptom onset and diagnosis (years)[2]	1.01 (0.67–1.67)	**0.76**	**(0.60, 0.96)**	**0.022**

Table of descriptive statistics for the overall ALS participant study population. For continuous variables, Median (25th – 75th percentile), and for categorical variables, N (%). Hazard ratios, 95% confidence intervals, and p-values correspond to univariable unadjusted Cox proportional hazards models.

Abbreviations: ALS, amyotrophic lateral sclerosis; BMI, body mass index; CI, confidence interval; GED, Graduate Equivalency Diploma; HR, hazard ratio; HS, high school; kg, kilograms; LS, lab supported; m, meters; NIV, non-invasive ventilation; Ref, reference category.

[*] Median, 25th percentile, and 75th percentile are computed using 165 cases (2 cases are missing).

[**] Median, 25th percentile, and 75th percentile are computed using 111 cases (56 cases are missing).

[***] Median, 25th percentile, and 75th percentile are computed using 164 cases (3 cases are missing).

[1] Hazard ratio corresponds to those with at least some post-high school education compared to those with no post-high school education.

[2] Hazard ratios correspond to log-transformed time between diagnosis and blood draw and log-transformed time between symptom onset and diagnosis.

Table 2.

Survival Model with ERS as the Predictor

ERS / Covariate	HR	95% CI	P-Value
Unadjusted Cox Model			
ERS Quartile 2	1.69	(0.99, 2.87)	0.054
ERS Quartile 3	1.47	(0.88, 2.47)	0.143
ERS Quartile 4	1.41	(0.86, 2.33)	0.177
Adjusted Cox Model			
ERS Quartile 2	1.50	(0.82, 2.76)	0.189
ERS Quartile 3	1.84	(0.96, 3.52)	0.065
ERS Quartile 4	**2.07**	**(1.13, 3.80)**	**0.018**
NIV user	0.85	(0.54, 1.34)	0.488
BMI at study entry (kg/m^2)	1.00	(0.95, 1.06)	0.994
Rate of change in BMI	0.82	(0.55, 1.23)	0.343
Age at diagnosis (years)	1.02	(0.99, 1.04)	0.159
Time between symptom onset and diagnosis (years) [*]	**0.59**	**(0.43, 0.81)**	**0.001**
Time between diagnosis and blood draw (years) [*]	**0.42**	**(0.30, 0.58)**	**<0.001**
Family history of ALS	1.21	(0.62, 2.36)	0.578
Military service	0.94	(0.44, 2.03)	0.877
Sex (reference = male)	1.03	(0.60, 1.75)	0.924
Former smoker	1.05	(0.62, 1.76)	0.864
Current smoker	1.25	(0.56, 2.77)	0.581
> HS/GED	1.27	(0.80, 2.02)	0.312
Bulbar onset	1.35	(0.76, 2.41)	0.309
Lumbar onset	0.87	(0.51, 1.47)	0.595
El Escorial – Probable	1.64	(0.78, 3.47)	0.194
El Escorial – Probable, LS	**2.30**	**(1.11, 4.77)**	**0.025**
El Escorial – Definite	**3.20**	**(1.52, 6.78)**	**0.002**

Hazard ratios correspond to Cox proportional hazards models with N = 167 Cases (119 observed deaths) pooled across 20 imputed datasets. Adjustment covariates include: non-invasive ventilation (NIV) use, age at diagnosis, time between symptom onset and diagnosis (log-transformed), time between diagnosis and blood draw (log-transformed), BMI at study entry, rate of change in BMI over the five years prior to study entry, family history of ALS, military service, sex, smoking, education, onset segment, and initial El Escorial criteria. Outcome for both models is survival since diagnosis.

Abbreviations: ALS, amyotrophic lateral sclerosis; BMI, body mass index; CI, confidence interval; ERS, environmental risk score; GED, graduate equivalency diploma; HR, hazard ratio; HS, high school; kg, kilograms; LS, lab supported; m, meters.

[*] Covariates are log-transformed.