A post hoc analysis of subgroup outcomes and creatinine in the phase III clinical trial (EMPOWER) of dexpramipexole in ALS

MICHAEL E. BOZIK1, HIROSHI MITSUMOTO2, BENJAMIN R. BROOKS 3, STACY A. RUDNICKI4, DAN H. MOORE5, BING ZHANG6, ALBERT LUDOLPH7, MERIT E. CUDKOWICZ8, LEONARD H. VAN DEN BERG9, JAMES MATHER1, THOMAS PETZINGER JR1 & DONALD ARCHIBALD1

1Knopp Biosciences LLC, Pittsburgh, Pennsylvania, USA, 2Columbia University Medical Center, New York, USA, 3Carolina Neuromuscular/ALS-MDA Center, Carolina Healthcare System, University of North Carolina School of Medicine – Charlotte Campus, Charlotte, North Carolina, USA, 4Department of Neurology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA, 5Calico Computing, Livermore, California, USA, 6MacroStat Inc., Hockessin, Delaware, USA, 7University of Ulm, Ulm, Germany, 8Massachusetts General Hospital, Neurology Clinical Research Institute, Boston, Massachusetts, USA, and 9Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract
Our objective was to compare the phase II and phase III (EMPOWER) studies of dexpramipexole in ALS and evaluate potential EMPOWER responder subgroups and biomarkers based on significant inter-study population differences. In a post hoc analysis, we compared the baseline population characteristics of both dexpramipexole studies and analyzed EMPOWER efficacy outcomes and laboratory measures in subgroups defined by significant inter-study differences. Results showed that, compared with phase II, the proportion of El Escorial criteria (EEC) definite participants decreased (p = 0.005), riluzole use increased (p = 0.002), and mean symptom duration increased (p = 0.037) significantly in EMPOWER. Baseline creatinine (p < 0.001) and on-study creatinine change (p < 0.001) correlated significantly with ALSFRS-R in EMPOWER. In the EMPOWER subgroup defined by EEC-definite ALS, riluzole use, and < median symptom duration (15.3 months), dexpramipexole-treated participants had reduced ALSFRS-R slope decline (p = 0.015), decreased mortality (p = 0.011), and reduced creatinine loss (p = 0.003). In conclusion, significant differences existed between the phase II and EMPOWER study populations in ALS clinical trials of dexpramipexole. In a post hoc analysis of EMPOWER subgroups defined by these differences, potential clinical benefits of dexpramipexole were identified in the subgroup of riluzole-treated, short-symptom duration, EEC-definite ALS participants. Creatinine loss correlated with disease progression and was reduced in dexpramipexole-treated participants, suggesting it as a candidate biomarker.

Key words: Clinical trials, biomarker, mitochondria, therapy, risk

Introduction
Every investigational drug tested in a phase III clinical trial in amyotrophic lateral sclerosis (ALS) during the last 15 years has failed to demonstrate efficacy. The most recent of these phase III trials, known as EMPOWER, evaluated the investigational treatment dexpramipexole in the most comprehensive clinical trial ever conducted in ALS (1). Dexpramipexole failed to meet any prespecified endpoints in this study of 943 ALS participants, despite an earlier two-part phase II study in 102 participants demonstrating dose-dependent trends in reducing the rate of functional decline over 12 weeks (Part 1) and a significant difference (p = 0.046) between the high- and low-dose treatment groups in a joint-rank test of mortality and function over 24 weeks (Part 2) (2). Dexpramipexole at total daily doses up to 300 mg was well-tolerated in both the phase II and phase III studies with no dose-limiting toxicities.

Accordingly, we compared the phase II and phase III dexpramipexole trials to determine if inter-study population differences may have contributed to the
divergent results. The goals of the analyses were to determine whether significant inter-study differences were present, whether such differences might suggest treatment-responsive EMPOWER subgroups and biomarkers of drug effect, and whether such findings could inform future ALS trial design.

The EMPOWER study was designed to replicate the phase II study in every feasible respect except size. The phase II study enrolled 102 participants at 20 U.S. centers in a primary safety and tolerability study. Prespecified efficacy endpoints were the ALSFRS-R slope tested by the mixed-effects linear slopes model and mortality tested by the Cox proportional hazards regression model. The phase III study enrolled 943 participants in 11 countries and tested efficacy on the primary endpoint, the Combined Assessment of Function and Survival (CAFS), and on separate tests of the CAFS components: functional decline, measured by ALSFRS-R change, and mortality. Both studies included riluzole use, the only drug approved for the treatment of ALS and which extends survival by approximately two to three months (3).

Baseline age, site of onset, ALSFRS-R score, and symptom duration are important covariates in predicting rates of disease progression and survival in ALS trials (4). Gender is also reported to affect disease course (5), though not consistently, and definite ALS classification by El Escorial criteria (EEC) is emerging as a prognostic factor (6,7). We evaluated the impact of these factors and riluzole use on clinical outcomes in the EMPOWER placebo population for consistency with previous reports.

The need for biomarkers to monitor and predict disease progression and better inform patient selection in ALS clinical trials is widely recognized (8). While not diagnostic, changes in certain biochemical parameters, including creatinine, have been correlated with disease outcomes (9). We correlated changes from baseline in clinical laboratory tests with EMPOWER outcomes and analyzed the effect of dexpramipexole on select parameters.

Significant progress in understanding the genetic basis of the ALS has been made since the discovery of the SOD1 gene mutation in association with familial ALS. Incorporating genetic markers in ALS clinical trials, in which approximately 10% of participants can be expected to have familial ALS, may be a useful tool for reducing the heterogeneity of clinical trial participants and improving clinical trial design and conduct (10,11).

Materials and methods

Study designs

The study designs, participants, endpoints and statistical analyses for the phase II and EMPOWER trials of dexpramipexole have been reported in detail previously (1,2).

Post hoc statistical analyses

Differences in baseline participant characteristics between the phase II and EMPOWER studies, between EMPOWER treatment groups, and between the definite and not-definite EEC categories were tested by χ² test for categorical variables and by t-test for continuous variables. Characteristics that differed significantly between phase II and phase III were used to select subgroups for efficacy testing. The order of subgroup testing – EEC-definite, riluzole use, median symptom duration – was based on 1) the unexpected, significantly decreased representation in EMPOWER of an aggressive disease phenotype (EEC-definite ALS); and 2) the magnitude of the relative differences between phase II and EMPOWER baseline characteristics from greatest to least. The predictive association between baseline variables and the CAFS was determined by multiple regression.

Baseline clinical laboratory tests were correlated with baseline clinical characteristics and changes from baseline in clinical laboratory tests were correlated with clinical outcomes by treatment groups. Dexpramipexole and riluzole effects on creatinine were estimated using a mixed-effects repeated-measures model to compare effects averaged from baseline through month 12. The model included terms for treatment, visit, and treatment by visit interaction, baseline laboratory value, and baseline laboratory value by visit interaction. The coefficient for treatment main effect represents the mean change in creatinine averaged over each visit (time-averaged creatinine loss). The difference in the coefficients for treatments (placebo minus dexpramipexole) represents the time-averaged difference in creatinine loss. Creatinine loss from baseline through month 12 for EEC-definite participants compared with not-definite participants was also analyzed using the mixed-effects repeated-measures model described above.

Testing for treatment effects on the primary efficacy endpoint in subgroups selected by significant inter-study differences used the methods prespecified in the EMPOWER statistical analysis plan. The CAFS is a joint-rank test of functional outcomes adjusted for mortality by ranking participants’ time to death or change from baseline in ALSFRS-R scores, using follow-up data through the end of the 12-month primary evaluation period (12,13). CAFS ranks were analyzed using an ANCOVA model with treatment as a fixed effect, adjusted for ALSFRS-R, duration from symptom onset to first dose of study treatment, onset site, and concomitant riluzole use as baseline covariates. Functional change was measured by the ALSFRS-R slopes through month 12 analyzed by a linear mixed-effects model, and time to death through month 18 was determined using the Cox proportional hazards model; both models adjusted for the same covariates used in the CAFS ANCOVA. The intent-to-treat (ITT) population
(i.e. randomized participants who received at least one dose of study drug, n = 942) was used for the survival analysis. The efficacy population (i.e. randomized participants receiving at least one dose of study drug and one on-study efficacy assessment, n = 941) was used for all other endpoint analyses. Treatment effects were evaluated both without adjustments for multiple testing and with a Bonferroni correction applied. Given the three subgroup variables (i.e. EEC-definite, riluzole use, median symptom duration), testing for treatment effects in the subgroup with these characteristics requires a nominal $p$-value of 0.007 to reach significance following Bonferroni correction ($p = \alpha/n; \alpha = 0.05, n = 7$; each variable individually (3) + each pair of variables (3) + the triplet variable (1)).

### Results

The dexpramipexole phase II and EMPOWER study populations were similar in baseline age (57.0 years vs. 57.1 years, $p = 0.932$), gender (64% male vs. 64% male, $p = 0.920$), ALSFRS-R (38.0 vs. 38.2, $p = 0.790$), and bulbar-onset ALS (18% vs. 23%, $p = 0.200$) (Table I). In contrast, there were significant baseline differences between phase II and EMPOWER EEC-definite ALS participants (46% vs. 32%, $p = 0.005$), riluzole use (61% vs. 75%, $p = 0.002$), and symptom duration (14.0 months vs. 15.2 months, $p = 0.037$).

Consistent with prior reports, EMPOWER EEC-definite participants in the placebo group (n = 156) had worse outcomes on CAFS (398.4 vs. 502.1; $p < 0.001$), ALSFRS-R slopes (~1.37 vs. ~1.06; $p < 0.001$), and mortality (H.R. 1.39; $p = 0.093$) compared with not-definite participants (n = 312) (6,7). Creatinine loss from baseline through month 12 was also greater for EEC-definite participants compared with not-definite participants (~7.0 $\mu$m/l vs. ~5.3 $\mu$m/l, $p = 0.056$). EMPOWER EEC-definite participants were significantly younger (55.8 years vs. 57.7 years, $p = 0.018$) and more female (40.9% vs. 33.3%, $p = 0.024$) than not-definite participants (Table II). They also had significantly faster pre-study ALSFRS-R progression rates (~0.80 vs. ~0.71, $p = 0.016$), and lower baseline ALSFRS-R scores (36.7 vs. 38.9, $p < 0.001$), predicted slow vital capacity (85.6% vs. 90.4%, $p < 0.001$), and plasma creatinine (68.4 $\mu$m/l vs. 71.3 $\mu$m/l, $p = 0.010$). There were no significant differences in riluzole use (76.2% vs. 74.8%, $p = 0.686$), bulbar onset (25.7% vs. 22.1%, $p = 0.216$), or symptom duration at baseline (15.3 months vs. 15.2 months, $p = 0.625$).

The effects of riluzole on mortality in EMPOWER were also consistent with previous reports (3). Among EMPOWER participants in the placebo group (n = 468), the hazard ratio for mortality was 0.72 (95% CI 0.48–1.09; $p = 0.123$) for participants on riluzole (n = 350) vs. no riluzole (n = 118). In participants receiving dexpramipexole (n = 474), the hazard ratio for mortality was reduced to 0.59 (95% CI 0.38–0.91; $p = 0.018$) for participants on concomitant riluzole (n = 359) vs. no riluzole (n = 115).

As shown in Figure 1A, participants receiving dexpramipexole in the EMPOWER subgroups selected by significant inter-study differences (EEC-definite; EEC-definite on riluzole; EEC-definite on riluzole with symptom duration <15.3 months) had improved CAFS outcomes compared with placebo-treated participants. In contrast to the overall efficacy population, in which no CAFS benefit was observed in dexpramipexole-treated participants (n = 473) compared with placebo (n = 468) (least square difference (l.s.d.) 2.8; 95% CI –28.9–34.4; $p = 0.863$), the CAFS outcome was improved for dexpramipexole-treated participants (n = 147) in the EEC-definite subgroup (n = 303) compared with placebo (n = 156) (l.s.d. 28.3; 95% CI 25.3–32.4; $p = 0.298$). The CAFS was further improved for dexpramipexole-treated participants (n = 113) in the EEC-definite subgroup on concomitant riluzole (n = 231) compared with placebo (n = 118) (l.s.d. 51.4; 95% CI –8.4–111.1; $p = 0.092$), and for dexpramipexole-treated participants (n = 54) in the EEC-definite ALS, riluzole use, and short symptom duration (<15.3 months) subgroup (n = 110).

### Table I. Covariate distribution across studies.

<table>
<thead>
<tr>
<th>Baseline characteristic</th>
<th>Phase II</th>
<th>EMPOWER</th>
<th>$p$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ALSFRS-R</td>
<td>38.0</td>
<td>38.2</td>
<td>0.790</td>
</tr>
<tr>
<td>Mean age</td>
<td>57.0</td>
<td>57.1</td>
<td>0.932</td>
</tr>
<tr>
<td>Site of onset (limb)</td>
<td>82%</td>
<td>77%</td>
<td>0.200</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>64%</td>
<td>64%</td>
<td>0.920</td>
</tr>
<tr>
<td>Mean symptom duration</td>
<td>14.0 months</td>
<td>15.2 months</td>
<td>0.037</td>
</tr>
<tr>
<td>Riluzole (yes)</td>
<td>61%</td>
<td>75%</td>
<td>0.002</td>
</tr>
<tr>
<td>EEC-definite</td>
<td>46%</td>
<td>32%</td>
<td>0.005</td>
</tr>
</tbody>
</table>

*p-value from $\chi^2$ test for categorical variables and by $t$-test for continuous variables.

### Table II. EMPOWER baseline covariates in EEC-definite and not-definite participants.

<table>
<thead>
<tr>
<th>Baseline characteristic</th>
<th>EEC-definite</th>
<th>EEC-not-definite</th>
<th>$p$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ALSFRS-R</td>
<td>36.7</td>
<td>38.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean age</td>
<td>55.8</td>
<td>57.7</td>
<td>0.018</td>
</tr>
<tr>
<td>Site of onset (limb)</td>
<td>74%</td>
<td>78%</td>
<td>0.216</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>41%</td>
<td>33%</td>
<td>0.024</td>
</tr>
<tr>
<td>Mean symptom duration</td>
<td>15.3 months</td>
<td>15.2 months</td>
<td>0.625</td>
</tr>
<tr>
<td>Riluzole (yes)</td>
<td>76%</td>
<td>75%</td>
<td>0.686</td>
</tr>
<tr>
<td>Pre-study progression rate</td>
<td>$-0.80$</td>
<td>$-0.71$</td>
<td>0.016</td>
</tr>
<tr>
<td>Predicted SVC</td>
<td>86%</td>
<td>90%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Plasma creatinine</td>
<td>68.4 $\mu$m/l</td>
<td>71.3 $\mu$m/l</td>
<td>0.010</td>
</tr>
</tbody>
</table>

*p-value from $\chi^2$ test for categorical variables and by $t$-test for continuous variables.
by reduced ALSFRS-R slopes decline, and in a decreased hazard for mortality, as shown in Figure 1B and Figure 1C. Compared with the efficacy population, in which there was no difference in the ALSFRS-R monthly slope between dexpramipexole- and placebo-treated subjects (0.02; 95% CI –0.09–0.13; \( p = 0.732 \)), the monthly ALSFRS-R slopes decline was improved for participants receiving dexpramipexole vs. placebo in the EEC-definite subgroup (0.12; 95% CI –0.08–0.32; \( p = 0.231 \)), the EEC-definite with concomitant riluzole subgroup (0.24; 95% CI 0.02–0.46, \( p = 0.029 \)), and the EEC-definite with concomitant riluzole and short symptom duration subgroup (0.49, 95% CI 0.10–0.87; \( p = 0.015 \)). Similarly, compared to the hazard for mortality in the ITT population (H.R. 0.98; 95% CI 0.75–1.30; \( p = 0.909 \)), the hazard for mortality for dexpramipexole-treated participants was reduced in the EEC-definite subgroup (H.R. 0.76; 95% CI 0.48–1.23; \( p = 0.263 \)), the EEC-definite with concomitant riluzole subgroup (H.R. 0.63; 95% CI 0.36–1.10, \( p = 0.103 \)), and the EEC-definite with concomitant riluzole and short symptom duration subgroup (H.R. 0.37, 95% CI 0.17–0.80; \( p = 0.011 \)). Following Bonferroni correction, the nominal \( p \)-values for dexpramipexole treatment effect on reducing the monthly ALSFRS-R slope decline and hazard for mortality in the EEC-definite with concomitant riluzole and short symptom duration subgroup were \( p = 0.105 \) and \( p = 0.077 \), respectively.

Creatinine levels are significantly lower in ALS subjects compared with healthy controls and low plasma creatinine has been correlated with worse ALS outcomes (8,14). In EMPOWER, there was a significant correlation between baseline creatinine and baseline ALSFRS-R (Pearson correlation coefficient = 0.305, \( p < 0.001 \)). There was an even stronger correlation between creatinine reduction and ALSFRS-R decline at month 6 among placebo participants (Pearson correlation coefficient = 0.380, \( p < 0.001 \)).

Creatinine loss from baseline through month 12 was greater for placebo participants vs. dexpramipexole participants in the EMPOWER ITT population (−6.0 \( \mu \)m/l vs. −4.8 \( \mu \)m/l, \( p = 0.036 \)). This finding is consistent with the phase II study, in which participants receiving dexpramipexole 300 mg daily had reduced creatinine loss at month 3 compared with placebo participants (0.0 \( \mu \)m/l vs. −5.3 \( \mu \)m/l, \( p = 0.011 \)). As shown in Figure 2A, creatinine loss through month 12 was greatest for placebo participants not on riluzole (−6.5 \( \mu \)m/l, 95% CI −8.1–5.0), followed by dexpramipexole participants not on riluzole (−5.8 \( \mu \)m/l, 95% CI −7.4 to −4.3), placebo participants on riluzole (−5.8 \( \mu \)m/l, 95% CI −6.7 to −4.9), and dexpramipexole participants on riluzole (−4.5 \( \mu \)m/l, 95% CI −5.4 to −3.6). Compared with subjects taking neither dexpramipexole nor riluzole, subjects taking both drugs had significantly less

Figure 1. Clinical outcomes in the EMPOWER overall population (ITT or Efficacy) and subgroups defined by inter-study differences; model based estimates of treatment effect and 95% confidence intervals. (A) Difference from placebo in mean CAFS score through month 12 for dexpramipexole treatment in the Efficacy, EEC-definite, EEC-definite + riluzole, and EEC-definite + riluzole and short symptom duration subgroups. The CAFS is a joint-rank test of functional outcomes adjusted for mortality by ranking participants’ time to death or change from baseline in ALSFRS-R scores, using follow-up data through the end of the 12-month primary evaluation period. Means determined from analysis of covariance. (B) Difference from placebo in ALSFRS-R slope estimates through month 12 for dexpramipexole treatment in the Efficacy, EEC-definite, EEC-definite + riluzole, and EEC-definite + riluzole and short symptom duration subgroups. Slopes analyzed by a linear mixed-effects model. (C) Hazard ratio for mortality (dexpramipexole/placebo) in the ITT, EEC-definite, EEC-definite + riluzole, and EEC-definite + riluzole and short symptom duration subgroups. Hazard ratio for mortality determined using the Cox proportional hazards model.

compared with placebo (\( n = 56 \)) (l.s.d. 72.1; 95% CI −6.8–150.9; \( p = 0.073 \)).

Corresponding dexpramipexole benefits were observed in these subgroups on function, defined
creatinine loss at 12 months (2.0 μm/l; 95% CI 0.2–3.8, \( p = 0.026 \)). Consistent with clinical outcomes in EMPOWER subgroups, creatinine sparing for dexpramipexole vs. placebo treated participants increased progressively from the ITT population (1.2 μm/l; 95% CI 0.1–2.2, \( p = 0.036 \)), to the EEC-definite subgroup (2.1 μm/l; 95% CI 0.1–4.0, \( p = 0.035 \)), to the EEC-definite with concomitant riluzole subgroup (2.8 μm/l; 95% CI 0.7–5.0, \( p = 0.009 \)), and was greatest in the EEC-definite with concomitant riluzole and short symptom duration subgroup (4.8 μm/l; 95% CI 1.6–8.0, \( p = 0.003 \)) (Figure 2B). The nominal \( p \)-value for dexpramipexole treatment effect on creatinine sparing in the EEC-definite with concomitant riluzole and short symptom duration subgroup after Bonferroni correction was \( p = 0.021 \).

Discussion

Despite significant progress in understanding the genetic basis of ALS (13,14) in characterizing its molecular pathogenesis (15), and in developing improved clinical trial outcome measures and effective clinical trial consortia (16), no new drug has been approved for the treatment of ALS since riluzole in 1995. Particularly vexing has been the failure of drugs in phase III trials that showed promise in phase II or, in the case of IGF1, failed to replicate a positive phase III trial (17). The reasons for these failures include the absence of biomarkers, disease heterogeneity, trial design limitation, participant selection inconsistency, and ineffective treatments (18).

The EMPOWER result has only increased the debate about strategies in ALS drug development. In some views, ineffective drugs are advancing to phase III trials because phase II trials are too small to rule out false-positive error (19). In other views, promising investigational drugs are failing in phase III because those trials are too large and non-selective to detect a treatment signal within a heterogeneous population (20).

These concerns in ALS clinical research are consistent with emerging regulatory policy regarding drug development in complex diseases. A current draft guidance from the U.S. Food and Drug Administration (FDA) encourages incorporating enrichment strategies into clinical trials, which the FDA defines as ‘the prospective use of any participant characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population’ (21). The key condition in this definition is that a drug effect is ‘in fact present’.

In phase II, dexpramipexole demonstrated several criteria important in advancing a drug candidate to late-stage testing: 1) evidence of a dose-dependent treatment effect; 2) replication of treatment effect across re-randomized cohorts; 3) drug exposures consistent with preclinical pharmacodynamic effects; and 4) consistency across endpoints. Despite these promising results, dexpramipexole failed to meet the prespecified efficacy endpoints in EMPOWER.

Notable in the phase II study of dexpramipexole was a placebo population with more rapidly progressing disease (ALSFRS-R monthly slope estimate of \(-1.28 \)) than reported for placebo groups in the recent phase III ALS trials of lithium, TCH346, and minocycline (ALSFRS-R slopes of \(-0.78 \), –0.94, and \(-1.04 \), respectively) (22–24) or observed in EMPOWER. Like ceftriaxone, which showed promise in a phase II ALS trial against a rapidly progressing placebo group but failed to show benefit in phase III (25,26), efficacy signals may be more easily detected in cohorts with aggressive disease simply because they exhibit greater dynamic range. More aggressive ALS may also represent a phenotype more responsive to specific interventions (27).
significantly smaller in EMPOWER. Symptom duration and EEC diagnosis were significant predictors of CAFS outcome, and the inter-study covariate differences may have contributed to the reduced rate of disease progression in EMPOWER vs. phase II. In particular, the marked decline in the proportion of EEC-definite participants had the effect of diluting the number of participants with a more aggressive disease phenotype (1).

In addition to being clinically distinct at baseline and on study, EEC-definite participants had significantly lower baseline creatinine levels and greater on-study creatinine loss than non-definite participants. More importantly, creatinine has re-emerged as a potential biomarker of ALS progression and of drug related (dexpramipexole-riluzole) clinical effects in the treated population. Specifically, increased creatinine sparing was observed in EMPOWER subgroups with improved clinical outcomes, an effect also seen in the phase II trial.

Interestingly, creatinine sparing was greater in participants receiving both dexpramipexole and riluzole than participants receiving either drug alone, and the creatinine sparing effect of the combined drugs was significantly increased compared with participants receiving neither agent. This additive effect on creatinine was mirrored in the EMPOWER clinical outcomes. Riluzole use was associated with a 28% reduction in the hazard ratio for mortality among placebo participants, while the magnitude of the riluzole treatment effect increased to 41% in the combined dexpramipexole-riluzole treatment group. In contrast to previous clinical trials that have shown no added benefit of riluzole in combination with the investigational drug or suggested a negative interaction (28,29), the EMPOWER results suggest the potential of an additive benefit of dexpramipexole with riluzole, a result that could be important in designing future clinical trials.

A potential limitation of using the CAFS as a primary endpoint in ALS clinical trials emerged from the post hoc analysis of EMPOWER. Under most circumstances, the sensitivity of the CAFS in detecting a significant difference between treatment groups is concordant with the sensitivity of its component tests. However, in the subgroup analysis of EEC-definite participants on riluzole with symptom duration < 15.3 months, the CAFS proved to be less powerful than either component test, the mixed effects model for ALSFRS-R slopes or the Cox proportional hazards model. The CAFS is a non-parametric rank test and, in this case, ranking the joint outcome difference of function adjusted for mortality significantly compressed estimate of treatment effect compared with either of the separate parametric tests of function or mortality.

Thirty-five percent (35%) of EMPOWER participants consented to genetic testing. The results of those tests in that subgroup are pending but the genetic marker data they yield could provide important correlations with clinical phenotypes (e.g. the frequency and extent of C9orf72 hexanucleotide repeat expansion in patients with EEC-definite disease), clinical outcomes, and drug response in EMPOWER.

Subgroup analysis, both prespecified and post hoc, is an important tool for the thorough understanding of clinical trials of investigational drugs and established treatments. In trials of disease-modifying agents in neurodegenerative disorders, for example, post hoc analyses have been reported extensively in assessing the heterogeneity of treatment effects and comparing the efficacy of multiple sclerosis therapeutics (30–32). Subgroup analyses can inform treatment guidelines, enhance future clinical trial designs, improve participant and biomarker selection, as well as generate hypotheses for follow-on clinical trials. Subgroup analyses must be considered with caution, however, as they can be over-interpreted and should be viewed against both the methodology used in the analyses and the total body of evidence for the investigational new drug or established treatment being evaluated (33).

Subgroup analyses, especially post hoc, are by definition subject to bias. However, some steps can be taken to reduce bias (34,35). These include conducting exploratory analyses on the basis of prior hypotheses, accounting for imbalances in the distribution of prognostic baseline characteristics within subgroups, reporting absolute and relative risk reductions with 95% confidence intervals, and adjusting for multiple comparisons. We hypothesized that significant differences in inter-study populations might have accounted for the divergent phase II and phase III dexpramipexole results and tested for potential responder subgroups in EMPOWER based on the identification of those differences. We also used the primary endpoint and the analysis techniques prespecified in the EMPOWER protocol, we reported results using 95% confidence intervals, and we conducted statistical testing both with and without adjustments for multiplicity.

There are important limitations in conducting subgroup analyses. EMPOWER, like other randomized controlled trials, was neither designed nor powered to test for treatment differences within subgroups or heterogeneity across subgroups, nor is the optimal adjustment for multiple testing clear although the Bonferroni correction is generally considered to be a conservative approach. Applying or interpreting statistical tests must be placed in the context of the rationale supporting subgroup testing. Subgroup testing is indicated when potential heterogeneity of treatment effect may result if multiple pathologies underlie a clinical syndrome or when treatment effect may be related to disease severity or stage in the natural history of a disease (36). It is also important to consider the biological and clinical plausibility of the findings. EEC-definite ALS and short symptom duration at diagnosis are both associated with faster progressing disease (4–7) in which a drug
effect, if present, is likely to be more apparent. The plausibility of an additive effect of dexpramipexole with riluzole on clinical endpoints is supported by the additive effect on creatinine of the drugs in combination. The potential beneficial effects of dexpramipexole observed in the EMPOWER subgroup defined by riluzole use, EEC-definite ALS, and short symptom duration should therefore be viewed as hypothesis generating and requiring further investigation. Significance testing does not establish the validity of subgroup analysis; only replication does.

ALS remains an intractable disease. There is considerable discussion as to whether ALS is a single disease or a spectrum of disorders manifesting a common phenotype. Consensus is emerging, however, following failures such as EMPOWER, that smaller, targeted phase III trials are necessary to reduce participant heterogeneity and enrich study populations to more effectively detect drug responses. EMPOWER attempted to replicate the phase II dexpramipexole study, but in scaling by nearly a factor of 10, significant population differences emerged that may have accounted for the divergent results. Future studies of dexpramipexole should be conducted against a background of riluzole use, be enriched for participants with EEC-definite, rapidly-progressing ALS, and incorporate creatinine as a putative biomarker.

Acknowledgements

The phase II and phase III dexpramipexole studies were sponsored by Knopp Biosciences and Biogen Idec, respectively. Knopp Biosciences was involved in the design and conduct (collection/analysis of data) of the phase II dexpramipexole trial and Biogen Idec and Knopp Biosciences were involved in the design and conduct (collection/analysis of data) of the EMPOWER trial.

Declaration of interest: MEB, TVP, JM and DA are employees of Knopp Biosciences. HM was a phase II and EMPOWER study investigator and received grants from Avanir, Knopp Biosciences, Biogen Idec, and Cytokinetics for clinical trials and honoraria for participating in advisory board meetings from Avanir, Sanofi-Aventis, Shionogi, and Biogen Idec. He is now a member of the Data Safety Monitoring Board for the NeuralStem clinical trial. He received honoraria from Sanofi-Aventis Japan for giving seminars at the annual meetings of the Japanese Neurological Society in 2009 and 2010. He received a conference grant (to Columbia University) for the 2011 International ALS Conference from the National Institute of Neurological Disorders and Stroke, the National Institutes of Health’s Office of Rare Diseases Research, Muscular Dystrophy Association, ALS Association, ALS Society of Canada, Adams Foundation, Ride for Life, ALS Hope Foundation, Les Turner Foundation, Sanofi-Aventis, Biogen Idec, Knopp Biosciences, and Avania.

BB was a phase II and EMPOWER study investigator and a consultant to Knopp Biosciences. SAR was a phase II and EMPOWER study investigator and a consultant to Knopp Biosciences. DHM has been a consultant to Knopp Biosciences. BZ is an employee of MacroStat. AL was an EMPOWER investigator and received consulting fees and travel support from Biogen Idec and Teva Pharmaceuticals, consulting fees from Lundbeck, Knopp Biosciences, GlaxoSmithKline, and Boehringer Ingelheim, and speaker honoraria from Biogen Idec and Merz.

MEC was the principal investigator of this study and has received a grant from Biogen Idec for this role. In the past five years she has served as a consultant for Teva Pharmaceuticals, GlaxoSmithKline, Millennium, Synapse (DSMB), Ono (DSMB), Link Medicine, and Trophos (DSMB) and has served on an advisory board for Biogen Idec. LHvdB was a study investigator and has served on an advisory board for Biogen Idec and Cytokinetics and as a consultant for Baxter.

The authors alone are responsible for the content and writing of the paper.

References


