

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Analysing the influencing factors on caregivers' burden among ALS patients in China based on data mining

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-066402
Article Type:	Original research
Date Submitted by the Author:	13-Jul-2022
Complete List of Authors:	Lian, Ling; Sun Yat-Sen University, Zheng, Minying; Sun Yat-sen University First Affiliated Hospital He, Ruojie; Sun Yat-sen University First Affiliated Hospital Lin, Jianing; Sun Yat-sen University First Affiliated Hospital Chen, Weineng; Sun Yat-sen University First Affiliated Hospital Pei, Zhong; Sun Yat-sen University First Affiliated Hospital Yao, Xiaoli; Sun Yat-sen University First Affiliated Hospital
Keywords:	NEUROLOGY, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PSYCHIATRY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title page

Article

Analysing the influencing factors on caregivers' burden among ALS patients in China based on data mining

Ling Lian¹, Minying Zheng¹, Ruojie He¹, Jianing Lin¹, Weineng Chen¹, Zhong Pei¹, and Xiaoli Yao^{1,*}

¹Ling Lian, lianling@mail.sysu.edu.cn, ORCID: 0000-0002-6431-5154

¹Minying Zheng, zhengmy@mail.sysu.edu.cn

¹Ruojie He, herj@mail2.sysu.edu.cn

¹Jianing Lin, linjn7@mail2.sysu.edu.cn

¹Weineng Chen, chenwn5@mail2.sysu.edu.cn

¹Zhong Pei, peizhong@mail.sysu.edu.cn

(Ling Lian and Minying Zheng equally contributed to the paper)

¹Department of Neurology, The First Affiliated Hospital, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Diagnosis and Treatment of Major Neurological Diseases, National Key Clinical Department and Key Discipline of Neurology, No.58 Zhongshan Road 2, Guangzhou, 510080, China

*Correspondence: ¹Xiaoli Yao, yaoxiaol@mail.sysu.edu.cn; Tel: +8620-87755766-8290; Fax: +8620-87750632;

Abstract:

Objective: There is significant burden on caregivers of patients with amyotrophic lateral sclerosis (ALS). However, only a few studies have focused on caregivers, and traditional research methods have obvious shortcomings in dealing with multiple influencing factors. This study was designed to explore influencing factors on caregiver burden among ALS patients and their caregivers from a new perspective.

Methods: Fifty-seven patients with ALS and their caregivers were investigated to analyse the factors influencing caregiver burden through data mining. Statistical analysis was performed using SPSS 24.0, and LASSO regression model was established by python 3.8.1 to minimize the effect of multicollinearity.

Results: According to LASSO model, we found ten variables had weights. Among them, MITOS stage(0-1) had the highest weight (-12.235), followed by younger age group(-3.198), lower-educated group (2.136), fatigue(1.687), and social support(-0.45). Variables including sleep quality, anxiety, depression, and sex(man) had moderate weights in this model. Economic status(common), economic status(better), household(city), household(village), educational level(high), sex(female), age(older), MITOS stage(2-4), had a weight of zero.

Conclusions: Our study demonstrates that caregiver burden is positively correlated with the degree of fatigue and negatively correlated with social support. Low educational status could increase caregiver burden. Caregivers of ALS patients may suffer less from caregiver burden when the patients was younger and less severe. Hopefully, more attention should be paid to caregivers of ALS and effective interventions can be developed to relieve this burden.

Keywords: ALS; caregiver burden; LASSO; social support; MITOS; fatigue

Words: 3947

Strengths and limitations of this study

- ✧ This was the first time we study influencing factors on caregiver burden among ALS patients and their caregivers through data-mining based on LASSO regression.
- ✧ The research method solves the problem of small sample size and multicollinearity among independent variables.
- ✧ Different kinds of continuous variables and categorical variables were involved in the study.
- ✧ This was a study conducted in south-eastern China, which may only represent a part of the Chinese population.
- ✧ Our analysis did not include intervention effects with quantified coping strategies in a prospective study.

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease with progressive muscle weakness, dysarthria, dyspnoea, dysphagia, and other associated symptoms which prevent patients from taking care of themselves. Patients become increasingly dependent on their partners, family members, and other relatives for increasing support and assistance. For one thing, the high cost of medical care, drugs, and other supportive services exert huge pressure on families with ALS patients. Moreover, the influence on family in caregiving of patients with ALS is great[1; 2]. The role of caregivers is critical because their health and well-being are not only important for patients[3] [4]. Moreover, as the disease progresses, patients will suffer from worsening health and diminishing quality of life. Thus, as the disease progresses, the concomitant psychological distress and mental stress of informal caregivers of ALS increase significantly[5]. This phenomenon has also been observed in other neurodegenerative diseases[6] [7]. Kim et al. found a negative correlation between the time spent in caregiving and quality of life of caregivers[8]. In Burke's study, patients' cognitive and behavioural status is of great importance as mediators of caregiver burden[9]. While these studies have demonstrated factors that were associated with an increasing burden for caregivers, we only found a few studies that have focussed on influencing factors of caregivers, such as physical health, psychological symptoms, social background, social support, and the patients' condition, and most of them were designed as cross-sectional studies or case-control studies[10-15]. Also, knowledge about the action of interventions in relieving caregiver burden is still limited.

This study aimed to determine the effect of factors that influence caregivers' burden and physical and mental health through data mining based on LASSO regression model. Hopefully, effective interventions can be developed to improve the health status of caregivers and help relieve caregiver burden.

2. Materials and Methods

2.1 Resource of data and participants

Caregivers of patients registered at First affiliated hospital of Sun yat-sen University(FAH-SYSU) from August 2015 to February 2017 and with a patient diagnosis of ALS according to the revised El Escorial criteria were recruited in the study. As this retrospective study was an analysis of existing data, which did not involve patients' or caregivers' diagnosis and treatment, no ethics approval was required from the Ethics Committee of FAH-SYSU. Even though, participants enrolled have been fully informed and consent was obtained. Healthy caregivers with at least 3 months' caregiving experience were included in the study, and caregivers were restricted to relatives such as spouses, parents, children, and siblings of the patients. Informal, casual, or multiple caregivers and those with confirmed physical and mental problems were excluded. Patients who had rapid deterioration since diagnosis or with bulbar/spinal/respiratory breakdown were excluded from the sample.

This study primarily assessed the influencing factor of caregiver burden including age, gender, education level, economic status, anxiety, depression, social support, fatigue, sleep quality, as well as

the duration and stage of disease(MITOS stage). The quantitative questionnaires of standardised design assessed caregiver burden on different dimensions. Standardised scales were used for the caregivers, listed in the following section. In total, 60 caregivers of ALS patients were selected for the survey through internationally accepted questionnaires. Among all the 60 questionnaires sent out to the caregivers, 57 valid responses were received, with a valid rate of 95%.

2.1.1 Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting, or dissemination plans of this research.

2.2 Assessment measures and data collection

2.2.1 Patients' status

Disease duration was measured by months. Disease severity was mainly measured by the clinical staging system of Milano-Torinos (MITOS)^[16]. According to this staging system, patients with ALS are categorised as follows: Stage 0, functional involvement but no loss of independence in any domain; Stages 1–4, which refer to the number of domains in which independence was lost, for example, movement, swallowing, communicating, and/or breathing; and Stage 5, death.

2.2.2 Caregiver burden

Zarit Caregiver Burden Interview (ZBI) was used to evaluate caregiver burden. ZBI consists of a self-administered 22-item questionnaire, including caregivers' psychological health, finances, emotional wellbeing, social and family life, and degree of control over their life. Total scores range between 0 and 88.

2.2.3 Social support

The Social Support Rating Scale (SSRS) is a brief questionnaire designed by Xiao Shuiyuan in 1986-1993, which mainly reflects social support and satisfaction with social support from the respondents.

2.2.4 Physical and mental health of caregivers

The sleep quality of caregivers was assessed using the Pittsburgh Sleep Quality Index (PSQI). Fatigue scale-14 (FS-14) was used to measure the severity of physical and mental fatigue and its effect on a person's activities and lifestyle. Anxiety and depression status were evaluated using the Hamilton Anxiety Scale (HAMA) and Hamilton Depression Scale-17 (HAMD-17).

2.3 Statistics

2.3.1 General characteristics

Statistical analysis was performed using SPSS 24.0 and Python 3.8.1. Descriptive statistics were used for categorical variables. Means with a standard deviation(SD) and interquartile range(IQR) were used for continuous variables. Bivariate correlations between continuous variables were performed by Pearson correlation analysis. Correlations between caregiver burden and categorical variables were analysed by visualization analysis. All tests were performed with an alpha value of 0.05.

2.3.2 Least Absolute Shrinkage and Selection Operator (LASSO) regression

Least Absolute Shrinkage and Selection Operator (LASSO) regression was used to establish a model to predict the relationship between caregiver burden and variables with high multicollinearity. LASSO was a regression technique for variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model that it produces. It was a constrained version of ordinary least-squares (OLS) regression and typically used for regression

of a single response variable y on a predictor matrix X , which was widely used in dealing with data situations of low sample size and high dimension, especially when there were both continuous variables and categorical variables in the regression[17].

Compared with other linear regression, lasso was more applicable to analyse complex multicollinear data by adding a penalty equal to the absolute value of the importance of coefficients and minimizing insignificant coefficients to 0, thus the high weight variables will be selected. The absolute size of the regression coefficients β was constrained. The higher the penalty, the more regression coefficients were shrunk toward zero. The regularized regression is quite sensitive to the selection of the penalty coefficients. [18-20]

To appropriately tune the best parameter, the approach was to estimate the performance with different values using cross-validation[21]. Compared with other methods, LASSO was fast and accurate with the advantage of avoiding overfitting automatically.

2.3.3 Prediction Performance Evaluation

The mean square error (MSE) between the predicted value and the actual value was used to evaluate prediction performance. The coefficient of determination R^2 was used to reflect the regression fitting effect of the prediction model. The coefficient R^2 is defined as $(1 - u/v)$, where u is the regression sum of squares $((y_{\text{true}} - y_{\text{pred}})^2).sum()$ and v is the residual sum of squares $((y_{\text{true}} - y_{\text{true.mean}})^2).sum()$. Best possible score is 1.0, and lower values are worse. Lasso best(alpha) parameter was obtained by cross-validation (CV).

The algorithms were implemented using the Lasso CV package of sklearn in Python 3.8.1.

2.4 Data processing and dummy variables

In order to incorporate categorical variables and continuous variables into the LASSO regression model, each variable was processed according to the research purpose. Since continuous variable "age" followed a skewed distribution by calculation of skewness and kurtosis, age was categorized into two groups as categorical variables. Categorical variables were reclassified and served as dummy variables to make the regression result with a positively explanative sense.

Dummy coding method represents group membership with dummy variables that take on values 0 and 1, meaning that membership in a particular group is coded one whereas non-membership in the group is coded zero. When dummy coding is used in the regression analysis, the overall results indicate whether there is a relationship between the dummy variables and the dependent variables[22]. The values of the intercept and the regression coefficients of the resulted regression model can be obtained using least squares estimation procedures. The coefficient of multiple determination, R^2 , for the regression model with dummy variables can be interpreted in terms of the proportion of variance in the dependent variable that is accounted for by the categorical independent variable[23]. The regression model from the dummy coding can be written as:

$$Y_{ij} = B_0 + \sum_{j=1}^{k-1} B_j D_{ij} + \varepsilon_{ij}$$

Where:

Y_{ij} : The score on the dependent variable for subject i in group j .

B_0 : The intercept that represents the mean of the group coded 0 on all the dummy variables.

k : The number of categories of the independent variable.

B_j : The regression coefficient associated with the j th group, and it represents the difference between the mean of the group coded 1 on the corresponding dummy variable and the mean of the group coded 0 on all the dummy variables.

D_{ij} : The numerical value assigned to subject i in the j th group.

ε_{ij} : The error associated with the i th subject in the j th group.

To be specific, categorical variables were recoded with dummy variables, including 2 dummy variables suggesting age, 2 dummy variables suggesting MITOS stage, 2 dummy variables suggesting economic status, 2 dummy variables suggesting household, and 2 dummy variables suggesting educational level(**Table 1**). The principle of the setting of dummy variables was to equalize numbers between groups and accord with the meaning of each variable value. For

example, educational level (low) represented the sum of educational status including illiterate, primary school, middle school, and high school, while educational level (high) represented the sum of educational status including junior college, undergraduate, and master. Overall, 6 continuous variables, together with 12 dummy variables of 6 categorical variables were introduced into the model.

Table 1 The variable weight of the LASSO regression and description of dummy variables

Variable	The variable weight of LASSO regression	Description
<i>Continuous variables</i>		
Duration	0.077	
PSQI	0.078	
FS-14	1.687	
Anxiety	0.043	
Depression	0.004	
Social support	-0.455	
<i>Categorical variables</i>		
Sex(male)	-0.171	
Sex(female)	0	
Age(younger)	-3.198	<35
Age(older)	0	≥35
MITOS(stage 0-1)	-12.235	MITOS stage 0,1
MITOS(stage 2-4)	0	MITOS stage 2,3,4
Economic status(common)	0	poor, average
Economic status(better)	0	good, excellent
Household(village)	0	suburban, rural
Household(city)	0	city
Educational level(low)	2.136	illiterate primary school middle school high school technical secondary school
Educational level(high)	0	junior college undergraduate master

201
202

203 3. Results

204 3.1 Descriptive analysis of characteristics of caregivers and ALS patients

205 The majority of this cohort was two-thirds female caregivers (38, 66.7%), and the other one-
 206 third were male caregivers (19, 33.3%). The mean age of all the caregivers was 45.32±13.74
 207 years (IQR: 33.5-56.5), ranging from 22-72 years. The average disease course was 28.98±15.36
 208 months (interquartile range, IQR: 18-36), with the longest being 60 months and the shortest being 6
 209 months. The average score of caregiver burden scale was 39.28±17.11 (IQR: 25-55). The average
 210 scores of PSQI and FS-14 were 7.89±4.57 (IQR: 4-10), 6.58±3.60 (IQR: 4-9), respectively. The average
 211 HAMA scores of patients were 10.70±6.62 (IQR: 5-15). The average HAMD scores of patients were
 212 10.16±8.78 (IQR: 4-14.5). The mean score of the Social Support Rating Scale was 40.18±7.47 (IQR: 33.5-
 213 46). Forty caregivers were considered "anxious" to different degrees, and the number of possible,
 214 definite, obvious, and severe anxiety status caregivers were 20, 15, 4, 1, respectively. Similarly, 27
 215 caregivers were considered "depressed", and the number of possible, moderate, and severe
 216 depression status caregivers were 16, 10, 1, respectively. For categorical variables, the number of

217 patients with MITOS from zero to stage 4 was 10,16,20,4,7, respectively. Most of the caregivers were
 218 junior high school graduates (17, 29.8%). Overall, 34 participants (59.6%) had an education level of
 219 at least high school, of which two held a master's degree. Urban dwellers constituted a proportion
 220 of 56.1% of all the caregivers, while the rest of the individuals were from suburban or rural areas.
 221 Economic status varied from general to good condition among the participants (**Table 2**).

222 **Table 2** Demographic features of caregivers

	N	N%		N	N%
gender			MITOS		
male	19	33.3	stage 0	10	17.5
female	38	66.7	stage 1	16	28.1
education			stage 2	20	35.1
illiterate	1	1.8	stage 3	4	7.0
primary school	4	7.0	stage 4	7	12.3
middle school	17	29.8			
technical	1	1.8	economic		
secondary			state		
school			poor	5	8.8
high school	12	21.1	average	21	36.8
junior college	10	17.5	good	9	15.8
undergraduate	10	17.5	excellent	22	38.6
master	2	3.5			
househol			Age		
d register			<35	15	26.3
Rural	18	31.6	≥35	42	73.7
suburban	7	12.3			
city	32	56.1	total	57	100%

223

224 3.2 Analysis of correlations between continuous variables

225 Correlations between continuous variables were demonstrated with Heatmap Chart shown in
 226 **Figure 1**. According to Spearman correlation analysis, positive correlations were seen between
 227 caregiver burden and independent variables including duration of disease, PSQI scores, degree of
 228 fatigue, anxiety, and depression ($r=0.52, 0.54, 0.76, 0.71, 0.62, p<0.05$), while a negative correlation was
 229 noted between caregiver burden and social support ($r= -0.41, p<0.05$). Interestingly, a slight negative
 230 correlation was noticed between social support and PSQI scores, degree of fatigue, and depression ($r=$
 231 $-0.33, -0.29, -0.37, p<0.05$), indicating that social support could be a protective factor. Strong pairwise
 232 positive correlations were seen among PSQI scores, degree of fatigue, anxiety, and depression ($p<0.05$).

233 3.3 Analysis of correlations between caregiver burden and categorical variables

234 Correlations between caregiver burden and categorical variables were shown as visualizations
 235 (**Figure 2**). We used layered data processing to rearrange categorical variables and convert them
 236 into dummy variables. The principle of the setting of dummy variables was to equalize numbers
 237 between groups. The descriptions of each dummy variable were shown in **Table 1**. **Figure 2 (1)**
 238 showed that there was a difference in caregiver burden between younger and older patients. Also,
 239 an obvious disparity in caregiver burden was seen between group MITOS stage(0-1) and group
 240 MITOS stage(2-4), suggesting that patients with ALS with severer symptoms created higher
 241 caregiver burden (**Figure 2 (5)**).

242 3.4 Model construction and evaluation by Lasso regression

243 3.4.1 Variable Selection

The values of each categorical attribute were converted into dummy variables to build a regression model. With caregiver burden as the dependent variable and duration, PSQI, FS-14, anxiety, depression, social support, sex(male), sex(female), age(younger), age(older), MITOS stage (0-1), MITOS stage(2-4), economic status(common), economic status(better), household(village), household(city), educational level(low), educational level(high) as independent variables, a regression model with multiple dummy variables was constructed shown in **Table 1**.

In the LASSO model, we found ten variables had weights. Among them, MITOS stage(0-1) had the highest weight (-12.235), followed by younger age group(-3.198) , lower-educated group (2.136), fatigue(1.687), and social support(-0.45). Variables including sleep quality, anxiety, depression, and sex(man) had moderate weights in this model. Economic status(common), economic status(better) ,household(city), household(village), educational level(high), sex(female), age(older), MITOS stage(2-4), had a weight of zero (**Table 1**).

3.4.2 Performance evaluation

In this research, the mean squared error (MSE) was selected as the evaluation index, and the best Alpha value was obtained through cross-validation. The relationship between the model MSE and Alpha is shown in **Figure 3**. When the best penalty factor 'Alpha'= 0.3, the MSE is the smallest, and the LASSO regression model has the highest accuracy. To evaluate prediction performance and fitting effect, the R^2 and the MSE of the LASSO regression model were provided. The Lasso score(R^2) and the lasso MSE were 0.831 and 48.496, respectively.

3.5 Liner regression analysis between MITOS stage(0-1), FS-14, and caregiver burden

3.5.1 Variable selection

In the previous section, we learned from the LASSO regression model that MITOS stage(0-1) had the highest weight among categorical variables, while FS-14 had the highest weight among continuous variables. Moreover, the coefficient of MITOS stage(0-1) was significantly higher than other influential dummy variables, and the coefficient of FS-14 was even hundreds of times higher than other continuous variables. Therefore, to furtherly validated the action of the most influential factors in the model, a linear regression model was constructed with caregiver burden as the dependent variable, and FS-14, MITOS stage (0-1) as independent variables.

3.5.2 Performance evaluation

As was shown in **Figure 4**, caregiver burden was positively correlated with FS-14 whenever dummy variable MITOS stage(0-1) was kept in the model or not. On the other hand, when FS-14 remained constant, caregiver burden was significantly higher when dummy variable MITOS stage(0-1) was absent. The distance between the red line and the green line demonstrated the influence of MITOS stage(0-1) on caregiver burden in the model. The intercepts of the linear regression model was 32.125. The Lasso score(R^2) and the lasso MSE were 0.783 and 62.46, respectively. The results were consistent with LASSO regression, suggesting that MITOS stage(0-1) had a strong effect on caregiver burden.

4. Discussion

Based on the LASSO model, 10 variables had weights, whereas 5 variables had a greater influence on the model. MITOS stage(0-1) had the highest weight (-12.235), followed by younger age group(-3.198) , lower-educated group (2.136), fatigue(1.687), and social support(-0.45). According to the results mentioned above, MITOS stage of patients with ALS was the most important factor influencing caregiver burden, and caregiver burden was significantly lower in patients of ALS with a milder type (MITO stage 0-1). Interestingly, when another variable duration of disease was brought into the regression model, it didn't show any effect on the model, indicating that compared to duration, the severity of patients with ALS had a greater influence on caregiver burden. Sarah K. Bublitz found that thick mucus was a highly distressing symptom for both

1
2
3 292 patients and caregivers, which caused a higher degree of bulbar impairment and had a strong
4 293 impact on their quality of life[24].

6 294 We draw a conclusion from our study that caregiver burden was positively correlated with the
7 295 degree of fatigue, while the correlation between caregiver burden and psychological factors such as
8 296 anxiety and depression was relatively weak. The research of Brizzi K T et.al also showed that the
9 297 pressure of caregiving in ALS originated from physical symptoms consisting of fatigue, pain,
10 298 weakness, shortness of breath, difficulty sleeping, and speech problems, as well as depression and
11 299 other mood changes. It was worth noting that fatigue and weakness were the most common
12 300 symptoms and happened in over 75% of respondents [25], which was in accordance with our study.
13 301 In our LASSO model, fatigue had higher weights than anxiety, depression, and sleep quality.
14 302 Physical exhaustion was also a typical feature of caregiver burden of other progressive diseases,
15 303 such as Alzheimer's disease, Parkinson's disease, and cancer[26-29]. Not surprisingly, the feelings
16 304 of fatigue probably came from long-distance travel to the hospital for treatment, strenuous work in
17 305 caregiving, and loss of rest time. This could well explain our result that younger caregivers suffered
18 306 less from caregiver burden because they were healthier and more energetic, although they needed
19 307 to consider time-consuming costs and balance between work and caring. For those younger
20 308 caregivers, their caring patients were also younger as most of the caregivers were their spouses.
21 309 This could be another reason that younger caregivers experienced less caregiver burden.

24 310 In this study, we found that caregiver burden was negatively correlated with social support,
25 311 which was consistent with a study in Korea [30]. According to the epidemiology of ALS in China, we
26 312 have a potentially lower incidence and prevalence of ALS than in foreign countries [31]. It is still a
27 313 rare disease so doctors with less experience may fail to make a correct diagnosis in time and provide
28 314 appropriate treatment. Even if the diagnosis is clear, effective treatment is not always available to
29 315 every family due to the high cost of treating the illness. Fortunately, ALS is gradually being covered
30 316 by health insurance in many provinces of China. The guarantee of a new policy is of great importance
31 317 in helping relieve the economic burden on caregivers. In addition to economic support, more
32 318 attention was paid to this group because of increasing caring activities, media reports, and social
33 319 propaganda such as the "Ice Bucket Challenge" all over the country[32]. Multiple-center clinical
34 320 studies of potential drugs based on animal experiments have also been conducted by a great number
35 321 of Chinese research teams [33; 34], bringing hope and light to patients and caregivers of ALS.

37 322 On the other hand, according to our study, social support only had a modest action on caregiver
38 323 burden. This phenomenon might be caused by insufficient and unbalanced social support in China.
39 324 Therefore, we speculated that greater social support would bring about less caregiver burden. In
40 325 some well-developed regions such as Hong Kong, the promotion of multidisciplinary teams (MDT)
41 326 provides effective support for the disease, which encourages patients and families of ALS in a highly
42 327 positive manner[35]. Family care is the main strategy for Chinese caregivers to look after ALS
43 328 patients, and older women and spouses still play an important role in providing family care, while
44 329 in some other countries, hospital care, nursing home care, and community care were made full use
45 330 of to help alleviate the families' burden [36; 37]. The lack of professional caring experience, staff, and
46 331 facilities could affect the efficiency of caring and thus lead to an increasing caregiver burden.
47 332 Although China has developed rapidly in recent years, there have been obvious differences in
48 333 medical development and medical investments between eastern and western areas. Little research
49 334 on ALS has been conducted in western areas, where economic conditions are relatively poor and
50 335 social support is inadequate. In contrast, modern countries, such as Japan and other European
51 336 countries can offer stronger social aid and help systems to ALS families through long-term care
52 337 insurance, home visit rehabilitation, and peer group rehabilitation, suggesting that more social
53 338 support could be provided to Chinese ALS caregivers to relieve their burden [13; 37-40]. In addition,
54 339 medical expenses and end-of-life care were guaranteed in well-developed countries like Japan.
55 340 Obviously, there is an urgent need for more economical, practical, physical, emotional, and spiritual
56 341 support for caregivers of ALS.

58 342 It was surprising to find in our study that educational level influenced caregiver burden. In this
59 343 study, caregivers of ALS patients were mainly their spouses who were of the same age (50-70y) and
60 344 had similar educational levels (middle school) to the patients. A possible explanation is that

understanding abilities varied among different caregivers and could be a limitation in low-educated groups. Under the current situation in China, patients were not able to address all their questions in a short period because of the large hospital outpatient quotas. For those caregivers who had a lower educational level, they might feel difficult to quickly understand the key points in caring and follow doctors' suggestions. In addition, they possibly lacked the abilities to search for information and teach themselves to improve the efficiency of home nursing work.

We investigated the correlations between caregiver burden and caregivers' mental states, including anxiety, depression, sleep quality index, and fatigue. Although we observed a slight correlation between anxiety, depression, and caregiver burden, they didn't show high weights in the LASSO regression model as we expected. In contrast with other studies conducted in China, mental problems do exist comprehensively in caregivers of ALS, and anxiety and depression were associated closely with each other[10; 41]. However, the actual condition was more complicated. Firstly, anxiety was not only seen in caregivers of ALS but was also common in regular populations, especially women in their peri- or postmenopausal period [42]. Based on the results of an epidemiological study of ALS in the Chinese population, the ratio of male patients to female patients in China was about 1.45.1.98:1 [31], which is consistent with our findings. As there were more male patients than female patients in our study, it was speculated that caregivers were mostly women aged 50–70, who had possible risks of being anxious in any event. Traditionally, Chinese women of this age also face the burden of taking care of their grandchildren, making it harder to balance family and their feelings. Secondly, as there was obvious collinearity among anxiety, depression, and quality of sleep, their actual influence on caregiver burden may be overestimated. Even though, non-drug intervention seems to be available and effective for alleviating caregiver burden among caregivers of patients with neuropsychiatric disorders [43; 44]. Similarly, we also observed an evident decline in HAMA scores and HAMD scores after non-drug interventions on caregivers such as psychological counselling classes, professional online communication groups, organizing lectures, and home visits, though a well-designed prospective study should be done to confirm that furtherly.

Limitations

There were four limitations in this study. The sample size was small and respondents could experience remission, so our observational analysis could have been biased. In addition, we failed to assess intervention effects with quantified coping strategies in a prospective study. Another limitation of the study was the lack of information about the patients and the lack of correlations between patients' physical or psychological aspects and caregiver burden. Lastly, this was a study conducted in south-eastern China, and caregivers may only represent a part of the Chinese population.

5. Conclusion

In conclusion, our study demonstrates that caregiver burden is positively correlated with the degree of fatigue and negatively correlated with social support. Low educational status could increase caregiver burden. Caregivers of ALS patients may suffer less from caregiver burden when the patient of ALS was younger and less severe.

Figure 1 Correlations among caregiver burden and continuous variables

Figure 2 Correlations between caregiver burden and categorical variables

Figure 3 Relationship between α and MSE

Figure 4 Linear regression between caregiver burden and FS-14

393

Footnotes**Contributions**

396 Drafting of the manuscript and data analysis: LL. Data collection and follow patients: MYZ. Formal
397 analysis: RJH. Paper editing and submission: JNL. Validation: WNC. Direct guidance, paper
398 editing, and reviewing: ZP. Project designing and guidance: XLY.
399 All authors commented on previous versions of the manuscript and approved the final manuscript.
400

Funding

401 This study was supported by grants from the Guangdong Provincial Key Laboratory of Diagnosis
402 and Treatment of Major Neurological Diseases (2020B1212060017), Guangdong Provincial Clinical
403 Research Center for Neurological Diseases (2020B111170002), Southern China International Joint
404 Research Center for Early Intervention and Functional Rehabilitation of Neurological Diseases
405 (2015B050501003 and 2020A0505020004), Guangdong Provincial Engineering Center for Major
406 Neurological Disease Treatment, Guangdong Provincial Translational Medicine Innovation
407 Platform for Diagnosis and Treatment of Major Neurological Disease, Guangzhou Clinical Research
408 and Translational Center for Major Neurological Diseases (201604020010), The Young Scientist
409 Fund of National Natural Science Foundation of China, grant number 82001332, The National Key
410 Research and Development Program of China, grant number 2017YFC0907703.
411

Competing interests

412
413
414 None declared.
415

Provenance and peer review

416 Not commissioned; externally peer-reviewed.
417
418

Patient consent for publication

419 Not required.
420
421

Ethics Statement

422 As this retrospective study is an analysis of existing data, which does not involve patients' or
423 caregivers' diagnosis and treatment, no ethics approval is required from the Ethics Committee of
424 First Affiliated hospital of Sun yat-sen University. Even though, participants enrolled have been
425 fully informed and consent was obtained.
426

Data sharing statement

427
428
429 Data are available upon reasonable request. The data that support the findings of this study are
430 available from the corresponding author on reasonable request.
431

432
433 **References**

- 434 1. Geng, D., Ou, R., Miao, X., Zhao, L., Wei, Q., Chen, X., Liang, Y., Shang, H., & Yang, R.
435 (2017). Patients' self-perceived burden, caregivers' burden and quality of life for
436 amyotrophic lateral sclerosis patients: a cross-sectional study. *J Clin Nurs*, 26(19-20), 3188-
437 3199.
- 438 2. Weisser, F. B., Bristowe, K., & Jackson, D. (2015). Experiences of burden, needs, rewards
439 and resilience in family caregivers of people living with Motor Neurone
440 Disease/Amyotrophic Lateral Sclerosis: A secondary thematic analysis of qualitative
441 interviews. *Palliat Med*, 29(8), 737-745.
- 442 3. Boerner, K., & Mock, S. E. (2012). Impact of patient suffering on caregiver well-being: the
443 case of amyotrophic lateral sclerosis patients and their caregivers. *Psychol Health Med*, 17(4),
444 457-466.
- 445 4. Chio, A., Gauthier, A., Calvo, A., Ghiglione, P., & Mutani, R. (2005). Caregiver burden and
446 patients' perception of being a burden in ALS. *Neurology*, 64(10), 1780-1782.
- 447 5. Creemers, H., Veldink, J. H., Grupstra, H., Nollet, F., Beelen, A., & van den Berg, L. H.
448 (2014). Cluster RCT of case management on patients' quality of life and caregiver strain in
449 ALS. *Neurology*, 82(1), 23-31.
- 450 6. Yuksel, B., Ak, P. D., Sen, A., Sariahmetoglu, H., Uslu, S. C., & Atakli, D. (2018). Caregiver
451 burden and quality of life in early stages of idiopathic Parkinson's disease. *Idegyogy Sz*,
452 71(9-10), 343-350.
- 453 7. Isik, A. T., Soysal, P., Solmi, M., & Veronese, N. (2018). Bidirectional relationship between
454 caregiver burden and neuropsychiatric symptoms in patients with Alzheimer's disease: A
455 narrative review. *Int J Geriatr Psychiatry*.
- 456 8. Kim, M. S., Shin, H. I., Min, Y., Kim, J. Y., & Kim, J. S. (2011). Correlation between Severe
457 ALS Patient-Caregiver Couples' Characteristics and Caregivers' Health Related Quality of
458 Life. *Journal of Korean Academy of Nursing*, 41(3), 354-363.
- 459 9. Burke, T., Elamin, M., Galvin, M., Hardiman, O., & Pender, N. (2015). Caregiver burden in
460 amyotrophic lateral sclerosis: a cross-sectional investigation of predictors. *J Neurol*, 262(6),
461 1526-1532.
- 462 10. Tang, S., Li, L., Xue, H., Cao, S., Li, C., Han, K., & Wang, B. (2021). Caregiver burden and
463 associated factors among primary caregivers of patients with ALS in home care: a cross-
464 sectional survey study. *BMJ Open*, 11(9), e050185.
- 465 11. Siciliano, M., Santangelo, G., Trojsi, F., Di Somma, C., Patrone, M., Femiano, C., Monsurro,
466 M. R., Trojano, L., & Tedeschi, G. (2017). Coping strategies and psychological distress in
467 caregivers of patients with Amyotrophic Lateral Sclerosis (ALS). *Amyotroph Lateral Scler*
468 *Frontotemporal Degener*, 18(5-6), 367-377.
- 469 12. de Wit, J., Schroder, C. D., El Mecky, J., Beelen, A., van den Berg, L. H., & Visser-Meily, J.
470 M. A. (2018). Support needs of caregivers of patients with amyotrophic lateral sclerosis: A
471 qualitative study. *Palliat Support Care*, 1-7.
- 472 13. Cipolletta, S., Gammino, G. R., Francescon, P., & Palmieri, A. (2018). Mutual support
473 groups for family caregivers of people with amyotrophic lateral sclerosis in Italy: A pilot
474 study. *Health Soc Care Community*, 26(4), 556-563.

- 1
2
3 475 14. Chu, H. S., Tak, Y. R., & Kim, S. H. (2018). Factors Influencing Psychosocial Well-Being in
4 476 Family Caregivers of People with Amyotrophic Lateral Sclerosis. *Journal of Korean Academy*
5 477 *of Nursing*, 48(4), 454-464.
- 6 478 15. Oh, J., & Kim, J. A. (2018). Factor analysis of the Zarit Burden Interview in family caregivers
7 479 of patients with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal*
8 480 *Degener*, 19(1-2), 50-56.
- 9 481 16. Chio, A., Hammond, E. R., Mora, G., Bonito, V., & Filippini, G. (2015). Development and
10 482 evaluation of a clinical staging system for amyotrophic lateral sclerosis. *Journal of Neurology*
11 483 *Neurosurgery and Psychiatry*, 86(1), 38-44.
- 12 484 17. Zhu, X. W., Xin, Y. J., & Ge, H. L. (2015). Recursive Random Forests Enable Better Predictive
13 485 Performance and Model Interpretation than Variable Selection by LASSO. *Journal of*
14 486 *Chemical Information and Modeling*, 55(4), 736-746.
- 15 487 18. Sun, C., Zhu, B., Zhu, S., Zhang, L., Du, X., & Tan, X. (2021). Risk Factors Analysis of Bone
16 488 Mineral Density Based on Lasso and Quantile Regression in America during 2015-2018. *Int*
17 489 *J Environ Res Public Health*, 19(1).
- 18 490 19. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal*
19 491 *Statistical Society*, 58(1), 267-288.
- 20 492 20. Rajaratnam, B., Roberts, S., Sparks, D., & Dalal, O. (2015). Lasso regression: estimation and
21 493 shrinkage via the limit of Gibbs sampling. *Journal of the Royal Statistical Society*, 78, 153-174.
- 22 494 21. Hassan, S. S., Farhan, M., Mangayil, R., Huttunen, H., & Aho, T. (2013). Bioprocess data
23 495 mining using regularized regression and random forests. *Bmc Systems Biology*, 7.
- 24 496 22. Myers, J. L., & Well, A. (2003). Research design and statistical analysis (2nd ed.). Mahwah,
25 497 N.J.: Lawrence Erlbaum Associates.
- 26 498 23. Alkharusi, H. (2012). Categorical Variables in Regression Analysis: A Comparison of
27 499 Dummy and Effect Coding. *International Journal of Education*, 4(2).
- 28 500 24. Bublitz, S. K., Mie, E., Wasner, M., Hapfelmeier, A., Geiseler, J., Lorenzl, S., & Winkler, A. S.
29 501 (2022). Thick Mucus in ALS: A Mixed-Method Study on Associated Factors and Its Impact
30 502 on Quality of Life of Patients and Caregivers. *Brain Sci*, 12(2).
- 31 503 25. Brizzi, K. T., Bridges, J. F. P., Yersak, J., Balas, C., Thakur, N., Galvin, M., Hardiman, O.,
32 504 Heatwole, C., Ravits, J., Simmons, Z., Bruijn, L., Chan, J., Bedlack, R., & Berry, J. D. (2020).
33 505 Understanding the needs of people with ALS: a national survey of patients and caregivers.
34 506 *Amyotroph Lateral Scler Frontotemporal Degener*, 21(5-6), 355-363.
- 35 507 26. Madaleno, T. R., Moriguti, J. C., Ferriolli, E., De Carlo, M., & Lima, N. K. C. (2019). Mood,
36 508 lifestyle and cardiovascular risk factors among older caregivers of patients with
37 509 Alzheimer's disease dementia: a case-control study. *Aging Clin Exp Res*.
- 38 510 27. Liu, S., Li, C., Shi, Z., Wang, X., Zhou, Y., Liu, S., Liu, J., Yu, T., & Ji, Y. (2017). Caregiver
39 511 burden and prevalence of depression, anxiety and sleep disturbances in Alzheimer's
40 512 disease caregivers in China. *J Clin Nurs*, 26(9-10), 1291-1300.
- 41 513 28. Torný, F., Videaud, H., Chatainier, P., Tarrade, C., Meissner, W. G., & Couratier, P. (2018).
42 514 Factors associated with spousal burden in Parkinson's disease. *Rev Neurol (Paris)*, 174(10),
43 515 711-715.
- 44 516 29. Molassiotis, A., Zheng, Y., Denton-Cardew, L., Swindell, R., & Brunton, L. (2010).
45 517 Symptoms experienced by cancer patients during the first year from diagnosis: patient and

- 1
2
3 518 informal caregiver ratings and agreement. *Palliat Support Care*, 8(3), 313-324.
- 4 519 30. Chu, H. S., Tak, Y. R., & Kim, S. H. (2018). [Factors Influencing Psychosocial Well-Being in
5 520 Family Caregivers of People with Amyotrophic Lateral Sclerosis]. *J Korean Acad Nurs*, 48(4),
6 521 454-464.
- 7 522 31. Liu, X., He, J., Gao, F. B., Gitler, A. D., & Fan, D. (2018). The epidemiology and genetics of
8 523 Amyotrophic lateral sclerosis in China. *Brain Res*, 1693(Pt A), 121-126.
- 9 524 32. Kuehn, B. M. (2018). Simple Models and Ice Bucket Challenge Fuel Progress in ALS
10 525 Treatment. *JAMA*, 319(6), 535-537.
- 11 526 33. Feng, X. H., Yuan, W., Peng, Y., Liu, M. S., & Cui, L. Y. (2012). Therapeutic effects of dl-3-n-
12 527 butylphthalide in a transgenic mouse model of amyotrophic lateral sclerosis. *Chin Med J*
13 528 (*Engl*), 125(10), 1760-1766.
- 14 529 34. Huang, C., Li, J., Zhang, G., Lin, Y., Li, C., Zheng, X., Song, X., Han, B., Guo, B., Tu, Z.,
15 530 Zhang, J., Sun, Y., Wang, Y., Zhang, Z., & Yan, S. (2021). TBN improves motor function and
16 531 prolongs survival in a TDP-43M337V mouse model of ALS. *Hum Mol Genet*, 30(16), 1484-
17 532 1496.
- 18 533 35. Cheng, H. W. B., Chan, K. Y., Chung, Y. K. J., Choi, C. W., Chan, C. H., Cheng, S. C., Chan,
19 534 W. H., Fung, K. S., Wong, K. Y., Chan, O. M. I., & Man, C. W. (2018). Supportive &
20 535 palliative interventions in motor neurone disease: what we know from current literature?
21 536 *Annals of Palliative Medicine*, 7(3), 320-331.
- 22 537 36. Levitsky, G. N., Chub, R. V., & Kryachkov, A. V. (2017). [The quality of medical care and
23 538 cooperation of patients with amyotrophic lateral sclerosis in the Russian Federation]. *Zh*
24 539 *Nevrol Psikhiatr Im S S Korsakova*, 117(1. Vyp. 2), 59-63.
- 25 540 37. Hamaguchi, Y., Hakuno, A., Saito, K., & Wakabayashi, H. (2019). [Home Visit
26 541 Rehabilitation for a Patient with ALS to Improve Social Participation]. *Gan To Kagaku Ryoho*,
27 542 46(Suppl 1), 46-48.
- 28 543 38. Arai, Y., & Zarit, S. H. (2011). Exploring strategies to alleviate caregiver burden: effects of
29 544 the National Long-Term Care insurance scheme in Japan. *Psychogeriatrics*, 11(3), 183-189.
- 30 545 39. Madsen, L. S., Jeppesen, J., & Handberg, C. (2019). "Understanding my ALS". Experiences
31 546 and reflections of persons with amyotrophic lateral sclerosis and relatives on participation
32 547 in peer group rehabilitation. *Disabil Rehabil*, 41(12), 1410-1418.
- 33 548 40. Ogino, M. (2008). [End of life care for patients with ALS in Japan]. *Rinsho Shinkeigaku*,
34 549 48(11), 973-975.
- 35 550 41. Chen, D., Guo, X., Zheng, Z., Wei, Q., Song, W., Cao, B., Huang, R., Yang, R., & Shang, H.
36 551 (2015). Depression and anxiety in amyotrophic lateral sclerosis: correlations between the
37 552 distress of patients and caregivers. *Muscle Nerve*, 51(3), 353-357.
- 38 553 42. Mulhall, S., Andel, R., & Anstey, K. J. (2018). Variation in symptoms of depression and
39 554 anxiety in midlife women by menopausal status. *Maturitas*, 108, 7-12.
- 40 555 43. Chen, H. M., Huang, M. F., Yeh, Y. C., Huang, W. H., & Chen, C. S. (2015). Effectiveness of
41 556 coping strategies intervention on caregiver burden among caregivers of elderly patients
42 557 with dementia. *Psychogeriatrics*, 15(1), 20-25.
- 43 558 44. Martin-Carrasco, M., Fernandez-Catalina, P., Dominguez-Panchon, A. I., Goncalves-
44 559 Pereira, M., Gonzalez-Fraile, E., Munoz-Hermoso, P., Ballesteros, J., & group, E.-I. (2016). A
45 560 randomized trial to assess the efficacy of a psychoeducational intervention on caregiver

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

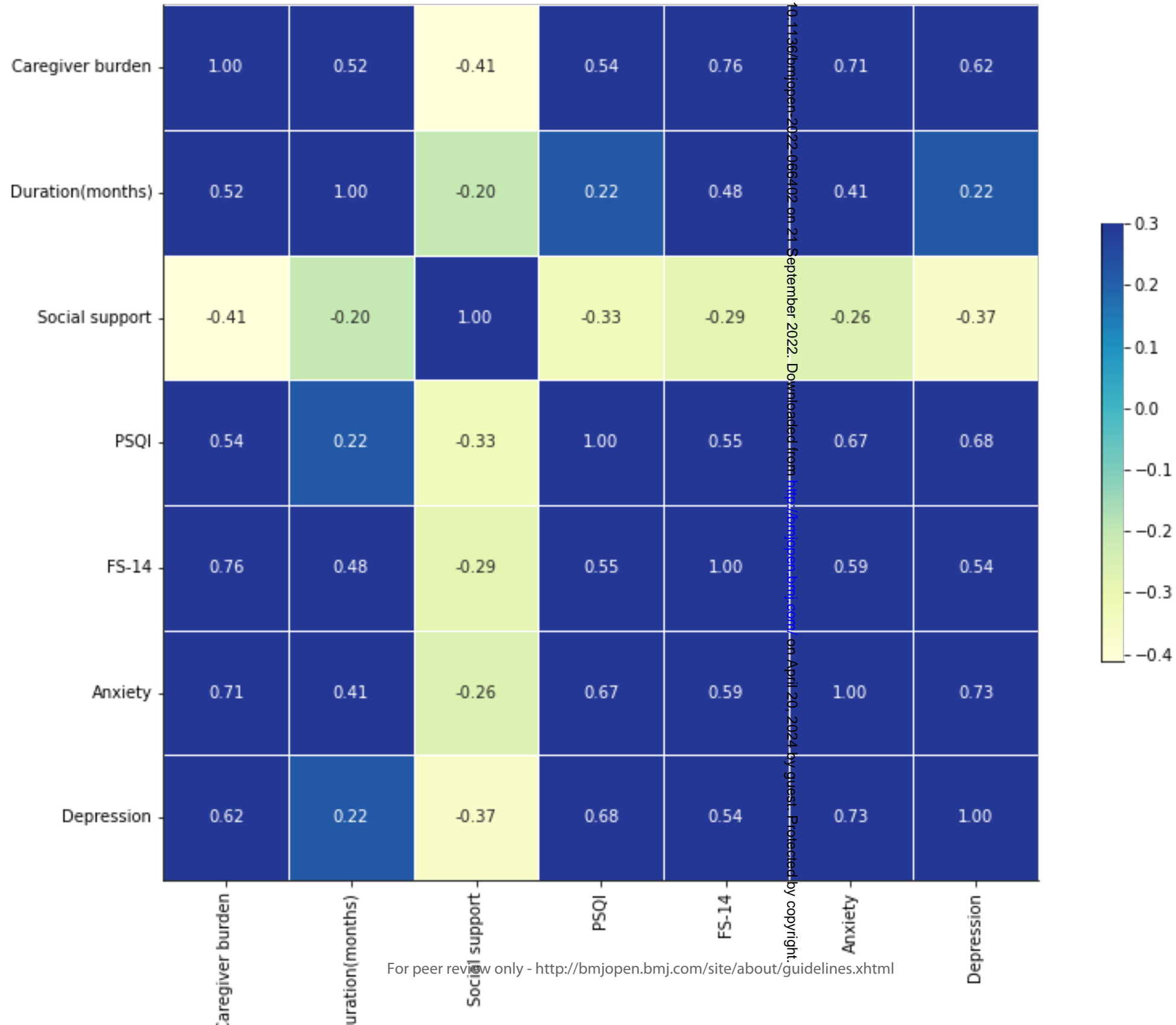
561 burden in schizophrenia. *Eur Psychiatry*, 33, 9-17.

562
563
564
565
566
567
568
569

Acknowledgments

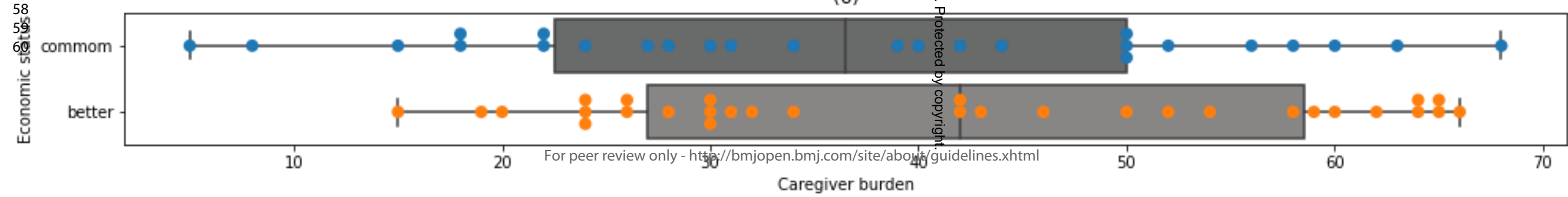
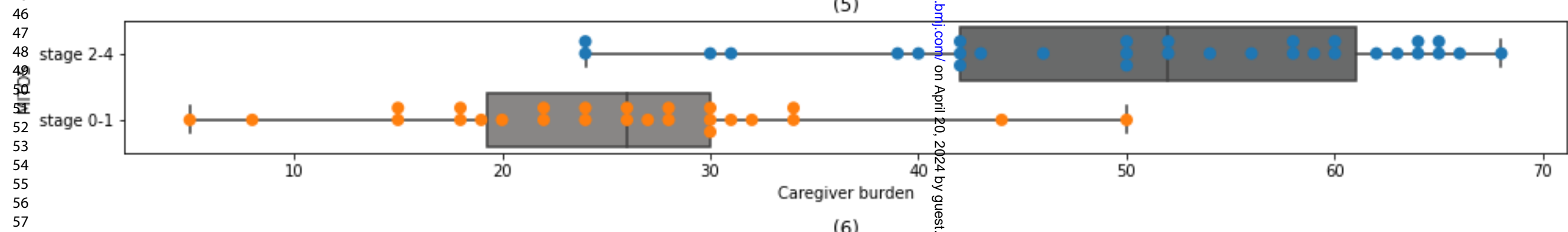
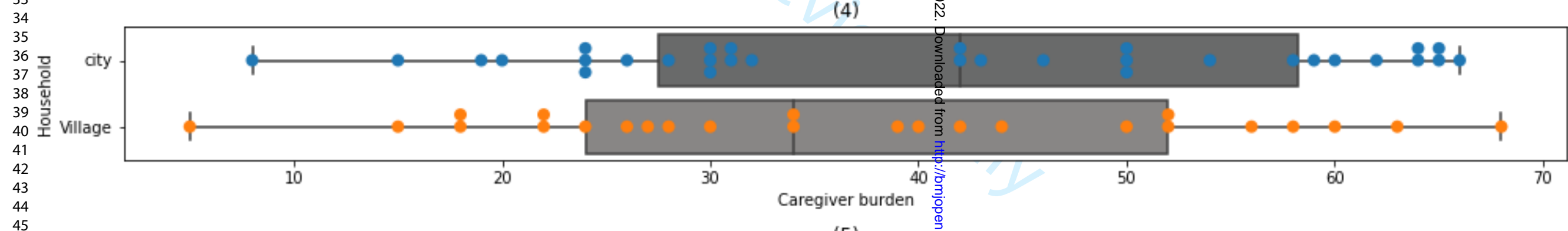
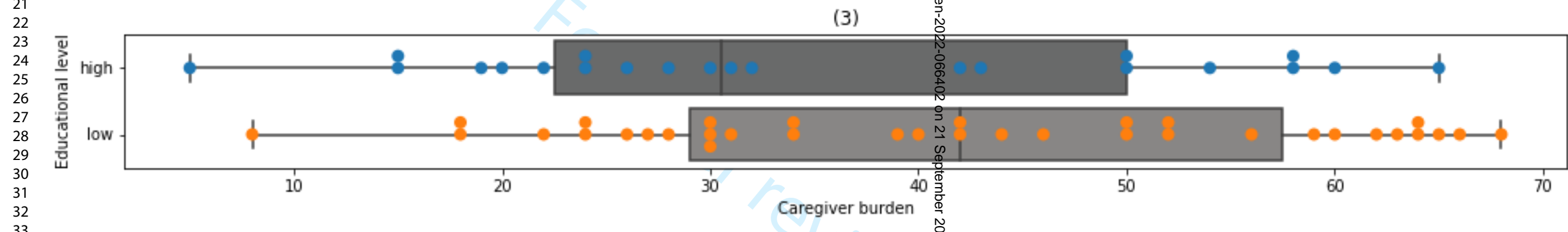
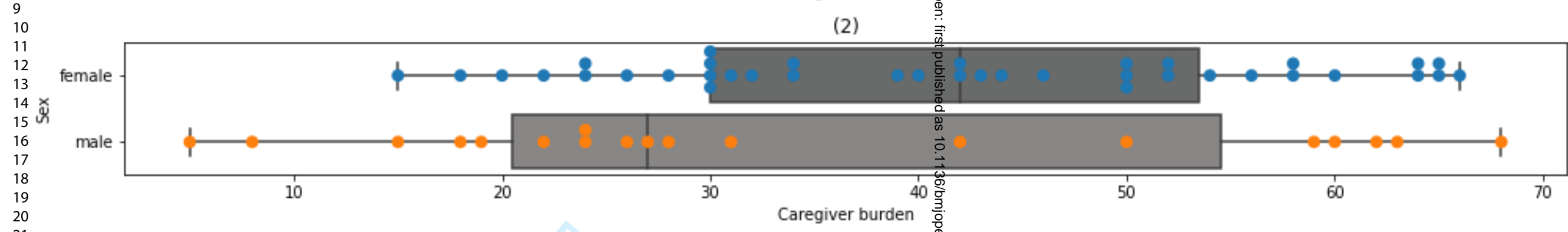
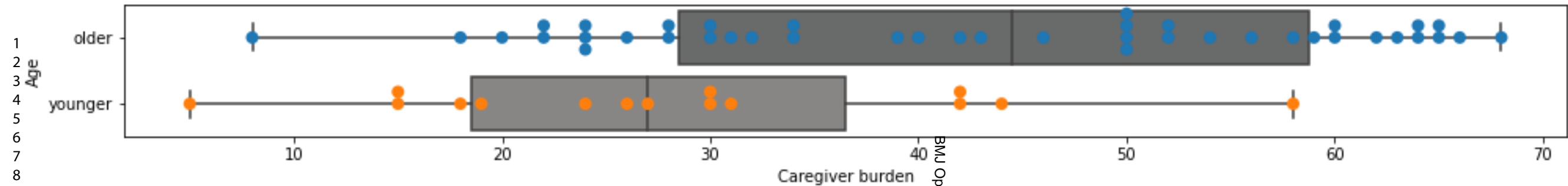
We thank all patients and their caregivers of ALS for taking part in this research. We gratefully acknowledge the contribution of the doctors, nurses, and other colleagues during the study.

For peer review only

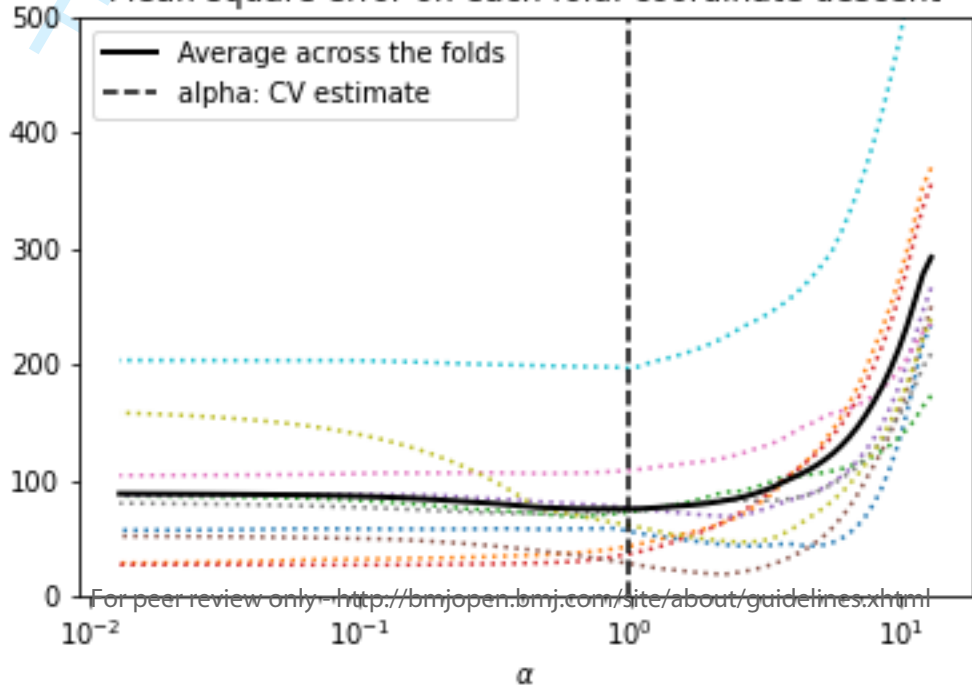


en: first published as 10.1136/bmjopen-2022-006402 on 21 September 2022. Downloaded from <http://bmjopen.bmj.com/> on April 20, 2024 by guest. Protected by copyright.

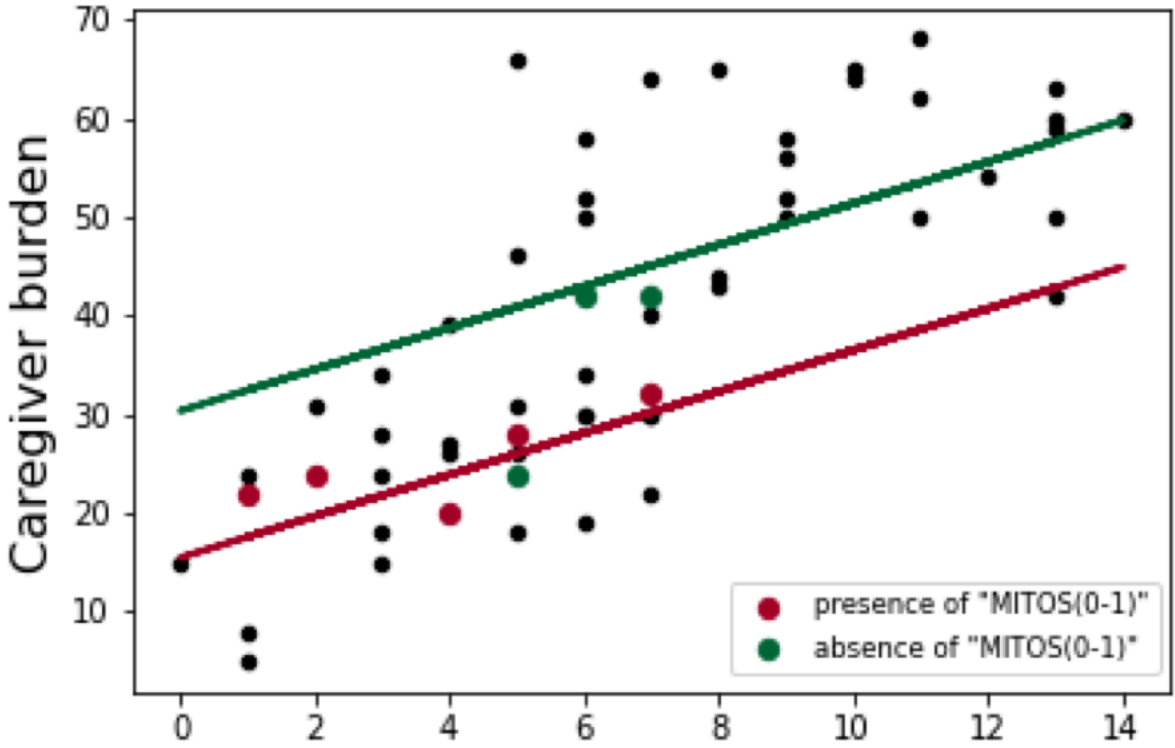
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56



Mean square error on each fold: coordinate descent

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28



FS-14

BMJ Open

Analysing the influencing factors on caregivers' burden among amyotrophic lateral sclerosis patients in China: a cross-sectional study based on data mining

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-066402.R1
Article Type:	Original research
Date Submitted by the Author:	30-Aug-2022
Complete List of Authors:	Lian, Ling; Sun Yat-Sen University, Zheng, Minying; Sun Yat-sen University First Affiliated Hospital He, Ruojie; Sun Yat-sen University First Affiliated Hospital Lin, Jianing; Sun Yat-sen University First Affiliated Hospital Chen, Weineng; Sun Yat-sen University First Affiliated Hospital Pei, Zhong; Sun Yat-sen University First Affiliated Hospital Yao, Xiaoli; Sun Yat-sen University First Affiliated Hospital
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Mental health, Palliative care
Keywords:	NEUROLOGY, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PSYCHIATRY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title page

Article

Analysing the influencing factors on caregivers' burden among amyotrophic lateral sclerosis patients in China: a cross-sectional study based on data mining

Ling Lian¹, Minying Zheng¹, Ruojie He¹, Jianing Lin¹, Weineng Chen¹, Zhong Pei¹, and Xiaoli Yao^{1,*}

¹Ling Lian, lianling@mail.sysu.edu.cn, ORCID: 0000-0002-6431-5154

¹Minying Zheng, zhengmy@mail.sysu.edu.cn

¹Ruojie He, herj@mail2.sysu.edu.cn

¹Jianing Lin, linjn7@mail2.sysu.edu.cn

¹Weineng Chen, chenwn5@mail2.sysu.edu.cn

¹Zhong Pei, peizhong@mail.sysu.edu.cn

(Ling Lian and Minying Zheng equally contributed to the paper)

¹Department of Neurology, The First Affiliated Hospital, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Diagnosis and Treatment of Major Neurological Diseases, National Key Clinical Department and Key Discipline of Neurology, No.58 Zhongshan Road 2, Guangzhou, 510080, China

*Correspondence: ¹Xiaoli Yao, yaoxiaol@mail.sysu.edu.cn; Tel: +8620-87755766-8290; Fax: +8620-87750632;

Abstract:

Objectives: There is significant burden on caregivers of patients with amyotrophic lateral sclerosis (ALS). However, only a few studies have focused on caregivers, and traditional research methods have obvious shortcomings in dealing with multiple influencing factors. This study was designed to explore influencing factors on caregiver burden among ALS patients and their caregivers from a new perspective.

Design: Cross-sectional study.

Setting: The data were collected at an affiliated hospital in Guangzhou, Guangdong, China.

Participants: Fifty-seven pairs of patients with ALS and their caregivers were investigated by standardised questionnaires.

Main outcome measures: This study primarily assessed the influencing factor of caregiver burden including age, gender, education level, economic status, anxiety, depression, social support, fatigue, sleep quality and stage of disease through data mining. Statistical analysis was performed using SPSS 24.0, and LASSO regression model was established by python 3.8.1 to minimize the effect of multicollinearity.

Results: According to LASSO model, we found ten variables had weights. Among them, MITOS stage(0-1) had the highest weight (-12.235), followed by younger age group(-3.198), lower-educated group (2.136), fatigue(1.687), and social support(-0.45). Variables including sleep quality, anxiety, depression, and sex(man) had moderate weights in this model. Economic status(common), economic status(better), household(city), household(village), educational level(high), sex(female), age(older), MITOS stage(2-4), had a weight of zero.

Conclusions: Our study demonstrates that the severity of ALS patients is the most influencing factor in caregiver burden. Caregivers of ALS patients may suffer less from caregiver burden when the patients are less severe and the caregivers are younger. Low educational status could increase caregiver burden. Caregiver burden is positively correlated with the degree of fatigue and negatively

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

46 correlated with social support. Hopefully, more attention should be paid to caregivers of ALS and
47 effective interventions can be developed to relieve this burden.

48 **Keywords:** ALS; caregiver burden; LASSO; social support; MITOS; fatigue

49
50 **Words:** 4089
51

For peer review only

Strengths and limitations of this study

- ✧ This was the first time we study influencing factors on caregiver burden among ALS patients and their caregivers through data-mining based on LASSO regression.
- ✧ The research method solves the problem of small sample size and multicollinearity among independent variables.
- ✧ Different kinds of continuous variables and categorical variables were involved in the study.
- ✧ This was a study conducted in south-eastern China, which may only represent a part of the Chinese population.
- ✧ Our analysis did not include intervention effects with quantified coping strategies in a prospective study.

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease with progressive muscle weakness, dysarthria, dyspnoea, dysphagia, and other associated symptoms which prevent patients from taking care of themselves. Patients become increasingly dependent on their partners, family members, and other relatives for increasing support and assistance. For one thing, the high cost of medical care, drugs, and other supportive services exert huge pressure on families with ALS patients. Moreover, the influence on family in caregiving of patients with ALS is great^{1,2}. The role of caregivers is critical because their health and well-being are not only important for patients^{3,4}. Moreover, as the disease progresses, patients will suffer from worsening health and diminishing quality of life. Thus, as the disease progresses, the concomitant psychological distress and mental stress of informal caregivers of ALS increase significantly⁵. This phenomenon has also been observed in other neurodegenerative diseases^{6,7}. Kim et al. found a negative correlation between the time spent in caregiving and quality of life of caregivers⁸. In Burke's study, patients' cognitive and behavioural status is of great importance as mediators of caregiver burden⁹. While these studies have demonstrated factors that were associated with an increasing burden for caregivers, we only found a few studies that have focussed on influencing factors of caregivers, such as physical health, psychological symptoms, social background, social support, and the patients' condition, and most of them were designed as cross-sectional studies or case-control studies¹⁰⁻¹⁵. Also, knowledge about the action of interventions in relieving caregiver burden is still limited.

This study aimed to determine the effect of factors that influence caregivers' burden and physical and mental health through data mining based on LASSO regression model. Hopefully, effective interventions can be developed to improve the health status of caregivers and help relieve caregiver burden.

2. Materials and Methods

2.1 Resource of data and participants

Caregivers of patients registered at First affiliated hospital of Sun yat-sen University (FAH-SYSU) from August 2015 to February 2017 and with a patient diagnosis of ALS according to the revised El Escorial criteria were recruited in the study. As this cross-sectional study was an analysis of existing data, which did not involve patients' or caregivers' diagnosis and treatment, no ethics approval was required from the Ethics Committee of FAH-SYSU. Even though, participants enrolled have been fully informed and consent was obtained. The informed consent claimed that the both the patients and their caregivers knew and agreed that their medical records was used in the study anonymously. Healthy caregivers with at least 3 months' caregiving experience were included in the study, and caregivers were restricted to relatives such as spouses, parents, children, and siblings of the patients. Informal, casual, or multiple caregivers and those with confirmed physical and mental problems were excluded. Patients who had rapid deterioration since diagnosis or with bulbar/spinal/respiratory breakdown were excluded from the sample.

This study primarily assessed the influencing factor of caregiver burden including age, gender, education level, economic status, anxiety, depression, social support, fatigue, sleep quality, as well as the duration and stage of disease (MITOS stage). The quantitative questionnaires of standardised design assessed caregiver burden on different dimensions. Standardised scales were used for the caregivers, listed in the following section. In total, 60 caregivers of ALS patients were selected for the survey through internationally accepted questionnaires. Among all the 60 questionnaires sent out to the caregivers, 57 valid responses were received, with a valid rate of 95%.

2.1.1 Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting, or dissemination plans of this research.

2.2 Assessment measures and data collection

2.2.1 Patients' status

Disease duration was measured by months. Disease severity was mainly measured by the clinical staging system of Milano-Torinos (MITOS)¹⁶. According to this staging system, patients with ALS are categorised as follows: Stage 0, functional involvement but no loss of independence in any domain; Stages 1–4, which refer to the number of domains in which independence was lost, for example, movement, swallowing, communicating, and/or breathing; and Stage 5, death.

2.2.2 Caregiver burden

Zarit Caregiver Burden Interview (ZBI) was used to evaluate caregiver burden^{17 18}. ZBI consists of a self-administered 22-item questionnaire, including caregivers' psychological health, finances, emotional wellbeing, social and family life, and degree of control over their life. Total scores range between 0 and 88.

2.2.3 Social support

The Social Support Rating Scale (SSRS) is a brief questionnaire designed by Xiao Shuiyuan in 1986-1993, which mainly reflects social support and satisfaction with social support from the respondents¹⁹.

2.2.4 Physical and mental health of caregivers

The sleep quality of caregivers was assessed using the Pittsburgh Sleep Quality Index (PSQI)¹⁹. Fatigue scale-14 (FS-14) was used to measure the severity of physical and mental fatigue and its effect on a person's activities and lifestyle²⁰. Anxiety and depression status were evaluated using the Hamilton Anxiety Scale (HAMA) and Hamilton Depression Scale-17 (HAMD-17)^{21 22}.

2.3 Statistics

2.3.1 General characteristics

Statistical analysis was performed using SPSS 24.0 and Python 3.8.1. Descriptive statistics were used for categorical variables. Means with a standard deviation (SD) and interquartile range (IQR) were used for continuous variables. Bivariate correlations between continuous variables were performed by Pearson correlation analysis. Correlations between caregiver burden and categorical variables were analysed by visualization analysis. All tests were performed with an alpha value of 0.05.

2.3.2 Least Absolute Shrinkage and Selection Operator (LASSO) regression

Least Absolute Shrinkage and Selection Operator (LASSO) regression was used to establish a model to predict the relationship between caregiver burden and variables with high multicollinearity. LASSO was a regression technique for variable selection and regularization to

147 enhance the prediction accuracy and interpretability of the statistical model that it produces. It was
 148 a constrained version of ordinary least-squares (OLS) regression and typically used for regression
 149 of a single response variable y on a predictor matrix X , which was widely used in dealing with data
 150 situations of low sample size and high dimension, especially when there were both continuous
 151 variables and categorical variables in the regression²³.

152 Compared with other linear regression, lasso was more applicable to analyse complex
 153 multicollinear data by adding a penalty equal to the absolute value of the importance of coefficients
 154 and minimizing insignificant coefficients to 0, thus the high weight variables will be selected. The
 155 absolute size of the regression coefficients β was constrained. The higher the penalty, the more
 156 regression coefficients were shrunk toward zero. The regularized regression is quite sensitive to the
 157 selection of the penalty coefficients.²⁴⁻²⁶

158 To appropriately tune the best parameter, the approach was to estimate the performance with
 159 different values using cross-validation²⁷. Compared with other methods, LASSO was fast and
 160 accurate with the advantage of avoiding overfitting automatically.

161

162 2.3.3 Prediction Performance Evaluation

163 The mean square error (MSE) between the predicted value and the actual value was used to
 164 evaluate prediction performance. The coefficient of determination R^2 was used to reflect the
 165 regression fitting effect of the prediction model. The coefficient R^2 is defined as $(1 - u/v)$, where u is
 166 the regression sum of squares $((y_{\text{true}} - y_{\text{pred}}) ** 2).sum()$ and v is the residual sum of squares
 167 $((y_{\text{true}} - y_{\text{true}.mean()}) ** 2).sum()$. Best possible score is 1.0, and lower values are worse. Lasso
 168 best(alpha) parameter was obtained by cross-validation (CV).

169 The algorithms were implemented using the Lasso CV package of sklearn in Python 3.8.1.

170

171 2.4 Data processing and dummy variables

172 In order to incorporate categorical variables and continuous variables into the LASSO
 173 regression model, each variable was processed according to the research purpose. Since continuous
 174 variable "age" followed a skewed distribution by calculation of skewness and kurtosis, age was
 175 categorized into two groups as categorical variables. Categorical variables were reclassified and
 176 served as dummy variables to make the regression result with a positively explanative sense.

177 Dummy coding method represents group membership with dummy variables that take on
 178 values 0 and 1, meaning that membership in a particular group is coded one whereas non-
 179 membership in the group is coded zero. When dummy coding is used in the regression analysis,
 180 the overall results indicate whether there is a relationship between the dummy variables and the
 181 dependent variables²⁸. The values of the intercept and the regression coefficients of the resulted
 182 regression model can be obtained using least squares estimation procedures. The coefficient of
 183 multiple determination, R^2 , for the regression model with dummy variables can be interpreted in
 184 terms of the proportion of variance in the dependent variable that is accounted for by the
 185 categorical independent variable²⁹. The regression model from the dummy coding can be written
 186 as:

187

$$Y_{ij} = B_0 + \sum_{j=1}^{k-1} B_j D_{ij} + \varepsilon_{ij}$$

188

189 Where:

190

191 Y_{ij} : The score on the dependent variable for subject i in group j .

192

193 B_0 : The intercept that represents the mean of the group coded 0 on all the dummy variables.

194

195 k : The number of categories of the independent variable.

196

197 B_j : The regression coefficient associated with the j th group, and it represents the difference
 198 between the mean of the group coded 1 on the corresponding dummy variable and the mean of the
 199 group coded 0 on all the dummy variables.

200

201 D_{ij} : The numerical value assigned to subject i in the j th group.

202

203 ε_{ij} : The error associated with the i th subject in the j th group.

204

205 To be specific, categorical variables were recoded with dummy variables, including 2 dummy
 206 variables suggesting age, 2 dummy variables suggesting MITOS stage, 2 dummy variables
 207 suggesting economic status, 2 dummy variables suggesting household, and 2 dummy variables

208

suggesting educational level(**Table 1**). The principle of the setting of dummy variables was to equalize numbers between groups and accord with the meaning of each variable value. For example, educational level (low) represented the sum of educational status including illiterate, primary school, middle school, and high school, while educational level(high) represented the sum of educational status including junior college, undergraduate, and master. Overall, 6 continuous variables, together with 12 dummy variables of 6 categorical variables were introduced into the model.

Table 1 The variable weight of the LASSO regression and description of dummy variables

Variable	The variable weight of LASSO regression	Description
<i>Continuous variables</i>		
Duration	0.077	
PSQI	0.078	
FS-14	1.687	
Anxiety	0.043	
Depression	0.004	
Social support	-0.455	
<i>Categorical variables</i>		
Sex(male)	-0.171	
Sex(female)	0	
Age(younger)	-3.198	<35
Age(older)	0	≥35
MITOS(stage 0-1)	-12.235	MITOS stage 0,1
MITOS(stage 2-4)	0	MITOS stage 2,3,4
Economic status(common)	0	poor, average
Economic status(better)	0	good, excellent
Household(village)	0	suburban, rural
Household(city)	0	city
Educational level(low)	2.136	illiterate primary school middle school high school technical secondary school
Educational level(high)	0	junior college undergraduate master

3. Results

3.1 Descriptive analysis of characteristics of caregivers and ALS patients

The majority of this cohort was two-thirds female caregivers (38, 66.7 %), and the other one-third were male caregivers (19, 33.3%). The mean age of all the caregivers was 45.32±13.74 years(IQR: 33.5-56.5), ranging from 22-72 years. The average disease course was 28.98±15.36 months(interquartile range, IQR: 18-36), with the longest being 60 months and the shortest being 6 months. The average score of caregiver burden scale was 39.28±17.11(IQR:25-55). The average scores of PSQI and FS-14 were 7.89±4.57(IQR:4-10), 6.58±3.60(IQR:4-9), respectively. The average HAMA scores of patients were 10.70±6.62(IQR:5-15). The average HAMD scores of patients were 10.16±8.78(IQR:4-14.5). The mean score of the Social Support Rating Scale was 40.18±7.47(IQR:33.5-46). Forty caregivers were considered "anxious" to different degrees, and the number of possible, definite, obvious, and severe anxiety status caregivers were 20,15,4,1, respectively. Similarly, 27

caregivers were considered "depressed", and the number of possible, moderate, and severe depression status caregivers were 16,10,1, respectively. For categorical variables, the number of patients with MITOS from zero to stage 4 was 10,16,20,4,7, respectively. Most of the caregivers were junior high school graduates (17, 29.8%). Overall, 34 participants (59.6%) had an education level of at least high school, of which two held a master's degree. Urban dwellers constituted a proportion of 56.1% of all the caregivers, while the rest of the individuals were from suburban or rural areas. Economic status varied from general to good condition among the participants (**Table 2**).

Table 2 Demographic features of caregivers

	N	N%		N	N%
gender			MITOS		
male	19	33.3	stage 0	10	17.5
female	38	66.7	stage 1	16	28.1
education			stage 2	20	35.1
illiterate	1	1.8	stage 3	4	7.0
primary school	4	7.0	stage 4	7	12.3
middle school	17	29.8			
technical	1	1.8	economic		
secondary school			state		
high school	12	21.1	poor	5	8.8
junior college	10	17.5	average	21	36.8
undergraduate	10	17.5	good	9	15.8
master	2	3.5	excellent	22	38.6
household register			Age		
Rural	18	31.6	<35	15	26.3
suburban	7	12.3	≥35	42	73.7
city	32	56.1	total	57	100%

232

233 3.2 Analysis of correlations between continuous variables

234 Correlations between continuous variables were demonstrated with Heatmap Chart shown in
 235 **Figure 1**. According to Spearman correlation analysis, positive correlations were seen between
 236 caregiver burden and independent variables including duration of disease, PSQI scores, degree of
 237 fatigue, anxiety, and depression ($r=0.52, 0.54, 0.76, 0.71, 0.62, p<0.05$), while a negative correlation was
 238 noted between caregiver burden and social support ($r= -0.41, p<0.05$). Interestingly, a slight negative
 239 correlation was noticed between social support and PSQI scores, degree of fatigue, and depression ($r=$
 240 $-0.33, -0.29, -0.37, p<0.05$), indicating that social support could be a protective factor. Strong pairwise
 241 positive correlations were seen among PSQI scores, degree of fatigue, anxiety, and depression ($p<0.05$).

242 3.3 Analysis of correlations between caregiver burden and categorical variables

243 Correlations between caregiver burden and categorical variables were shown as visualizations
 244 (**Figure 2**). We used layered data processing to rearrange categorical variables and convert them
 245 into dummy variables. The principle of the setting of dummy variables was to equalize numbers
 246 between groups. The descriptions of each dummy variable were shown in **Table 1**. **Figure 2 (1)**
 247 showed that there was a difference in caregiver burden between younger and older patients. Also,
 248 an obvious disparity in caregiver burden was seen between group MITOS stage(0-1) and group
 249 MITOS stage(2-4), suggesting that patients with ALS with severer symptoms created higher
 250 caregiver burden (**Figure 2 (5)**).

251 3.4 Model construction and evaluation by Lasso regression

3.4.1 Variable Selection

The values of each categorical attribute were converted into dummy variables to build a regression model. With caregiver burden as the dependent variable and duration, PSQI, FS-14, anxiety, depression, social support, sex(male), sex(female), age(younger), age(older), MITOS stage (0-1), MITOS stage(2-4), economic status(common), economic status(better), household(village), household(city), educational level(low), educational level(high) as independent variables, a regression model with multiple dummy variables was constructed shown in **Table 1**.

In the LASSO model, we found ten variables had weights. Among them, MITOS stage(0-1) had the highest weight (-12.235), followed by younger age group(-3.198), lower-educated group (2.136), fatigue(1.687), and social support(-0.45). Variables including sleep quality, anxiety, depression, and sex(man) had moderate weights in this model. Economic status(common), economic status(better), household(city), household(village), educational level(high), sex(female), age(older), MITOS stage(2-4), had a weight of zero (**Table 1**).

3.4.2 Performance evaluation

In this research, the mean squared error (MSE) was selected as the evaluation index, and the best Alpha value was obtained through cross-validation. The relationship between the model MSE and Alpha is shown in **Figure 3**. When the best penalty factor 'Alpha'= 0.3, the MSE is the smallest, and the LASSO regression model has the highest accuracy. To evaluate prediction performance and fitting effect, the R^2 and the MSE of the LASSO regression model were provided. The Lasso score(R^2) and the lasso MSE were 0.831 and 48.496, respectively.

3.5 Linear regression analysis between MITOS stage(0-1), FS-14, and caregiver burden

3.5.1 Variable selection

In the previous section, we learned from the LASSO regression model that MITOS stage(0-1) had the highest weight among categorical variables, while FS-14 had the highest weight among continuous variables. Moreover, the coefficient of MITOS stage(0-1) was significantly higher than other influential dummy variables, and the coefficient of FS-14 was even hundreds of times higher than other continuous variables. Therefore, to furtherly validated the action of the most influential factors in the model, a linear regression model was constructed with caregiver burden as the dependent variable, and FS-14, MITOS stage (0-1) as independent variables.

3.5.2 Performance evaluation

As was shown in **Figure 4**, caregiver burden was positively correlated with FS-14 whenever dummy variable MITOS stage(0-1) was kept in the model or not. On the other hand, when FS-14 remained constant, caregiver burden was significantly higher when dummy variable MITOS stage(0-1) was absent. The distance between the red line and the green line demonstrated the influence of MITOS stage(0-1) on caregiver burden in the model. The intercepts of the linear regression model was 32.125. The Lasso score(R^2) and the lasso MSE were 0.783 and 62.46, respectively. The results were consistent with LASSO regression, suggesting that MITOS stage(0-1) had a strong effect on caregiver burden.

4. Discussion

Based on the LASSO model, 10 variables had weights, whereas 5 variables had a greater influence on the model. MITOS stage(0-1) had the highest weight (-12.235), followed by younger age group(-3.198), lower-educated group (2.136), fatigue(1.687), and social support(-0.45). According to the results mentioned above, MITOS stage of patients with ALS was the most important factor influencing caregiver burden, and caregiver burden was significantly lower in patients of ALS with a milder type (MITO stage 0-1). Interestingly, when another variable duration of disease was brought into the regression model, it didn't show any effect on the model, indicating that compared to duration, the severity of patients with ALS had a greater influence on caregiver

burden. Sarah K. Bublitz found that thick mucus was a highly distressing symptom for both patients and caregivers, which caused a higher degree of bulbar impairment and had a strong impact on their quality of life³⁰.

We draw a conclusion from our study that caregiver burden was positively correlated with the degree of fatigue, while the correlation between caregiver burden and psychological factors such as anxiety and depression was relatively weak. The research of Brizzi K T et.al also showed that the pressure of caregiving in ALS originated from physical symptoms consisting of fatigue, pain, weakness, shortness of breath, difficulty sleeping, and speech problems, as well as depression and other mood changes. It was worth noting that fatigue and weakness were the most common symptoms and happened in over 75% of respondents³¹, which was in accordance with our study. In our LASSO model, fatigue had higher weights than anxiety, depression, and sleep quality. Physical exhaustion was also a typical feature of caregiver burden of other progressive diseases, such as Alzheimer's disease, Parkinson's disease, and cancer³²⁻³⁵. Not surprisingly, the feelings of fatigue probably came from long-distance travel to the hospital for treatment, strenuous work in caregiving, and loss of rest time. This could well explain our result that younger caregivers suffered less from caregiver burden because they were healthier and more energetic, although they needed to consider time-consuming costs and balance between work and caring. For those younger caregivers, their caring patients were also younger as most of the caregivers were their spouses. This could be another reason that younger caregivers experienced less caregiver burden.

In this study, we found that caregiver burden was negatively correlated with social support, which was consistent with a study in Korea³⁶. According to the epidemiology of ALS in China, we have a potentially lower incidence and prevalence of ALS than in foreign countries³⁷. It is still a rare disease so doctors with less experience may fail to make a correct diagnosis in time and provide appropriate treatment. Even if the diagnosis is clear, effective treatment is not always available to every family due to the high cost of treating the illness. Fortunately, ALS is gradually being covered by health insurance in many provinces of China. The guarantee of a new policy is of great importance in helping relieve the economic burden on caregivers. In addition to economic support, more attention was paid to this group because of increasing caring activities, media reports, and social propaganda such as the "Ice Bucket Challenge" all over the country³⁸. Multiple-center clinical studies of potential drugs based on animal experiments have also been conducted by a great number of Chinese research teams^{39,40}, bringing hope and light to patients and caregivers of ALS.

On the other hand, according to our study, social support only had a modest action on caregiver burden. This phenomenon might be caused by insufficient and unbalanced social support in China. Therefore, we speculated that greater social support would bring about less caregiver burden. In some well-developed regions such as Hong Kong, the promotion of multidisciplinary teams (MDT) provides effective support for the disease, which encourages patients and families of ALS in a highly positive manner⁴¹. Family care is the main strategy for Chinese caregivers to look after ALS patients, and older women and spouses still play an important role in providing family care, while in some other countries, hospital care, nursing home care, and community care were made full use of to help alleviate the families' burden^{42,43}. The lack of professional caring experience, staff, and facilities could affect the efficiency of caring and thus lead to an increasing caregiver burden. Although China has developed rapidly in recent years, there have been obvious differences in medical development and medical investments between eastern and western areas. Little research on ALS has been conducted in western areas, where economic conditions are relatively poor and social support is inadequate. In contrast, modern countries, such as Japan and other European countries can offer stronger social aid and help systems to ALS families through long-term care insurance, home visit rehabilitation, and peer group rehabilitation, suggesting that more social support could be provided to Chinese ALS caregivers to relieve their burden^{13, 43-46}. In addition, medical expenses and end-of-life care were guaranteed in well-developed countries like Japan. Obviously, there is an urgent need for more economical, practical, physical, emotional, and spiritual support for caregivers of ALS.

It was surprising to find in our study that educational level influenced caregiver burden. In this study, caregivers of ALS patients were mainly their spouses who were of the same age (50-70y) and had similar educational levels (middle school) to the patients. A possible explanation is that

1
2
3 353 understanding abilities varied among different caregivers and could be a limitation in low-educated
4 354 groups. Under the current situation in China, patients were not able to address all their questions in
5 355 a short period because of the large hospital outpatient quotas. For those caregivers who had a lower
6 356 educational level, they might feel difficult to quickly understand the key points in caring and follow
7 357 doctors' suggestions. In addition, they possibly lacked the abilities to search for information and teach
8 358 themselves to improve the efficiency of home nursing work.

9 359 We investigated the correlations between caregiver burden and caregivers' mental states,
10 360 including anxiety, depression, sleep quality index, and fatigue. Although we observed a slight
11 361 correlation between anxiety, depression, and caregiver burden, they didn't show high weights in the
12 362 LASSO regression model as we expected. In contrast with other studies conducted in China, mental
13 363 problems do exist comprehensively in caregivers of ALS, and anxiety and depression were associated
14 364 closely with each other^{10,47}. However, the actual condition was more complicated. Firstly, anxiety was
15 365 not only seen in caregivers of ALS but was also common in regular populations, especially women
16 366 in their peri- or postmenopausal period⁴⁸. Based on the results of an epidemiological study of ALS
17 367 in the Chinese population, the ratio of male patients to female patients in China was about 1.45:1.98:1
18 368³⁷, which is consistent with our findings. As there were more male patients than female patients in
19 369 our study, it was speculated that caregivers were mostly women aged 50–70, who had possible risks
20 370 of being anxious in any event. Traditionally, Chinese women of this age also face the burden of taking
21 371 care of their grandchildren, making it harder to balance family and their feelings. Secondly, as there
22 372 was obvious collinearity among anxiety, depression, and quality of sleep, their actual influence on
23 373 caregiver burden may be overestimated. Even though, non-drug intervention seems to be available
24 374 and effective for alleviating caregiver burden among caregivers of patients with neuropsychiatric
25 375 disorders^{49,50}.

26 376 From the conclusion obtained above, practical actions should be done for healthcare system to
27 377 help relieve caregiver burden of ALS. Financial assistance and professional nursing guidance,
28 378 including long-term care insurance, medical investments, professional online communication
29 379 groups, home visit rehabilitation, nursing home care and community care etc., may be particularly
30 380 beneficial as part of ALS care^{13,43,45}. Regional ALS centres participated by doctors, nurses,
31 381 psychologists, rehabilitation trainers, social workers, nutritionists, communities, speech trainers,
32 382 families, and researchers, are necessary for healthcare system to provide complete and efficient
33 383 support^{10,51}. Physical, psychological, emotional, and social support are potential effective non-
34 384 medication interventions to help relieve caregiver burden^{42,52-54}. Healthcare system should pay more
35 385 attention to the health status of both the patients and their caregivers proactively. Hopefully,
36 386 positive and effective strategies can be developed as soon as possible to improve the situation.

37 387 38 388 Limitations

39 389 There were four limitations in this study. The sample size was small and respondents could
40 390 experience remission, so our observational analysis could have been biased. In addition, we failed to
41 391 assess intervention effects with quantified coping strategies in a prospective study. Another
42 392 limitation of the study was the lack of information about the patients and the lack of correlations
43 393 between patients' physical or psychological aspects and caregiver burden. Lastly, this was a study
44 394 conducted in south-eastern China, and caregivers may only represent a part of the Chinese
45 395 population.

46 396 5. Conclusion

47 397 In conclusion, our study demonstrates that the most influencing factor of caregiver burden is the
48 398 severity of ALS patient. Caregivers of ALS patients may suffer less from caregiver burden when the
49 399 patients are less severe and the caregivers are younger. Low educational status could increase
50 400 caregiver burden. Caregiver burden is positively correlated with the degree of fatigue and negatively
51 401 correlated with social support.

52 402
53 403
54 404
55 405 **Figure 1** Correlations among caregiver burden and continuous variables

1
2
3 406 **Figure 2** Correlations between caregiver burden and categorical variables

4 407 **Figure 3** Relationship between α and MSE

5 408 **Figure 4** Linear regression between caregiver burden and FS-14

6 409

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3 410
45 **Footnotes**6 **Contributions**

7
8 413 Drafting of the manuscript and data analysis: LL. Data collection and follow patients: MYZ. Formal
9 414 analysis: RJH. Paper editing and submission: JNL. Validation: WNC. Direct guidance, paper
10 415 editing, and reviewing: ZP. Project designing and guidance: XLY.
11 416 All authors commented on previous versions of the manuscript and approved the final manuscript.
12 417

13
14 **Funding**

15 419 This study was supported by grants from the Guangdong Provincial Key Laboratory of Diagnosis
16 420 and Treatment of Major Neurological Diseases (2020B1212060017), Guangdong Provincial Clinical
17 421 Research Center for Neurological Diseases (2020B111170002), Southern China International Joint
18 422 Research Center for Early Intervention and Functional Rehabilitation of Neurological Diseases
19 423 (2015B050501003 and 2020A0505020004), Guangdong Provincial Engineering Center for Major
20 424 Neurological Disease Treatment, Guangdong Provincial Translational Medicine Innovation
21 425 Platform for Diagnosis and Treatment of Major Neurological Disease, Guangzhou Clinical Research
22 426 and Translational Center for Major Neurological Diseases (201604020010), The Young Scientist
23 427 Fund of National Natural Science Foundation of China, grant number 82001332, The National Key
24 428 Research and Development Program of China, grant number 2017YFC0907703.
25 429

26 430 **Competing interests**

27
28
29 431 None declared.
30 432

31
32 **Provenance and peer review**

33 434 Not commissioned; externally peer-reviewed.
34 435

35
36 **Patient consent for publication**

37 437 Not required.
38 438

39
40 **Ethics Statement**

41 439
42 440 As this descriptive study is an analysis of existing data, which does not involve patients' or
43 441 caregivers' diagnosis and treatment, no ethics approval is required from the Ethics Committee of
44 442 First Affiliated hospital of Sun yat-sen University. Even though, participants enrolled have been
45 443 fully informed and consent was obtained.
46 444

47
48 **Data sharing statement**

49 446 Data are available upon reasonable request. The data that support the findings of this study are
50 447 available from the corresponding author on reasonable request.
51 448

1
2
3 449
4 450 **References**

- 5
6 451 1. Geng D, Ou R, Miao X, et al. Patients' self-perceived burden, caregivers' burden and quality of
7 452 life for amyotrophic lateral sclerosis patients: a cross-sectional study. *J Clin Nurs* 2017;26(19-
8 453 20):3188-99. doi: 10.1111/jocn.13667
- 9 454 2. Weisser FB, Bristowe K, Jackson D. Experiences of burden, needs, rewards and resilience in
10 455 family caregivers of people living with Motor Neurone Disease/Amyotrophic Lateral
11 456 Sclerosis: A secondary thematic analysis of qualitative interviews. *Palliat Med*
12 457 2015;29(8):737-45. doi: 10.1177/0269216315575851
- 13 458 3. Boerner K, Mock SE. Impact of patient suffering on caregiver well-being: the case of amyotrophic
14 459 lateral sclerosis patients and their caregivers. *Psychol Health Med* 2012;17(4):457-66. doi:
15 460 10.1080/13548506.2011.613942
- 16 461 4. Chio A, Gauthier A, Calvo A, et al. Caregiver burden and patients' perception of being a burden
17 462 in ALS. *Neurology* 2005;64(10):1780-82. doi: Doi 10.1212/01.Wnl.0000162034.06268.37
- 18 463 5. Creemers H, Veldink JH, Grupstra H, et al. Cluster RCT of case management on patients' quality
19 464 of life and caregiver strain in ALS. *Neurology* 2014;82(1):23-31. doi:
20 465 10.1212/01.wnl.0000438227.48470.62
- 21 466 6. Yuksel B, Ak PD, Sen A, et al. Caregiver burden and quality of life in early stages of idiopathic
22 467 Parkinson's disease. *Idegyogy Sz* 2018;71(9-10):343-50. doi: 10.18071/isz.71.0343
- 23 468 7. Isik AT, Soysal P, Solmi M, et al. Bidirectional relationship between caregiver burden and
24 469 neuropsychiatric symptoms in patients with Alzheimer's disease: A narrative review. *Int J*
25 470 *Geriatr Psychiatry* 2018 doi: 10.1002/gps.4965
- 26 471 8. Kim MS, Shin HI, Min Y, et al. Correlation between Severe ALS Patient-Caregiver Couples'
27 472 Characteristics and Caregivers' Health Related Quality of Life. *J Korean Acad Nurs*
28 473 2011;41(3):354-63. doi: 10.4040/jkan.2011.41.3.354
- 29 474 9. Burke T, Elamin M, Galvin M, et al. Caregiver burden in amyotrophic lateral sclerosis: a cross-
30 475 sectional investigation of predictors. *J Neurol* 2015;262(6):1526-32. doi: 10.1007/s00415-015-
31 476 7746-z
- 32 477 10. Tang S, Li L, Xue H, et al. Caregiver burden and associated factors among primary caregivers of
33 478 patients with ALS in home care: a cross-sectional survey study. *BMJ Open*
34 479 2021;11(9):e050185. doi: 10.1136/bmjopen-2021-050185 [published Online First: 2021/10/01]
- 35 480 11. Siciliano M, Santangelo G, Trojsi F, et al. Coping strategies and psychological distress in
36 481 caregivers of patients with Amyotrophic Lateral Sclerosis (ALS). *Amyotroph Lateral Scler*
37 482 *Frontotemporal Degener* 2017;18(5-6):367-77. doi: 10.1080/21678421.2017.1285316 [published
38 483 Online First: 2017/06/21]
- 39 484 12. de Wit J, Schroder CD, El Mecky J, et al. Support needs of caregivers of patients with
40 485 amyotrophic lateral sclerosis: A qualitative study. *Palliat Support Care* 2018:1-7. doi:
41 486 10.1017/S1478951517001213
- 42 487 13. Cipolletta S, Gammino GR, Francescon P, et al. Mutual support groups for family caregivers of
43 488 people with amyotrophic lateral sclerosis in Italy: A pilot study. *Health Soc Care Community*
44 489 2018;26(4):556-63. doi: 10.1111/hsc.12558 [published Online First: 2018/02/27]
- 45 490 14. Chu HS, Tak YR, Kim SH. Factors Influencing Psychosocial Well-Being in Family Caregivers of
46 491 People with Amyotrophic Lateral Sclerosis. *J Korean Acad Nurs* 2018;48(4):454-64. doi:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

492 10.4040/jkan.2018.48.4.454

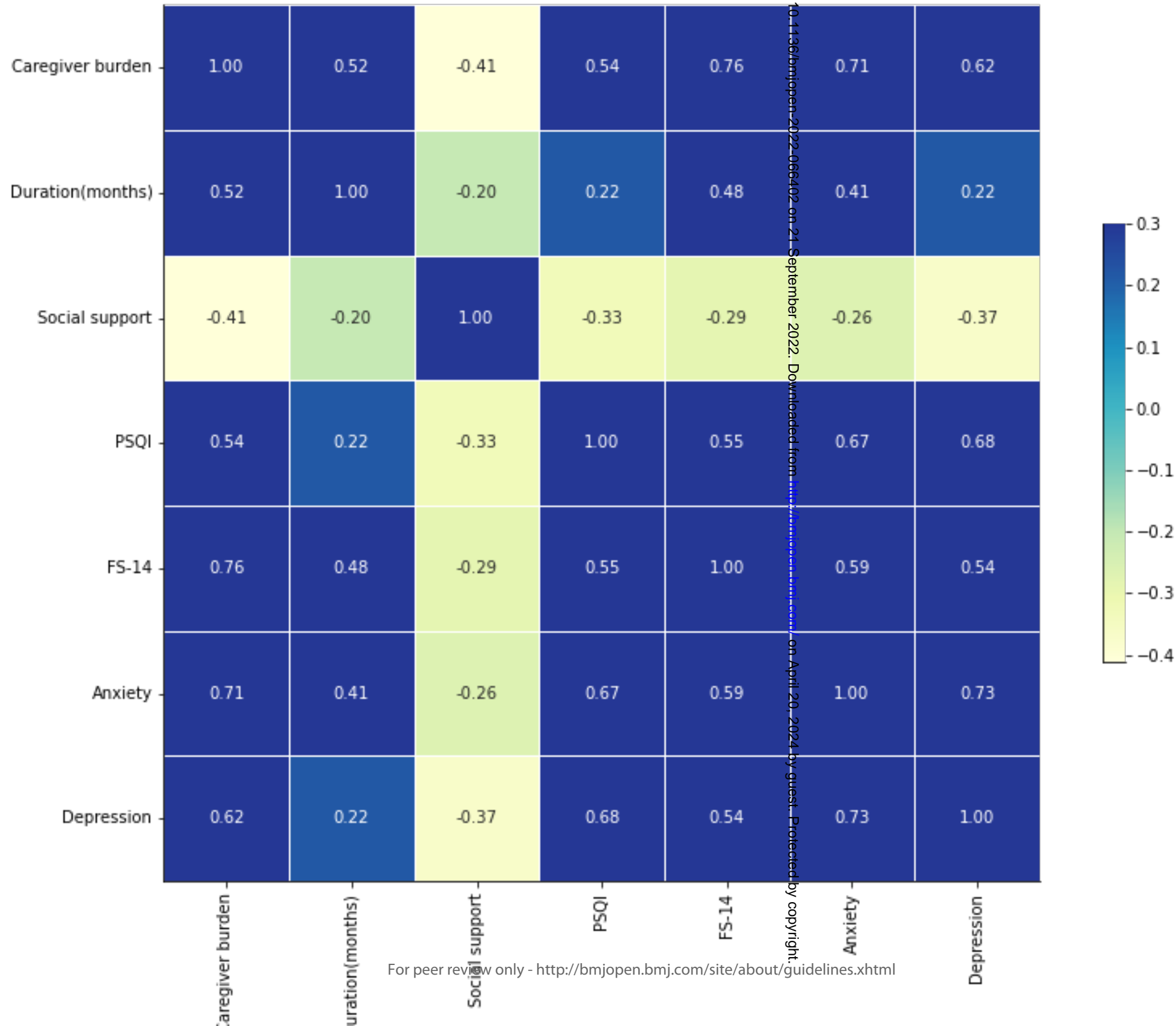
- 493 15. Oh J, Kim JA. Factor analysis of the Zarit Burden Interview in family caregivers of patients with
494 amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener* 2018;19(1-2):50-
495 56. doi: 10.1080/21678421.2017.1385636
- 496 16. Chio A, Hammond ER, Mora G, et al. Development and evaluation of a clinical staging system
497 for amyotrophic lateral sclerosis. *J Neurol Neurosurg Ps* 2015;86(1):38-44. doi: 10.1136/jnnp-
498 2013-306589
- 499 17. Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of
500 burden. *Gerontologist* 1980;20(6):649-55. doi: 10.1093/geront/20.6.649 [published Online First:
501 1980/12/01]
- 502 18. Arai Y, Kudo K, Hosokawa T, et al. Reliability and validity of the Japanese version of the Zarit
503 Caregiver Burden interview. *Psychiatry Clin Neurosci* 1997;51(5):281-7. doi: 10.1111/j.1440-
504 1819.1997.tb03199.x [published Online First: 1997/12/31]
- 505 19. Xiao SY. The theoretical basis and application of social support questionnaire. *J Clin Psychi*
506 1994(4):98-100.
- 507 20. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. *J Psychosom Res*
508 1993;37(2):147-53. doi: 10.1016/0022-3999(93)90081-p [published Online First: 1993/01/01]
- 509 21. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959;32(1):50-5. doi:
510 10.1111/j.2044-8341.1959.tb00467.x [published Online First: 1959/01/01]
- 511 22. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62. doi:
512 10.1136/jnnp.23.1.56 [published Online First: 1960/02/01]
- 513 23. Zhu XW, Xin YJ, Ge HL. Recursive Random Forests Enable Better Predictive Performance and
514 Model Interpretation than Variable Selection by LASSO. *J Chem Inf Model* 2015;55(4):736-46.
515 doi: 10.1021/ci500715e
- 516 24. Sun C, Zhu B, Zhu S, et al. Risk Factors Analysis of Bone Mineral Density Based on Lasso and
517 Quantile Regression in America during 2015-2018. *Int J Environ Res Public Health* 2021;19(1)
518 doi: 10.3390/ijerph19010355 [published Online First: 2022/01/12]
- 519 25. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical*
520 *Society* 1996;58(1):267-88.
- 521 26. Rajaratnam B, Roberts S, Sparks D, et al. Lasso regression: estimation and shrinkage via the limit
522 of Gibbs sampling. *Journal of the Royal Statistical Society* 2015;78:153-74.
- 523 27. Hassan SS, Farhan M, Mangayil R, et al. Bioprocess data mining using regularized regression
524 and random forests. *Bmc Syst Biol* 2013;7 doi: Artn S5
525 10.1186/1752-0509-7-S1-S5
- 526 28. Myers JL, Well A. Research design and statistical analysis. 2nd ed. Mahwah, N.J.: Lawrence
527 Erlbaum Associates 2003.
- 528 29. Alkharusi H. Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect
529 Coding. *International Journal of Education* 2012;4(2) doi: 10.5296/ije.v4i2.1962
- 530 30. Bublitz SK, Mie E, Wasner M, et al. Thick Mucus in ALS: A Mixed-Method Study on Associated
531 Factors and Its Impact on Quality of Life of Patients and Caregivers. *Brain Sci* 2022;12(2)
532 doi: 10.3390/brainsci12020252 [published Online First: 2022/02/26]
- 533 31. Brizzi KT, Bridges JFP, Yersak J, et al. Understanding the needs of people with ALS: a national
534 survey of patients and caregivers. *Amyotroph Lateral Scler Frontotemporal Degener* 2020;21(5-

- 1
2
3 535 6):355-63. doi: 10.1080/21678421.2020.1760889 [published Online First: 2020/05/13]
4
5 536 32. Madaleno TR, Moriguti JC, Ferriolli E, et al. Mood, lifestyle and cardiovascular risk factors
6 537 among older caregivers of patients with Alzheimer's disease dementia: a case-control study.
7 538 *Aging Clin Exp Res* 2019 doi: 10.1007/s40520-019-01212-8 [published Online First: 2019/05/11]
8
9 539 33. Liu S, Li C, Shi Z, et al. Caregiver burden and prevalence of depression, anxiety and sleep
10 540 disturbances in Alzheimer's disease caregivers in China. *J Clin Nurs* 2017;26(9-10):1291-300.
11 541 doi: 10.1111/jocn.13601 [published Online First: 2016/09/30]
12
13 542 34. Torny F, Videaud H, Chatainier P, et al. Factors associated with spousal burden in Parkinson's
14 543 disease. *Rev Neurol (Paris)* 2018;174(10):711-15. doi: 10.1016/j.neurol.2018.01.372 [published
15 544 Online First: 2018/07/24]
16
17 545 35. Molassiotis A, Zheng Y, Denton-Cardew L, et al. Symptoms experienced by cancer patients
18 546 during the first year from diagnosis: patient and informal caregiver ratings and agreement.
19 547 *Palliat Support Care* 2010;8(3):313-24. doi: 10.1017/S1478951510000118 [published Online
20 548 First: 2010/09/30]
21
22 549 36. Chu HS, Tak YR, Kim SH. [Factors Influencing Psychosocial Well-Being in Family Caregivers of
23 550 People with Amyotrophic Lateral Sclerosis]. *J Korean Acad Nurs* 2018;48(4):454-64. doi:
24 551 10.4040/jkan.2018.48.4.454 [published Online First: 2018/09/13]
25
26 552 37. Liu X, He J, Gao FB, et al. The epidemiology and genetics of Amyotrophic lateral sclerosis in
27 553 China. *Brain Res* 2018;1693(Pt A):121-26. doi: 10.1016/j.brainres.2018.02.035 [published
28 554 Online First: 2018/03/05]
29
30 555 38. Kuehn BM. Simple Models and Ice Bucket Challenge Fuel Progress in ALS Treatment. *JAMA*
31 556 2018;319(6):535-37. doi: 10.1001/jama.2017.20704 [published Online First: 2018/01/25]
32
33 557 39. Feng XH, Yuan W, Peng Y, et al. Therapeutic effects of dl-3-n-butylphthalide in a transgenic
34 558 mouse model of amyotrophic lateral sclerosis. *Chin Med J (Engl)* 2012;125(10):1760-6.
35 559 [published Online First: 2012/07/18]
36
37 560 40. Huang C, Li J, Zhang G, et al. TBN improves motor function and prolongs survival in a TDP-
38 561 43M337V mouse model of ALS. *Hum Mol Genet* 2021;30(16):1484-96. doi:
39 562 10.1093/hmg/ddab101 [published Online First: 2021/05/01]
40
41 563 41. Cheng HWB, Chan KY, Chung YKJ, et al. Supportive & palliative interventions in motor
42 564 neurone disease: what we know from current literature? *Ann Palliat Med* 2018;7(3):320-31.
43 565 doi: 10.21037/apm.2017.10.01
44
45 566 42. Levitsky GN, Chub RV, Kryachkov AV. [The quality of medical care and cooperation of patients
46 567 with amyotrophic lateral sclerosis in the Russian Federation]. *Zh Nevrol Psikhiatr Im S S*
47 568 *Korsakova* 2017;117(1. Vyp. 2):59-63. doi: 10.17116/jnevro20171171259-63 [published Online
48 569 First: 2017/05/18]
49
50 570 43. Hamaguchi Y, Hakuno A, Saito K, et al. [Home Visit Rehabilitation for a Patient with ALS to
51 571 Improve Social Participation]. *Gan To Kagaku Ryoho* 2019;46(Suppl 1):46-48. [published
52 572 Online First: 2019/06/14]
53
54 573 44. Arai Y, Zarit SH. Exploring strategies to alleviate caregiver burden: effects of the National Long-
55 574 Term Care insurance scheme in Japan. *Psychogeriatrics* 2011;11(3):183-9. doi: 10.1111/j.1479-
56 575 8301.2011.00367.x [published Online First: 2011/09/29]
57
58 576 45. Madsen LS, Jeppesen J, Handberg C. "Understanding my ALS". Experiences and reflections of
59 577 persons with amyotrophic lateral sclerosis and relatives on participation in peer group

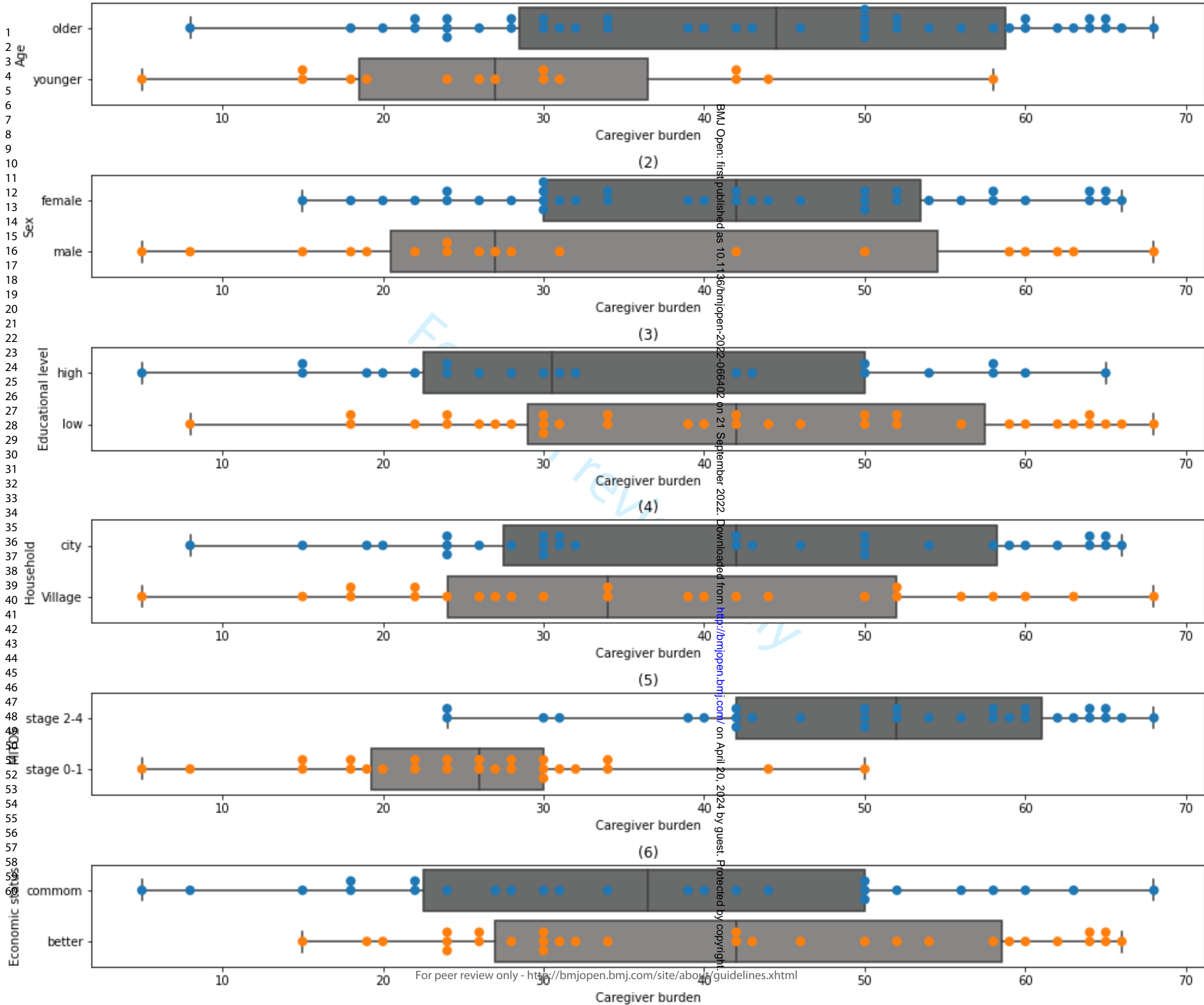
- 1
2
3 578 rehabilitation. *Disabil Rehabil* 2019;41(12):1410-18. doi: 10.1080/09638288.2018.1429499
4 579 [published Online First: 2018/01/28]
5
6 580 46. Ogino M. [End of life care for patients with ALS in Japan]. *Rinsho Shinkeigaku* 2008;48(11):973-5.
7 581 doi: 10.5692/clinicalneuro.48.973 [published Online First: 2009/02/10]
8
9 582 47. Chen D, Guo X, Zheng Z, et al. Depression and anxiety in amyotrophic lateral sclerosis:
10 583 correlations between the distress of patients and caregivers. *Muscle Nerve* 2015;51(3):353-7.
11 584 doi: 10.1002/mus.24325 [published Online First: 2014/07/01]
12
13 585 48. Mulhall S, Andel R, Anstey KJ. Variation in symptoms of depression and anxiety in midlife
14 586 women by menopausal status. *Maturitas* 2018;108:7-12. doi: 10.1016/j.maturitas.2017.11.005
15 587 [published Online First: 2018/01/02]
16
17 588 49. Chen HM, Huang MF, Yeh YC, et al. Effectiveness of coping strategies intervention on caregiver
18 589 burden among caregivers of elderly patients with dementia. *Psychogeriatrics* 2015;15(1):20-
19 590 25. doi: 10.1111/psyg.12071
20
21 591 50. Martin-Carrasco M, Fernandez-Catalina P, Dominguez-Panchon AI, et al. A randomized trial to
22 592 assess the efficacy of a psychoeducational intervention on caregiver burden in
23 593 schizophrenia. *Eur Psychiatry* 2016;33:9-17. doi: 10.1016/j.eurpsy.2016.01.003
24
25 594 51. Caga J, Zoing MC, Foxe D, et al. Problem-focused coping underlying lower caregiver burden in
26 595 ALS-FTD: implications for caregiver intervention. *Amyotroph Lateral Scler Frontotemporal*
27 596 *Degener* 2021;22(5-6):434-41. doi: 10.1080/21678421.2020.1867180 [published Online First:
28 597 2021/01/14]
29
30 598 52. de Wit J, Beelen A, Drossaert CHC, et al. A blended psychosocial support program for partners
31 599 of patients with amyotrophic lateral sclerosis and progressive muscular atrophy: protocol
32 600 of a randomized controlled trial. *BMC Psychol* 2018;6(1):20. doi: 10.1186/s40359-018-0232-5
33 601 [published Online First: 2018/05/03]
34
35 602 53. Ang K, Umaphathi T, Tong J, et al. Healthcare Needs of Patients with Amyotrophic Lateral
36 603 Sclerosis (ALS) in Singapore: A patient-centred qualitative study from multiple
37 604 perspectives. *J Palliat Care* 2015;31(3):150-7. [published Online First: 2015/10/31]
38
39 605 54. Giusiano S, Peotta L, Iazzolino B, et al. Amyotrophic lateral sclerosis caregiver burden and
40 606 patients' quality of life during COVID-19 pandemic. *Amyotroph Lateral Scler Frontotemporal*
41 607 *Degener* 2022;23(1-2):146-48. doi: 10.1080/21678421.2021.1912772 [published Online First:
42 608 2021/04/17]
43
44 609
45 610
46 611
47 612
48 613

Acknowledgments

49 614 We thank all patients and their caregivers of ALS for taking part in this research. We gratefully
50 615 acknowledge the contribution of the doctors, nurses, and other colleagues during the study.
51 616

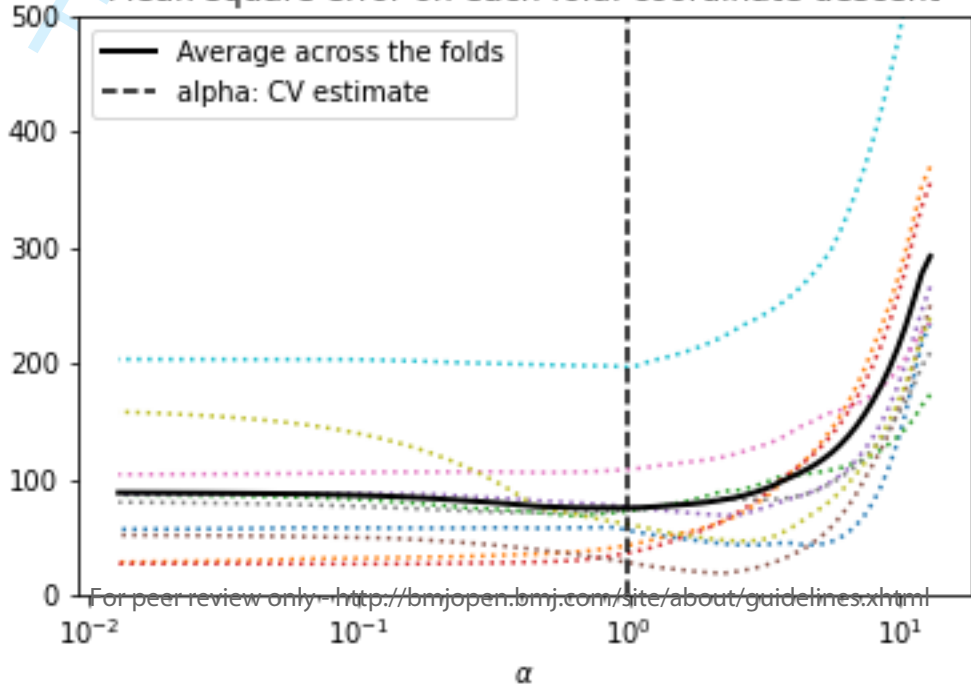


en: first published as 10.1136/bmjopen-2022-006402 on 21 September 2022. Downloaded from <http://bmjopen.bmj.com/> on April 20, 2024 by guest. Protected by copyright.

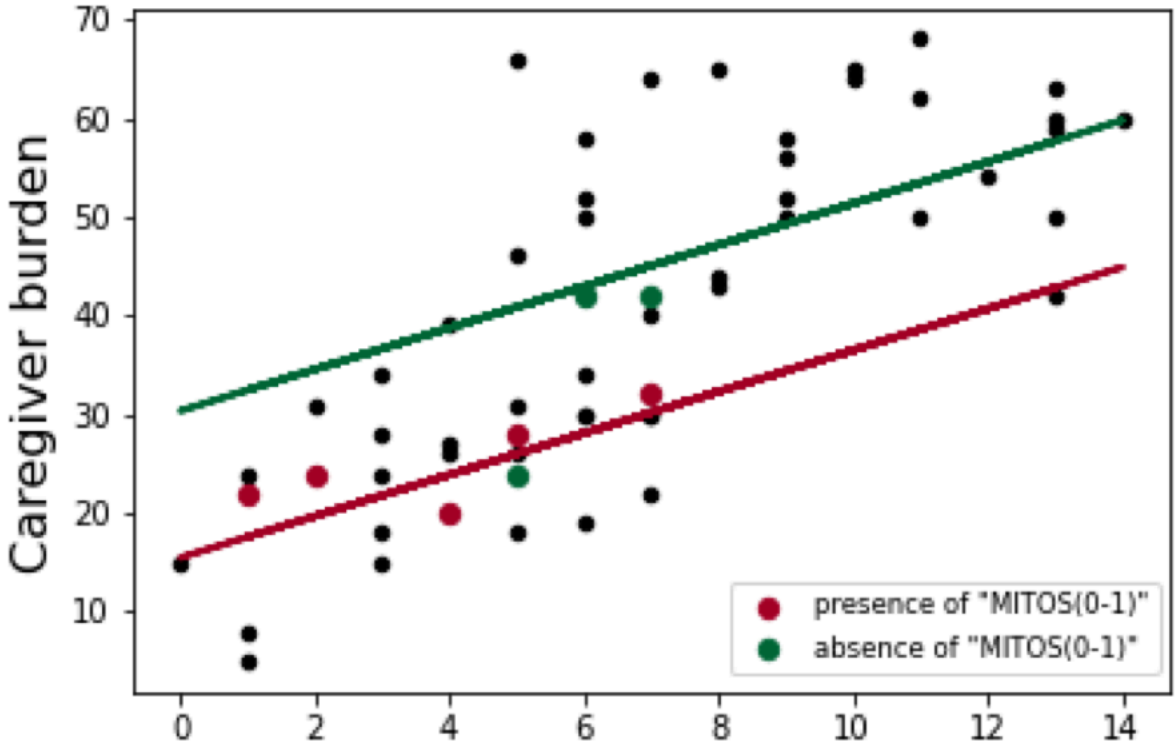


BMJ Open: first published as 10.1136/bmjopen-2022-066402 on 21 September 2022. Downloaded from <http://bmjopen.bmj.com/> on April 20, 2024 by guest. Protected by copyright.

Mean square error on each fold: coordinate descent

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28



FS-14

STROBE Statement—Checklist of items that should be included in reports of *cross-sectional studies*

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract line 5 (b) Provide in the abstract an informative and balanced summary of what was done and what was found line 22-47
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported line 67-82
Objectives	3	State specific objectives, including any prespecified hypotheses line 83-86
Methods		
Study design	4	Present key elements of study design early in the paper line 92
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection line 90-92,102-104
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants line 91-92, 97-101
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable NA
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group line 113-141
Bias	9	Describe any efforts to address potential sources of bias line 107-108
Study size	10	Explain how the study size was arrived at line 106-108
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why line 143-206
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding line 143-206 (b) Describe any methods used to examine subgroups and interactions line 143-206 (c) Explain how missing data were addressed NA (d) If applicable, describe analytical methods taking account of sampling strategy NA (e) Describe any sensitivity analyses NA
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed line 107-108 (b) Give reasons for non-participation at each stage NA (c) Consider use of a flow diagram NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders line 214-232 (b) Indicate number of participants with missing data for each variable of interest line 107-108
Outcome data	15*	Report numbers of outcome events or summary measures table 1
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included line 233-250, 251-290

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(b) Report category boundaries when continuous variables were categorized [line 251-290](#)

(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period [NA](#)

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses line 251-290
----------------	----	---

Discussion

Key results	18	Summarise key results with reference to study objectives line 397-401
-------------	----	---

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias line 389-395
-------------	----	---

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence line 292-375
----------------	----	---

Generalisability	21	Discuss the generalisability (external validity) of the study results line 376-386
------------------	----	--

Other information

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based line 418-428
---------	----	--

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.