

THE CURIOSITY CUP 2022

A Global SAS[®] Student Competition

Depression: Does Caregiving Put You at Risk? A European-Wide Study

Team OverSeAS

ABSTRACT

The Cambridge Academic Content Dictionary¹ defines a caregiver as "*Someone who provides for the needs of children or of people who are ill or cannot provide for their own needs.*" Caregivers therefore suffer for both the illness of a loved person and for the physical and mental burden of having to take care of her or him. In this report, we investigate the relationship between the different types of caring activity and the level of depression of carers who live in the different areas of Europe. We do this by estimating a cumulative proportional hazards model which relates depression to caregiving while adjusting for the area of residence and a number of demographic and socio-economic characteristics of the subjects. The final model will be used to predict the probability of depression for the varying types of caregiving and across Europe.

INTRODUCTION

Depression is a common mental disorder, characterized by persistent sadness, that affects a person's ability to function and live a rewarding life. According to the World Health Organization, 5% of the global population suffer from this condition². A major cause of depression is taking care of a loved person, especially when one cannot rely on state assistance. Indeed, different areas of Europe have different welfare policies regarding Long Term Care. These welfare policies are more developed in Northern Europe and become less effective as we cross Europe towards the Mediterranean states (Brenna and Di Novi, 2016).

PROBLEM

It is well known from the scientific literature that being a caregiver results in professional, personal and social sacrifices that can lead to high physical and mental stress (Silverstein, Gans and Yang, 2006; Bookwala, 2009). This project wants to explore the impact of caring activities on the psychological health of carers. In particular, we are interested in assessing to what extent the relationship with the person they look after affects the degree of depression and whether this effect differs across Europe.

DATA

We use data taken from SHARE³ (*Survey of Health, Ageing, and Retirement in Europe*). This multidisciplinary and longitudinal survey provides information about the health, the socio-economic status and the social and familiar relationships of European citizens. Information is collected in 28 European countries and in Israel (which we excluded from our analysis as it is not a member of the European Community) on 140,000 subjects aged over 50. Specifically, the database we used for the analysis was built in SAS by joining two different datasets using the **MERGE** statement. The primary dataset is called "EasyShare"⁴ and provides the socio-economic and demographic information on the subjects. This dataset was integrated using information from the social support questionnaire of the SHARE survey on

¹<https://dictionary.cambridge.org/it/dizionario/inglese/caregiver>

²<https://www.who.int/health-topics/depression>

³<http://www.share-project.org/data-documentation/waves-overview/wave-5.html>

⁴<http://www.share-project.org/special-data-sets/easyshare.html>

the psychological situation and the assistance given to different people. We focused on the 5th wave of the survey (released in 2015) for both datasets, as this is the one which includes the highest number of caregivers.

DATA CLEANING AND VALIDATION

OUTCOME OF INTEREST

The outcome of interest of our analysis is the level of depression of caregivers. This is measured in the SHARE survey through the EURO-D index (Prince et al., 1999). This index combines twelve markers which are related to the mental health conditions of a subject. In particular, the EURO-D index considers the following 12 aspects: *depression, pessimism, suicidal tendency, guilt, sleep, interest, irritability, appetite, tiredness, concentration, fun, and crying*. Figure 1 shows how the Euro-D index distributes in our database. It ranges from the minimum value 0 ("not depressed") to 12, which represents the highest value of depression.

Our analysis considers three levels of depression: not-depressed subjects ($\text{EURO-D} < 4$), mildly depressed subjects ($\text{EURO-D} \geq 4$ and < 7), and severely depressed subjects ($\text{EURO-D} \geq 7$). The two cut-off values 4 and 7 were chosen according to scientific input: the value 4 is commonly used in the literature to dichotomize the EURO-D index, while the value 7 allows us to achieve a proportion of severely depressed subjects, which is coherent with WHO statistics. In all, 72.12% of subjects in our database are not depressed, 20.31% are mildly depressed, and 6.56% are severely depressed.

RISK FACTOR

In our analysis, caregiving is a multilevel risk factor. The SHARE survey distinguishes between caregivers who provide care to a person they live with and carers of a person who lives in another household. To avoid self-selection bias, we decided to exclude the latter ones. Indeed, caregivers who provide care to a person outside their household necessarily are in a physically and mentally saner health status. The final database consists of 23,672 records. The different levels of our risk factor represent the relationship between the caregiver and the assisted person. In order to numerically balance the different classes, we recoded the original categories of the SHARE survey into four main types depending on whether care is provided to: a partner, to children, to parents or siblings or to other people. The fifth class refers to non-caregivers. The univariate percentage distribution of this five-level risk factor for our database is shown in Figure 2.

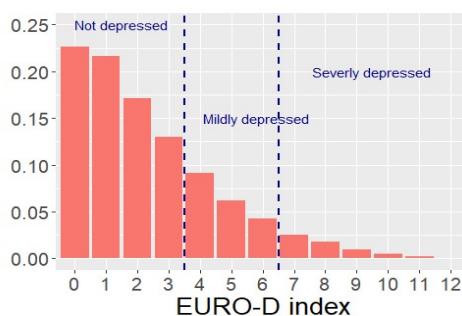


Figure 1. Univariate distribution of the EURO-D index. The cut-off values are shown as vertical dashed lines⁵.

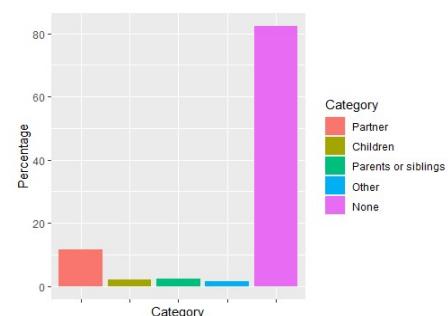


Figure 2. Bar chart of the univariate percentage distribution of the five-level risk factor representing caregiving⁵.

⁵Figure created in R (www.r-project.org).

STRATIFICATION BY GEOGRAPHICAL AREA

The third key variable of our analysis is the geographical area of residence. We have grouped the 17 European states considered by the SHARE survey into four macro-areas, based both on their geographic location and on their welfare policies. The four areas considered are:

- Northern Europe (Sweden, Denmark and Germany);
- Central Europe (Austria, France, Switzerland, Belgium, the Netherlands and Luxembourg);
- Eastern Europe (Poland, Czech Republic, Estonia and Latvia);
- Southern Europe (Italy, Spain, and Greece).

CONFOUNDERS AND EFFECT MODIFIERS

We also included other socio-economic and demographic variables which represent potential confounders and/or effect modifiers. These are:

- **Sex**, with levels "male" or "female";
- **Age**, with levels " ≤ 60 years", " > 60 but ≤ 70 years" and " > 70 years";
- **Civil status**, with levels "partnered" or "partnerless." Subjects who declared not to be in a relationship, but also classified themselves as caregivers of their partner, were recoded to "partnered";
- **Education**, with levels "low" for people who just attended middle school, "medium" for people who have a high school diploma and "high" for people who have a university degree;
- **Economic status**, to distinguish between people that have ("disadvantaged") or don't have ("advantaged") economic difficulties;
- **Employment status**, with levels "employed" and "not employed";
- **Physical activity**, with levels "yes" and "no";
- **Chronic diseases**, with levels "present" and "absent".

We chose to consider fewer classes than in the SHARE survey for these variables to avoid separation or quasi-separation problems in the statistical analysis. All recoding was dealt by **PROC FORMAT**.

ANALYSIS

METHODS

The statistical analysis was carried out in three step using **SAS OnDemand for Academics**. Significance is assessed at the 5% significance level and confidence intervals are computed with 95% confidence.

The first step is an exploratory data analysis to gain insight into the relationship between the type of caring activity and the level of depression of the carer. Odds ratios of mild and severe depression against healthy status were computed for each caregiver type with **PROC FREQ**. Mentally healthy non-caregivers represent the reference group.

We then fitted a cumulative logit model with proportional hazards to investigate how the macro-area affects the relationship between depression and caregiving activity using **PROC LOGISTIC**. Odds ratios were stratified by geographical area and furthermore adjusted to account for possible confounding and/or effect modification due to demographic and socio-economic variables. Stepwise model selection was used to select the statistically significant subset of covariates. The **PROC LOGISTIC** step is portrayed in the Appendix.

In the final analysis step, we identified the worst covariate profile, that is, the configuration of the demographic and socio-economic variable classes which characterizes subjects with highest probability of being depressed. We hence computed the probabilities of severe depression, mild depression and no depression for all five caregiver types and for the four different geographical areas. These were obtained with the **OUTPUT** statement of **PROC LOGISTIC**.

RESULTS

Exploratory data analysis. Table 1 reports the odds ratios, with 95% confidence intervals, for severe and mild depression computed at Step 1 (exploratory analysis). All odds ratios are statistically significant at the 5% level apart from the odds ratio of mild depression for subjects who take care of people other than family members. Among the severely depressed people, those who take care of their children have the highest risk of illness. In fact, with an odds ratio of 4.03, their risk of depression is 4 times higher than the corresponding risk for subjects who are not caregivers. Among the mildly depressed people, we observe the highest odds ratio of 2.12 for those who take care of their partner. This tells us that they are twice at risk for depression compared to non-caregivers.

	Severely depressed	Mildly depressed
Children vs. none	4.03 (3.13 - 5.18)	1.88 (1.53 - 2.32)
Partner vs. none	2.89 (2.53 - 3.30)	2.12 (1.94 - 2.33)
Parents vs. none	1.86 (1.38 - 2.53)	1.47 (1.20 - 1.79)
Other vs. none	1.49 (1.01 - 2.20)	1.05 (0.81 - 1.37)

Table 1. Odds-ratio of severe depression and mild depression against health status with 95% confidence intervals. Reference class: non-caregivers. In bold: statistically significant estimates at the 5% level.

Statistical model. Suitable post-processing confirmed the proportional hazards assumption of the cumulative logit model. In the final model fit, all demographic and socio-economic variables act as confounders. In addition, the effect of the macro-area is modified by our risk factor and by economic status, by the presence of chronic diseases and by employment status. This means that the relationship between depression and caregiving activity changes depending on where one lives, as initially conjectured. Furthermore, in Eastern Europe, being economically disadvantaged seems to weigh less than in the other three areas, while being employed and suffering from at least one chronic disease raises the odds ratio of depression. The latter is also true in Southern Europe. The parameter estimates, together with their standard errors and p-values of significance, are reported in Table A.1 of the Appendix.

Subject profiling. The worst profile corresponds to a woman aged more than 70, partnerless, disadvantaged, not employed, with a low education level, suffering from at least one chronic disease and who doesn't practice any physical activity. Figure 3 shows how the probabilities of severe depression, mild depression and no depression vary for this subject profile according to our risk factor and the geographical area. We can identify a distinct North-South gradient for the mental well-being of caregivers. People in Northern Europe are globally less depressed, most likely thanks to the more efficient welfare policies of these countries. In particular, in Northern Europe assisting people outside one's own family nucleus does not affect the psychological well-being. Indeed, for this category of caregivers the probability of depression is even lower than the same calculated for the non-caregivers. Caregivers from Central Europe share a similar health status with people living in Eastern countries. The worst situation is experienced by caregivers from Southern Europe who, in particular, suffer of the highest probability of severe depression.

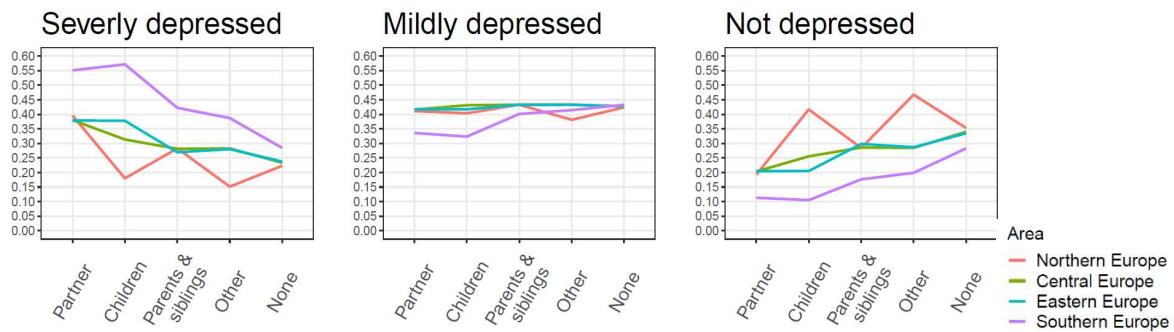


Figure 3. Probability of depression as estimated with the cumulative proportional hazards model for varying caregiver and macro-area. Confounders and effects modifiers are set to the worst profile. Left: severe depression. Middle: mild depression. Right: no depression⁵.

SUGGESTIONS FOR FUTURE STUDIES

Three aspects of our research are worth of further attention. Firstly, exploiting the longitudinal nature of the SHARE survey would provide insight on how the duration of the caring activity impacts on the psychological well-being of the carer. Secondly, accounting for the level of depression pre to the beginning of caregiving would allow us to trace the evolvement of depression on caregivers. Lastly, our model can easily be extended to analyze mental health problems during the current Covid-19 pandemic.

CONCLUSION

In conclusion, being a caregiver definitely worsens mental health by increasing one's risk of depression. The most weighty caregiving activity is taking care of children, followed by, in decreasing order of heaviness, looking after a partner, after parents and siblings and after other people. These results are coherent with our expectation given the emotional connection between caregivers and the person they take care of. We furthermore found statistical evidence of a North-South gradient for the impact on the psychological well-being: caregivers from Southern Europe are in general more depressed, followed by caregivers from Central and Eastern Europe who appear to be similar to each other. The best situation is faced by caregivers from Northern Europe, most likely thanks to their highly developed welfare policies.

REFERENCES

- Bookwala, J. (2009). "The impact of parent care on marital quality and well-being in adult daughters and sons." *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, **64(3)**:339–347.
- Brenna, E., Di Novi, C. (2016) "Is caring for older parents detrimental to women's mental health? The role of the European North-South gradient." *Rev Econ Household* **14**, 745–778.
- Silverstein, M., Gans, D. and Yang, F. (2006). "Intergenerational support to aging parents: the role of norms and needs." *Journal of Family Issues*, **27(8)**, 1068–1084.
- Prince, M. J., Reischies, F., Beekman, A. T. F., et al. (1999). "Development of the EURO-D scale – a European Union initiative to compare symptoms of depression in 14 European centres." *The British Journal of Psychiatry*, **174(4)**, 330–338.

APPENDIX

Logistic procedure for the cumulative logit proportional hazards model:

```
proc logistic data = d.wave5;
    class age (ref = "Less than 60") sex (ref = "Male")
        education (ref = "University") civ_stat (ref = "Partnered")
        economic (ref = "Advantaged") chronic_dis (ref = "Absent")
        physical_act (ref = "Yes") employment (ref = "Not employed")
        country (ref = "1_North_Europe") carer (ref = "5_No") /
        param = ref;
    model eurod (event = "1_Depressed") = age sex education civ_stat
        economic chronic_dis physical_act employment
        carer country
        carer * country age * country carer * age
        education * country economic * country
        chronic_dis * country physical_act * country
        carer * employment
        carer * sex/ selection = stepwise;
    format age age. sex sex. education edu. civ_statcivstat.
        economic economic. chronic_dis chronic.
        physical_actphysact. employment emp. country area.
        carer who. euroddepord. ;
    output out = d.pred_ord_area predprobs = individual;
run;
```

Coefficient of		Value	Value	Estimate	Standard error	P-value
Health status	Severly Depressed			-4.0258	0.1338	<.0001
	Mildly depressed			-2.1724	0.1312	<.0001
Age	[60-70)			-0.7156	0.1155	<.0001
	≥ 70			-0.5639	0.1238	<.0001
Sex	Female			0.6531	0.0326	<.0001
Education	Low			0.3188	0.0446	<.0001
	Medium			0.0750	0.0449	0.0947
Civil status	Partnerless			0.4792	0.0396	<.0001
Economic status	Disadvantaged			0.9152	0.0933	<.0001
Chronic diseases	Present			0.3795	0.0891	<.0001
Physical activity	No			0.5965	0.0335	<.0001
Employment status	Employed			-0.4519	0.1151	<.0001
	Partner			0.8252	0.1094	<.0001
Caregiving activity	Children			-0.2693	0.3346	0.4210
	Parents and siblings			0.3168	0.3012	0.2928
	Other			-0.4754	0.4191	0.2567
	Central Europe			-0.1312	0.1480	0.3754
Macro-area	Southern Europe			-0.6222	0.1615	0.0001
	Eastern Europe			-0.5376	0.1646	0.0011
Macro-area: Caregiving activities	Central Europe	Partner		-0.1242	0.1306	0.3418
		Children		0.6748	0.3835	0.0784
		Parents and siblings		-0.0641	0.3737	0.8637
		Other		0.7345	0.4788	0.1250
	Southern Europe	Partner		0.3013	0.1439	0.0363
		Children		1.4800	0.3682	<.0001
		Parents and siblings		0.2919	0.3380	0.3878
		Other		0.9375	0.4659	0.0442
	Eastern Europe	Partner		-0.1548	0.1370	0.2586
		Children		0.9368	0.3838	0.0147
		Parents and siblings		-0.1477	0.3472	0.6706
		Other		0.6984	0.4859	0.1507
Age:Macro-area	[60-70)	Central Europe		0.2427	0.1397	0.0824
		Southern Europe		0.5315	0.1493	0.0004
		Eastern Europe		0.4625	0.1475	0.0017
		Central Europe		0.2633	0.1479	0.0751
	≥ 70	Southern Europe		0.7180	0.1561	<.0001
		Eastern Europe		0.6643	0.1551	<.0001
		Central Europe		-0.1887	0.1109	0.0890
		Southern Europe		-0.0329	0.1128	0.7703
Economic status: Macro-area	Disadvantaged	Eastern Europe		-0.3530	0.1124	0.0017
		Central Europe		0.1142	0.1062	0.2821
		Southern Europe		0.2621	0.1154	0.0232
		Eastern Europe		0.3073	0.1172	0.0088
Chronic diseases: Macro-area	Present	Employed	Central Europe	0.00209	0.1409	0.9882
		Employed	Southern Europe	-0.1311	0.1560	0.4006
		Employed	Eastern Europe	0.2672	0.1515	0.0778

Table A.1. Parameter estimates for the cumulative logit model with proportional hazards fitted with PROC LOGISTIC. In bold: statistically significant estimates at the 5% level.