

Online Supplementary Materials - Appendix B

Evaluation of meaningful effects for men and women

Our work was focused on the increase in sadness/depression compared with one's condition before the beginning of the COVID-19 pandemic rather than on just reporting people's state of being sad/depressed during the interview period. Therefore, it was difficult to find previous studies in this research area working with Average Marginal Effects (AMEs) that might suggest the evaluation of *meaningful* effect sizes. Consequently, we applied global conventional benchmarks. To do this, we needed to take the nature of the phenomenon under investigation and the scale of our dependent variable into account. More specifically, as the interest was in the increase of the probability to be sadder or more depressed, we needed to evaluate when such increment can be considered as *meaningful*.

The standard SHARE questionnaire collects the data required to construct the EURO-D mental scale (Prince et al., 1999), which is a validated depression scale comprising 12 items (depression, pessimism, wishing death, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness). It is a symptom-oriented, self-assessed scale that determines the presence of various depressive or emotional manifestations on a scale from 0 (the lowest level of depression) to 12 (the highest). In particular, a EURO-D score of 4 or more is used to discriminate between *depressed* and *non-depressed* individuals (Dewey and Prince, 2005).

We exploited this EURO-D scale (in its ordinal version) to assess the *meaningful* effect. Indeed, both the depression categories of Gennaro et al. (2021) used as covariates in our work and the dependent variable of our models were strictly related to the EURO-D scale: the depression categories derived from the information collected on all EURO-D items, while our dependent variable was constructed from the answers to the first item of this EURO-D scale.

Separately by gender, we estimated a logistic model of our dependent variable on the EURO-D scale (based on the SHARE wave 6 data) and calculated probabilities of the outcome at different values of such scale.

According to the EURO-D measure, increasing the score from 3 to 4 means moving from a *non-depressed* to a *depressed* classification. Likewise, a *meaningful* effect may be obtained comparing the predicted probabilities of being sadder/more depressed at the scores of 3 and 4: to highlight substantive significance of the estimates (Rainey, 2014; Bernardi et al., 2016), the *meaningful* effect was computed as the difference between the upper bound of the confidence interval at 95% level for the predicted probability at score 4 and the lower bound of the confidence interval at 95% level for the predicted probability at score 3.

Main results are reported in Table B1. According to these estimates, we assessed a *meaningful* effect of 3.85 percentage points for men and 5.38 percentage points for women (rounded to 3.9 and 5.4 percentage points, respectively).

Taking our procedure and the nature of the phenomenon under investigation into account, the interest was addressed to move from a *non-depressed* classifi-

cation to a *depressed* condition. Therefore, it was difficult to define and evaluate small, medium, and large effects as in the Cohen (1988) recommendations. For this reason, one effect was just reported.

Table B1: Predicted probabilities at the scores of 3 and 4 of the EURO-D depression scale after the estimation of a logistic model of reporting to be sadder/more depressed during the first wave of the COVID-19 pandemic than before the outbreak

Predicted probability of score	Males (15,538 units)		Females (21,937 units)	
	Estimate	CI 95%	Estimate	CI 95%
3	.1124	[.1004 .1243]	.1794	[.1670 .1917]
4	.1236	[.1083 .1389]	.2066	[.1924 .2208]

References

- Bernardi, F., L. Chakhaia, and L. Leopold (2016). ‘sing me a song with social significance’: The (mis)use of statistical significance testing in european sociological research. *European Sociological Review* 33, 1–15.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). New York, NY: Academic Press.
- Dewey, M. and M. Prince (2005). Mental health. In A. Börsch-Supan, A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist, and G. Weber (Eds.), *Health, Ageing and Retirement in Europe. First Results from the Survey on Health, Ageing and Retirement in Europe*, pp. 108–117. Mannheim: MEA.
- Gennaro, C., O. Paccagnella, and P. Zaninotto (2021). A model-driven approach to better identify older people at risk of depression. *Ageing & Society* 41(2), 339–361.
- Prince, M., F. Reischies, A. Beekman, R. Fuhrer, S. Jonker, C. Kivelä, B. Lawlor, A. Lobo, H. Magnusson, I. Fichter, H. van Oyen, M. Roelands, I. Skoog, C. Turrina, and J. Copeland (1999). Development of the euro-d scale - a european union initiative to compare symptoms of depression in 14 european centres. *British Journal of Psychiatry* 174(4), 330–338.
- Rainey, C. (2014). Arguing for a negligible effect. *American Journal of Political Science* 58, 1083–1091.