Research Question 1: Binary Logistic Regression

Objective of analysis

This study examines the impact of family status, job type, and age on depression among healthcare workers during the COVID-19 pandemic, using a binary logistic regression model. Additionally, potential interactions are explored to determine whether the effect of job type depends on the level of family status. Family status reflects the influence of social support on mental health, while job type highlights differences in stress and responsibilities between nurses and physicians (**?**). By including the interaction term, the study aims to determine whether the protective or risk factors associated with family status differ for nurses and physicians, providing insights for more tailored mental health interventions for healthcare workers.

The data, derived from **?**, was collected through a cross-sectional study between July and November 2022 in Spain. The dataset is available in the file 'depression_data.csv' and includes 2685 observations. Participants ranged in age from 18 to 73 years ($Median = 38$; $Mean = 39.2$; $SD = 13$). **Age** was treated as a continuous variable due to its quantitative nature. **Family status** was treated as a categorical variable with four groups: **Single** (25.4%; $n = 682$), **Married** (57.8%; $n = 1553$), **Divorced** (11.7%; $n = 314$), or **Widowhood** (5.1%; $n = 136$). The categorical treatment allows for capturing group differences and their interaction with job type. **Job type** was also treated as a categorical variable, with participants categorized as **Nurses** (64.8%; $n = 1739$), or **Physicians** (35.2%; $n = 946$). The dichotomous nature of this variable reflects the distinct workplace roles and stressors associated with each job type (Figure **??**). **Depression**, the binary outcome variable, was coded as 0 (Absent) and 1 (Present), with 17.7% ($n = 474$) participants reported depression and 82.3% ($n = 2211$) not reporting it. The mean prevalence of depression in the sample was 0.18 ($SD = 0.38$).

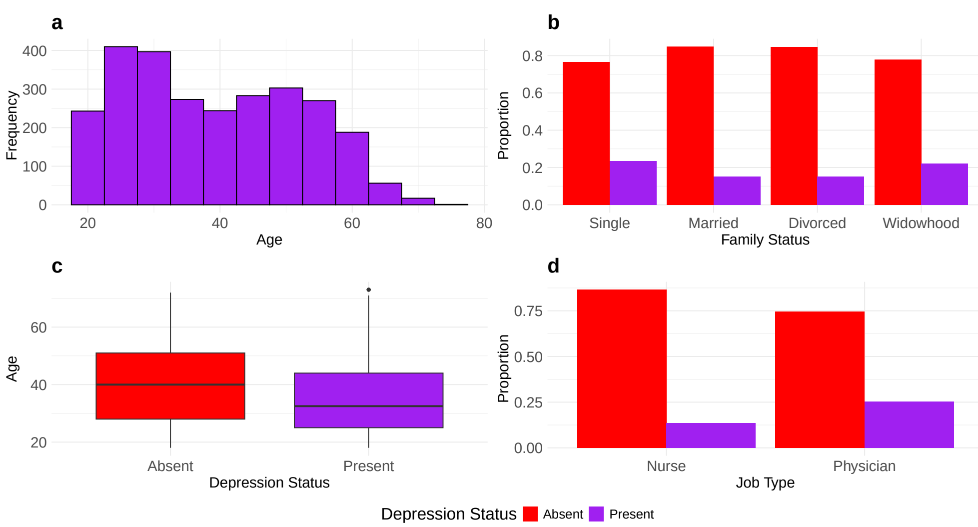

Figure 1: (a) The Age Distribution of Participants, (b) Proportions of Family Status, (c) Age by Depression Status, (d) Proportions by Job Type.

Statistical model

Binary logistic regression is a Generalized Linear Model used to model binary outcomes, making it ideal for predicting depression (*absent vs. present*). It estimates the probability of depression as a function of predictor variables while ensuring the predicted probabilities remain within the valid range of 0 to 1. This is achieved through the logit link function, which transforms probabilities into log-odds to linearize the relationship between predictors and the outcome. The formula for the logistic regression model is as follows:

$$y_i \sim \text{Bernoulli}(\theta_i), \quad \text{logit}(\theta_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}, \quad \text{for } i = 1, \ldots, n.$$

where y_i is the binary outcome for individual i, θ_i represents the probability of depression i, $\text{logit}(\theta_i) = \ln\left(\frac{\theta_i}{1-\theta_i}\right)$ is the log-odds of depression for individual i, β_0 is the intercept, representing the log-odds of depression when all predictors are at their reference values, and $\beta_k x_{ki}$ represents the effect of predictor k (e.g., age, job type, family status) on the log-odds of depression. For instance, a positive β_k increases the log-odds, while a negative β_k decreases it.

Logistic regression generalizes linear regression by allowing non-normal response variables and incorporating appropriate link functions. Unlike general linear models, which assume normally distributed outcomes, logistic regression uses the Bernoulli distribution for binary outcomes, ensuring accurate probability estimation through odds ratios. In contrast, multilevel logistic regression accounts for hierarchical structures with random effects but is unnecessary due to independent observations and non-clustered data. Odds ratios, derived by exponentiating coefficients, quantify the effect of predictors on depression likelihood, by measuring the relative change in the odds of depression associated with each predictor. For example, if job type has a positive coefficient, it implies an increase in the odds of depression for a specific job type compared to the reference category. The logistic regression model is implemented in R using glm():

```r
bm_model <- glm(depression ~ age + job * family_status, data = depression_df,
                family = binomial)
```

The linearity assumption was assessed visually. Non-linearity for age was observed, justifying the inclusion of a quadratic term (Age^2), aligning with prior research (**?**). Variance Inflation Factor analysis confirmed no significant multicollinearity among predictors.

Model Comparison

```r
bm_null_model <- glm(depression ~ 1, data = depression_df, family = binomial)
bm_model_linear <- glm(depression ~ age + job * family_status,
                       data = depression_df, family = binomial)
anova(bm_model_linear, bm_model_quadratic, test = "Chisq")
```

Table 1: Model Comparison Results.

Term	Deviance	AIC	LRT	P-Value
Full Model	2366.36	2386.36	—	—
Age (Linear Term)	2368.22	2386.22	1.86	0.17
Age (Quadratic Term)	2366.48	2384.48	0.12	0.73
Job: Family Status	2372.13	2386.13	5.77	0.12

Table **??** presents the model comparison results revealing a minimal improvement in fit with the quadratic term ($AIC = 2384.48$, $p = 0.73$), indicating a limited impact on depression outcomes. Similarly, adding interaction terms between job type and family status led to slight increases in deviance and AIC. Although the interaction term did not significantly affect the model, it helps capture subtle relationships between job roles and family support, aligning with prior research on their relevance to mental health (**??**). The model's overall explanatory power, as indicated by McFadden $R^2 = 0.05$ (5.46%), reflects a modest improvement over the null model. Although the predictors did not significantly reduce deviance ($\chi^2(1) = 0.12$, $p > 0.05$), they contribute to understanding potential factors influencing depression and justify further investigation.

Model Evaluation and Interpretation

Table 2: Logistic Regression Results.

Predictor	Estimate (Beta)	Odds Ratio	Std. Error	Z-Value	P-Value	95% CI
Intercept	-0.28	0.75	0.54	-0.52	0.602	[0.26, 2.16]
Age	-0.04	0.96	0.03	-1.37	0.17	[0.9, 1.02]
Age (Quadratic)	0.00	1.00	0.00	0.34	0.734	[1, 1]
Job: Physician	0.51	1.67	0.19	2.74	0.006**	[1.16, 2.41]
Family Status: Married	-0.40	0.67	0.17	-2.36	0.018*	[0.48, 0.94]
Family Status: Divorced	-0.48	0.62	0.27	-1.73	0.083	[0.35, 1.05]
Family Status: Widowhood	0.50	1.64	0.31	1.59	0.112	[0.87, 2.97]
Physician X Married	0.37	1.44	0.24	1.55	0.122	[0.91, 2.3]
Physician X Divorced	0.86	2.36	0.38	2.26	0.024*	[1.13, 4.99]
Physician X Widowhood	0.45	1.57	0.48	0.92	0.355	[0.6, 4.06]

Note. Significance levels: $*** \ p < .001$, $** \ p < .01$, $* \ p < .05$.

Table **??** presents the logistic regression results. The intercept (OR = 0.75) represents the baseline likelihood of depression, but since an age of 0 is not realistic, it serves as a reference point. The intercept is not statistically significant ($p = 0.6$), indicating no meaningful difference in depression odds at the baseline level of predictors. **Job type** emerged as the most significant factor, with physicians having significantly higher odds of depression compared to nurses. The estimate for job type (physician) is 0.51 with an odds ratio of 1.67 ($p < 0.001$), indicating a 67.21% increase in the odds of depression for physicians. **Family status** also influenced depression risk. Married individuals exhibited reduced odds of depression compared to singles. The estimate for married individuals is -0.4, with an odds ratio of 0.67, indicating a -33.03% reduction in the odds of depression for married individuals ($p = 0.02$). Overall, this suggests that marriage provides a protective effect against depression. However, widowhood showed that while the odds of depression were nearly doubled (OR = 1.64), the confidence interval includes 1, aligning with a non-significant p-value ($p = 0.1$). This suggests no conclusive evidence for an effect. Divorced individuals did not exhibit a statistically significant difference ($p = 0.08$). Additionally, **Age** showed no significant linear effect on depression. The estimate for age is -0.04, and the odds ratio is 0.96, which suggests that for each year increase in age, the odds of depression decrease by about 4.07%. However, this result is not statistically significant ($p = 0.17$). Similarly, the quadratic term for age (OR = 1) did not significantly improve model fit ($p = 0.73$).

The **interaction** between job type and family status revealed that Married physicians significantly exhibited lower odds of depression compared to single nurses. The estimate for this interaction term is 0.37, and the odds ratio is 1.44, suggesting that married physicians have 44.32% lower odds of depression compared to single nurses. However, the confidence interval includes 1 (95% CI: [0.91, 2.3]), indicating no significant interaction effect ($p = 0.12$). For the interaction effect with widowhood, the odds ratio is 1.57, and the confidence interval for this interaction is quite wide, indicating limited precision. The interaction is not statistically significant ($p = 0.36$). Finally, for divorced physicians, there is a significant increase in the odds of depression compared to single nurses (OR = 2.36, $p = 0.02$), indicating that being a physician and being divorced increases the likelihood of depression compared to being a single nurse.

The standard errors for significant predictors, such as physician ($SE = 0.19$) and married ($SE = 0.17$), were relatively small, reflecting precise estimates. However, for many predictors like the interaction between physician and widowhood ($SE = 0.48$), the larger standard error may contribute to its lack of significance, indicating greater uncertainty in its effect. The model's explanatory power is evidenced by a substantial reduction in deviance from 2503 (null deviance) on 2684 degrees of freedom to 2366.4 (residual deviance) on 2675 degrees of freedom. This reduction indicates that the included predictors substantially improve the model's fit, demonstrating their collective relevance in explaining depression risk among healthcare workers.

Model Predictions

```
predict(bm_model_quadratic, newdata = prediction_data, type = "response")
```

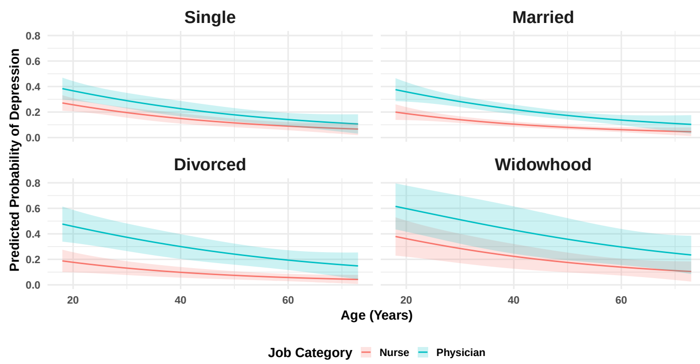

Figure 2: Predicted Probability of Depression by Job, Age, and Family Status.

Figure **??** illustrates that physicians consistently show higher depression probabilities than nurses, likely due to the increased workplace stress, highlighting the need for targeted mental health support for physicians. The non-linear relationship between age and depression is evident, with younger workers likely affected by early-career stress and older workers facing burnout or health-related challenges. Additionally, married individuals display lower probabilities of depression, suggesting protective effects of social support. In contrast, widowed individuals have higher predicted probabilities, warranting tailored emotional and social support programs for this group. Confidence intervals highlight the precision of predictions, with narrower intervals for married individuals and wider intervals for widowed individuals, reflecting greater uncertainty for some subgroups.

Conclusion

This analysis identifies job type as a key predictor of depression, with physicians at higher risk than nurses, suggesting the need for tailored mental health interventions. Marital status appears protective, highlighting the importance of social support systems in mitigating depression. However, several limitations must be noted. The cross-sectional design limits causal inference, as relationships observed cannot determine temporal effects. Unmeasured confounders, such as workplace overload variability or individual resilience, may influence depression risk. Additionally, the homogeneity of the sample, limited to healthcare workers in Spain, reduces generalisability to other occupations or cultural contexts. The findings provide actionable insights for developing targeted policies, including stress management programs tailored to physicians, mentorship initiatives for early career workers, resilience training for healthcare staff, and strengthening social support networks within healthcare organisations. These strategies could address disparities in mental health risk and improve overall workplace well-being. Future studies should evaluate the effectiveness of targeted intervention and refine strategies through subgroup analyses to better address the diverse needs of healthcare workers.