**STUDY Proposal**

Investigating the effect of variables SS and PS on Variable ALI

Dependent variable: Variable ALI

Independent variable: Variables SS & PS

Covariates: Variables Ag, LOS, C19, Rank, PA

Moderating Variable: SF1, MS1

Independent variable: SS (continuous) & PS (continuous)

Covariates:

Primary: Age , LOS (continuous)

Secondary: C19 (Categorical) , PA(categorical), Rank (categorical)

Dependent variable

ALI (continuous variable)

Moderating variable:

Primary: SF (continuous)

Secondary: MS (continuous)

**Study overview**

The primary aim of our study was to provide an improved understanding of how the psychosocial stressors (PS & SS) faced by individuals X may become biologically embedded over time in such a way as to increase their risk of developing adverse health outcomes (ALI). Hence we will explore the relationship between stress (PS & SS) and ALI.

In addition, we would account for factors such as age and length of service on the relatiotionship between psychosocial stress and ALI because these are factors that have shown the strongest relationship wit ALI in previous studies.

Further, we believe certain demographic factors of our occupational group may need to be accounted for depending on their relationship with ALI. They include education and their occupational rank. Other behavioral habits such as MJ (marijuana use), physical activity (PA), alcoholi intake (AIF), and Sleep (Slp) were also considered to investigate if they significantly correlated with ALI.

Also, given that Covid-19 (Covid-19) happened between the two phases of data collection we hope to see whether this factor played a role between change in stress-ALI relationship from time 1 to 2

Finally, we hoped to examine the moderating effect of social support (MS) and (SF) may on the psychosocial stress and allostatic load relationship.

**Analytical plan**

We performed all statistical analyses using either SPSS statistical software for macOS (version 25; IBM SPSS, 2017) or R Studio (version 3.4.1; RStudio Inc, 2017).

For descriptive analyses, means and standard deviations (SD) were calculated for continuous variables, while frequency (*n*) and percentages were determined for categorical variables. For bivariate analysis, Pearson's correlation coefficient was used to investigate the association between age, LOS, and the allostatic load index. The second set examined the association between age, LOS, stress instrument scores, and social support instruments. Two-tailed, independent Student's t-tests were used to compare mean allostatic load index (ALI) differences between 2-item categorical variables (gender, physical activity). In addition, one-factor analysis of variance (ANOVA) was used to compare mean ALI differences across three or more item categorical variables (alcohol intake, time spent sleeping each night).

Subsequently, we conducted linear regression analysis to test our hypothesis. Using a mixed linear regression model with random effect. A time variable (1) was used to represent phase 1 and variable (2) for phase 2. Then we conducted an unconditional growth model to determine how similar the ALI remains over time; hence, determine the ICC. Secondly, we conducted a between-subject model (conditional growth model) to determine the effect of time on the ALI. Finally, we conducted regression analyses consisted test the relationship between ALI as the outcome variable and psychosocial work-related stress (PSS and SOOS-14) as the predictor variables while adjusting for length of service, age, and level of perceived social support (moderating variable). Given the complexity of the model, if the sample size is not able to power the models, we may use only the baseline data (phase 1 data)of the predictor variables and covariates to investigate ALI at follow-up.

**Linear mixed models (or suggested best model, maybe simple T-testing especially if the sample size cannot carry the proposed regression model)**

* The main aim here is to investigate if the SS and PS predict the ALI. Also predict change over time. I am open to dropping which ever of the primary predictors don’t do a better job of predicting ALI. So I will suggest you run them independently of each other before considering both together in the same model.
* Secondly we did bivariate analysis to see which of the time 1 or 2 covariates, age, LOS, rank, C19, PA, AIF, and Slp associate with the dependent variable.
* When I initially tried to run both time 1 and 2 measures of all the predictors and covariates, the model couldn’t carry it. So we decided to run time 1 (PS and/or SS), run age and LOS at time 2, run C19 (covid, yes or NO), PA (physical activity), and rank. We didn’t add the other variables because they didn’t correlate with ALI ( Please run this to confirm).
* Based on the best predictors and covariates, we can develop a model that fits and can produce findings.

**Analyses to run**

**First stage.** Data cleaning. checks of data integrity (Preprocessing)

- Test assumptions

- Missing data (amount, patterns)

**Second stage** - Bivariate analysis to investigate associations and potential predictors to add to the regression analysis. Pearson’s, T-testing, and ANOVAs (3-categorical)

**Third stage**

1st stage = Within-subject model = The unconditional means model. Determine the ICC.

2nd stage = Between-subject model = Unconditional growth model.

3rd stage = Regression analysis (here, we will run variables to check for model fit. Most likely, we will end up using phase 1 data to predict changes in the outcome variable due to sample size)

4th stage = Moderation analysis. The primary analysis here is to determine if SF moderates the relationship between PS/SS and ALI. You should run MS also, but if it does not find anything we can remove it from the model.

Interacting terms:

SF1,

MS1

Dependent variable

ALI

Independent variable

SS and PS