Preregistration Form: Analysis Plan

Caspar J. van Lissa

03 Jan 2021

1. Title

Many Analysts Replication Project

2. Analysis Team

Surname	Name	Affiliation	Email.address
Van Lissa	Caspar J.	Utrecht University	c.j.vanlissa@uu.nl

3. Research Question

a. Do religious people have higher well-being?

This RQ can only be answered straightforwardly if there is no interaction between religiosity and perceived cultural norms.

b. Does the relationship between religiosity and well-being depend on perceived cultural norms of religion?

This research question must actually be answered first. If the answer is yes, then RQ 3a is a moot point; the effect then depends on perceived cultural norms.

4. Hypotheses

List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.

I have no specific hypotheses. I will perform and interpret two-tailed tests of the random intercept and slope of the effects of 1) the interaction between religiosity and perceived cultural norms (RQ 3b) and 2) religiosity on well-being. The random slopes indicate whether these effects vary accross country.

5. Variables

a. Dependent variable(s)

State which key dependent variable(s) you will use in your analysis. Name the specific column names of these variables as stated in the data documentation.

For the dependent variable wellbeing, we will compare a one-factor model to a three-factor model. We would prefer a simpler one-factor model, but if the fit of the three-factor model is notably better (subjective decision), we will treat the subscales of wellbeing as three correlated DVs. The latent variables are constructed using only the items as indicators, not the general wellbeing scores or the mean scale scores.

```
cfa_well <- "well =~ wb_phys_1 + wb_phys_2 + wb_phys_3 + wb_phys_4 + wb_phys_5
+ wb_phys_6 + wb_phys_7 + wb_psych_1 + wb_psych_2 + wb_psych_3 + wb_psych_4
+ wb_psych_5 + wb_psych_6 + wb_soc_1 + wb_soc_2 + wb_soc_3"
fit_cfa_well <- cfa(cfa_well, df)
cfa_well3 <- "phys =~ wb_phys_1 + wb_phys_2 + wb_phys_3 + wb_phys_4 + wb_phys_5
+ wb_phys_6 + wb_phys_7
psych =~ wb_psych_1 + wb_psych_2 + wb_psych_3 + wb_psych_4  + wb_psych_5 +
wb_psych_6
soc =~ wb_soc_1 + wb_soc_2 + wb_soc_3"
fit_cfa_well3 <- cfa(cfa_well3, df)
library(semTools)
compareFit(well = fit_cfa_well, well3 = fit_cfa_well3)
```

b. Predictor variable(s)

State which predictor variable (including moderators and covariates) you will use in your analysis. Name the specific column names of these variables as stated in the data documentation.

For the main predictor, religiosity, we will conduct parallel analysis to determine the dimensionality of the construct. If parallel analysis suggests >1 factors, we will conduct an EFA for the suggested number of factors, and construct a CFA model based on the pattern of factor loadings. We will then compare a one-factor model to the EFA-based model. We would prefer a simpler one-factor model, but if the fit of the EFA-based model is notably better (subjective decision), we will treat the subscales of religiosity as separate IVs.

```
cfa_rel <- "rel =~ rel_1 + rel_2 + rel_3 + rel_4 + rel_5 + rel_6 + rel_7 + rel_8 + rel_9"
```

c. Indices

If applicable, please define how measures will be combined into an index (or even a mean) and what measures will be used. Include either a formula or a precise description of the method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), please note that here but describe the exact method in the analysis plan section.

```
library(lavaan)

cfa_rel <- "religiosity =~ rel_1 + rel_2 + rel_3 + rel_4 +
rel_5 + rel_6 + rel_7 + rel_8 + rel_9"
cfa(cfa_rel, df)

cfa_cnorm <- "cnorm =~ cnorm_1 + cnorm_2"
# Note that this model is not identified as such.
# If it leads to problems in the larger measurement model,
# we will use the mean score as observed variable.

# Well-being
# Will use only the indicators, for three subscales separately.
cfa_well <- "phys =~ wb_phys_1 + wb_phys_2 + wb_phys_3 +
wb_phys_4 + wb_phys_5 + wb_phys_6 + wb_phys_7
psych =~ wb_psych_1 + wb_psych_2 + wb_psych_3 + wb_psych_4
+ wb_psych_5 + wb_psych_6
soc =~ wb_soc_1 + wb_soc_2 + wb_soc_3"
```

```
cfa(cfa_well, df)
# Note that wb_soc_3 might be dropped if the loading is poor;
# religious participants might be less likely to report sexual satisfaction,
# leading to MNAR
```

6. Analysis Plan

For the planned analysis script, see `analysis.R`. Importantly, I will split the data into a 70% training sample and 30% testing sample using a preregistered random seed. Thus, which cases are used for training and which for testing is already predetermined before I have received the data.

The reason for creating separate training and testing samples is that, despite my best intentions to to follow the preregistered analyses as closely as possible, adjustments may be deemed necessary after seeing the data. For example, certain assumptions might be violated, syntax might be mis-specified, or models may fail to converge. The training sample will be used to make such necessary adjustments to the planned analyses. After the definitive models have been developed using the training sample, the testing sample will be used only to obtain unbiased estimates of model fit. This prevents us from overfitting or p-hacking the training sample.

First, we will definitively decide upon the set of models to be compared based on the training sample. Then, these models will be evaluated on the 30% testing sample to obtain unbiased estimates of model fit. We will select the best-fitting model (in the testing sample) using a combination of BIC statistics, to rank the models by fit, and likelihood ratio tests. For the BIC statistic, a lower number indicates better fit. The preference would thus generally be for the model with the smallest BIC. However, sometimes the difference between models might be negligible. Therefore, we complement the BIC with likelihood ratio tests. If two models both have low BICs, the more parsimoneous model would be preferred. If this model does not differ significantly from the one with the lowest BIC, then we will select the more parsimoneous one. The selected final model will be evaluated on the full sample to obtain as much power as possible, and the results of that analysis are used for inference.

Some exploratory analyses may be conducted in lavaan, which (running in R) is more convenient than Mplus. Because the main analyses involve a multilevel model with random slopes, I will use Mplus for these (random slopes are not available in lavaan).

a. Statistical models

What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, RMANOVA, MANOVA, multiple regression, SEM, etc) and the specification of the model. This includes each variable that will be included, all interactions, subgroup analyses, pairwise or complex contrasts, and any follow-up tests from omnibus tests. Provide enough detail so that another person could run the same analysis with the information provided. Remember that in your final article any test not included here must be noted as exploratory and that you must report the results of all tests.

Note: This is perhaps the most important and most complicated question within the preregistration. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? Be aware for instances where the statistical models appear specific, but actually leave openings for the precise test.

The general modeling framework is a multilevel model, because the data are nested (respondents within countries). Both research questions are at the individual level, however. Therefore, I do not estimate a model at the country level. Instead, I estimate random intercepts and slopes for the effects of religiosity and the interaction between religiosity and perceived cultural norms on wellbeing. The random intercept informs us about the average effect of these predictors on wellbeing across countries, and the random slope (optionally) informs us about the variance of these effects across countries.

See the script `analysis.R`. The coefficients used for inference are:

```
  "[sreli];",
  "[sint];",
  "sreli;",
  "sint;",
```

If the final model is not the model with the interaction, the terms related to `sint` are omitted.

Exploratory analyses

In addition to the planned deductive analyses, I will conduct an inductive (exploratory) analysis using random forests, predicting a mean score of wellbeing from the individual-level predictors. This analysis should reveal what the most important predictors of wellbeing are in the dataset provided, and can therefore help contextualize the planned analyses. Is religion indeed important? Are perceived cultural norms? How important are they, compared to other measured variables?

We will examine predictive explained variance in the testing sample, variable importance, univariate partial dependence plots, and a bivariate partial dependence plot for the interaction between religion and perceived cultural norms.

b. Transformations

If you plan on transforming, centering, recoding the data, or requiring a coding scheme for categorical variables, please describe that process.

See the script `prepare_data.R`.

c. Inference criteria

What criteria will you use to make inferences? Please describe the information you'll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?

I will use the standard $p < .05$ criterion with two-sided tests for all inferences. I will not account for multiple comparisons.

d. Data exclusion

How will you determine which data points or samples if any to exclude from your analyses, for instance, based on the attention check or missing data? How will outliers be handled?

I will screen for data out of range on the observed variables. Such values will be coded as missing. Outliers will be included in the analysis.

For the main analysis, I will exclude participants who failed the attention check. I will conduct a sensitivity analysis, re-running the final model with all participants included.

e. Missing data

How will you deal with incomplete or missing data? Note: For the well-being and religiosity measures there are no missing data. See the data documentation for a more detailed overview of missing data for each variable.

I will use single imputation using random forests, using the `missForest` algorithm by Stekhoven & Bühlmann (2012), in the `missRanger` implementation.

Stekhoven, D.J. and Buehlmann, P. (2012). 'MissForest - nonparametric missing value imputation for mixed-type data', Bioinformatics, 28(1) 2012, 112-118. https://doi.org/10.1093/bioinformatics/btr597.

7. Other

If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other helpful context would be appropriate here.

To make my project fully reproducible, I will use the Workflow for Open Reproducible Code in Science (WORCS; Van Lissa et al., 2020).

Van Lissa, C. J., Brandmaier, A. M., Brinkman, L., Lamprecht, A.-L., Peikert, A., Struiksma, M. E., & Vreede, B. (in press). WORCS: A Workflow for Open Reproducible Code in Science. Data Science, in press. https://doi.org/10.17605/OSF.IO/ZCVBS