Replication of a Research Claim from Gerhold et al. (2020),
from PsyArXiv

Replication Team: Finders: Anirudh Tagat, Hansika Kapoor
Analyst: Jamie Field

Research Scientist: Nick Fox

Action Editor: Kevin Esterling

Independent Reviewers
(add name below when you initiate review, comment “DONE” on your name when you finish):

Reviewer #1: [Dino Krupić]

Reviewer #2: [NAME]

Reviewer #3: [NAME]

Review Period: July 28 - August 3

View-only links to: Original Paper, Replication Materials

Instructions for Data Analysts

The preregistration for this replication study was started by a separate team of researchers who were responsible for identifying data sources and constructing them into a replication dataset(s) for your use in the analysis. They have completed sections 1-13 of the preregistration below, and included additional materials in the OSF project that document how the dataset was constructed.

In cases where all of the underlying data sources were able to be freely shared and posted, the constructed dataset(s) have been posted to the OSF as well, which you are free to use in designing the analysis plan (see below for details). In cases where some or all of the data sources could *not* be freely shared or posted, the replication dataset(s) are not provided on the OSF. Rather, you will need to follow the instructions and code to first reconstruct the datasets, and then proceed with your work. In such cases, the team responsible for creating the dataset(s) has provided summary statistics in the OSF that correspond to the constructed datasets, so you can verify that the datasets you create match what they intended.

You'll be responsible for filling out sections 16-25 of the preregistration below. Before you do so, **please review the original study, sections 1-15 of the preregistration, and the materials provided on the OSF**, so that you are familiar with all of the decisions that have been made to date. In many cases, the 'data preparer' will have left you instructions and suggestions on how the provided data can be used in the analysis, as well as idiosyncrasies and discrepancies in the data that you should be aware of. The data preparers have tried to be thorough in including all variables that you might need, but please keep in mind the following:

- Some of the variables included in the constructed dataset(s) may not be needed in the final analysis, so please do not feel the need to necessarily use all of the provided variables.
- Some of the variables needed might have mistakenly been excluded from the constructed datasets. If you find that this is the case, please let Andrew or Anna know, and they will work with you to supplement the datasets as needed.

For these secondary data replications, we would like the analysis plan to be completed before the preregistration goes through review, so that after review, the only remaining steps are registration and running the analysis code on the full datasets. To facilitate that, we are asking that you include in section 19 a link to the code you will use that takes the constructed dataset(s) provided to you and produces the focal analysis (including all of the cleaning, merging, and transforming required). When developing your analysis plan and code, please randomly sample 5% of the data for use in your work, and **do not use the rest of the data until it is time to run the final analysis**. In section 19, you will find a statement that we are asking you to bold that confirms you've only used 5% of the data when developing and testing your code. If this approach will not work for any reason, please let Andrew or Anna know and disclose deviations from this plan somewhere in the preregistration.

- In cases where we are providing you a complete dataset, you can just sample out 5% of the observations and hold the rest out until you are ready to perform the final analysis.
- In cases where we are providing you multiple datasets that need to be combined prior to analysis, please sample out 5% of the observations in whatever way is most sensible.
    - For example, in cases where each dataset contains complete observations on its own (a typical 'row bind' situation), it makes the most sense to sample out 5% of each dataset separately and then combine them together to develop and test your code.
    - In cases where datasets need to be merged in order to create complete observations (a typical 'column bind' situation), it makes the most sense to merge the separate datasets

Commented [1]: This characterizes the Gerhold replication.

Commented [2]: OK.

into a full dataset first, and then sample out the 5% before proceeding with the rest of the analysis code.

- We leave the decision on how to sample out the random subset of data to you, so long as (a) you are not performing any analyses on the complete dataset until after your study is registered and (b) whatever decision you make is documented in the preregistration.

Finally, in cases where the replication data combines observations from the original study with observations that were not used in the original study (what we are calling 'hybrid replications'), please perform two analyses (details immediately below). This will likely require you to subset your data into the two groups described immediately below, based on the description of the original analysis provided in the study.

- When the 'new' data alone can clear the minimum power threshold, please perform one analysis that combines all available data, and a second that only uses the 'new' data. Please make sure both analyses are documented (with code) in section 19 below.
- When the 'new' data alone *cannot* clear the minimum power threshold, please perform one analysis that combines all available data, and a second that only uses the old data. Please make sure both analyses are documented (with code) in section 19 below.

Please contact Andrew or Anna if you have any questions. After you've completed the remaining sections of the preregistration and uploaded all the necessary materials to the OSF, please contact the SCORE coordinators regarding next steps.

Commented [3]: This does not apply to the Gerhold replication, so this paragraph can be ignored.

Commented [4]: OK

Preregistration of Gerhold_covid_Azg9

Existing Data Replication

Study Information

1. Title (provided by SCORE)

RR TEAM INSTRUCTIONS: *This has been determined by SCORE.*

Replication of a research claim from Gerhold (2020).

2. Authors and affiliations

RR TEAM INSTRUCTIONS: *Fill in the names and affiliations of your team below.*

Anirudh Tagat[1]
Hansika Kapoor[2]

1 Department of Economics, Monk Prayogshala, Mumbai
2 Department of Psychology, Monk Prayogshala, Mumbai

3. Description of study (provided by SCORE)

RR TEAM INSTRUCTIONS: *This description has been provided by SCORE. Please review and make a SCORE project coordinator aware of any edits, additions, and corrections you would suggest to the paragraph. You are free to add additional descriptions of your project in a separate paragraph.*

The claim selected for replication from Gerhold (2020) is that women are more concerned about COVID-19 than men. This reflects the following statement from the paper's abstract: "Women are more concerned about COVID-19 than men." This claim was tested by comparing agreement to the statement “The COVID-19 (Coronavirus SARS-CoV-2) worries me” between men and women, measured using a 5-point Likert Scale. Further details are not described in the original manuscript. The authors found that they are worried about COVID-19 in general (women=68.2%, men=55.7%, $p < .01$).

4. Hypotheses (provided by SCORE with possible Data Analyst additions)

RR TEAM INSTRUCTIONS: *The focal test for SCORE is indicated as H*. If you will test additional hypotheses (or use alternate analyses) that help you to evaluate the claim your replication/reproduction is testing, number them H1, H2, H3 etc. (You can place H* in the list wherever makes sense). Please make sure that any additional hypotheses are logical*

deductions/operationalizations of the selected SCORE claim or are necessary to properly interpret the focal H hypothesis. Research that is outside this scope should be described in a separate preregistration.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Are the listed hypotheses specific, concise, clearly testable, and specified at the level of operationalized variables?*
- *Are hypotheses identified as directional or non-directional, and, if applicable, have the direction of hypotheses been stated? (Example: "Customers' mean choice satisfaction will be higher in the CvSS architecture condition than in the standard attribute-by-attribute architecture condition.")*
- *Does the list of hypotheses/tests indicate whether additional hypotheses are taken from the original study or modified/added by the team?*

H*: Women will express more concern about COVID-19 than men.

Design Plan

5. Study type

NOTE: *The study type selected should be based on the data collected for the replication, and not necessarily the data used in the original study.*

- Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.
- **Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.**
- Meta-Analysis - A systematic review of published studies.
- Other

6. Blinding

RR TEAM INSTRUCTIONS: *Select any/all of the below that apply for your study by bolding them. You will give a longer description in the next question.*

- **No blinding is involved in this study.**
- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
- Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)
- Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

[QUESTION 6 - BOLD YOUR RESPONSE ABOVE]

7. Blinding

RR TEAM INSTRUCTIONS: *Since all existing data replications are based on data that has already been collected, in most cases it will not be necessary to comment on participant blinding. In the rare instance when an existing experiment is being re-analyzed for an existing data replication and blinding is a relevant consideration, please provide below any details regarding blinding that are important for a reviewer to be aware of.*

No blinding was involved to the secondary data collectors’ knowledge.

8. Study Design

RR TEAM INSTRUCTIONS: *Please describe how data was collected in the original study and how it compares to the data that was selected for the replication attempt. Explain why the data selected for the replication study is suitable for a replication and if any substantial deviations exist between the two.*

If the data used in the replication combines observations from the original study with new observations (e.g. if the data selected for the replication attempt comes from the same longitudinal survey as the original study), describe how 'original' and 'new' observations relate to each other and an estimate for what proportion of the final dataset's observations will be comprised of original vs. new observations.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify the unit of analysis?*
- *Does the preregistration provide sufficient detail about how the data selected for the replication attempt deviates from or is congruent with the data employed in the original study?*
- *Does the preregistration describe whether and how 'original' and 'new observations' are combined together for the replication dataset?*

The data was collected online from a German sample between 19 and 23 March, 2020. All respondents to the online survey were above 18 years of age and did not have COVID-19 (as far as they were aware). Of the 1300 respondents, 1242 remained in the analysis (the author only retained data from respondents who took more than seven minutes to complete the questionnaire), and all major regions of the country were represented, with the average age of participants being 46 years.

For replication, we propose to use data collected by Fetzer et al. (2020) that was deployed online starting March 20, 2020 and has at least 200 respondents from 58 countries. Their dataset also contains information on worries, as well as gender, age, and other demographic information. As of May 2020, the proposed replication dataset has 10,192 respondents to the online survey from Germany alone, but data on sub-national regions is not available. The main deviation from the original study is that the data on worries in Fetzer et al. (2020) does not explicitly ask for worries *regarding* COVID-19.

In the original study, the question asks to what extent respondents agree with the statement: "The COVID-19 (Coronavirus SARS-CoV-2) worries me." In the proposed replication dataset, the question asks the extent to which the following statement applies to the respondent: "I am nervous when I think about current circumstances." The other question that will be used is "I am worried about my health." Apart from this, there is no other deviation from the original study. Data from the original study and Fetzer et al. (2020) will not be combined for replication.

9. Randomization (free response)

RR TEAM INSTRUCTIONS: *If the variables used for this replication attempt were randomized, state how they were randomized, and at what level.*

NA. There is no publicly available information about randomization.

Sampling Plan

This section describes how the data sources for the replication were selected, how they were prepared into a replication dataset, and the number of observations that will be analyzed from these data. Please keep in mind that the data described in this section are the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

10. Existing data (multiple choice question, provided by SCORE)

1.1.1. Registration prior to creation of data
1.1.2. Registration prior to any human observation of the data
1.1.3. Registration prior to accessing the data
1.1.4. Registration prior to analysis of the data
1.1.5. Registration following analysis of the data

11. Explanation of existing data

NOTE: *For replications that rely on existing data sources, this question refers to the data that will be used for the replication analysis (i.e. the final replication dataset), and not (a) the data from the original study or (b) the data sources accessed to construct the replication dataset. Since no new data will be created for 'existing data replications,' 1.1.1 should never be selected. Since all analyses will occur after registration, 1.1.5 should also never be selected.*

The data from Fetzer et al. (2020) has been accessed and prepared to match the sample criteria (German sample) prior to registration. The variables were selected to match with the requirements of the focal claim to be replicated. These are two measures on worry related to COVID-19, and one variable on gender of the respondent.

12. Data collection procedures

RR TEAM INSTRUCTIONS: *Please describe the process for constructing the replication dataset in as much detail as you can. The sections below should be used to provide the following information:*

- *Which variables are needed from the original study to perform a good-faith, high-quality replication.*
- *Which data sources were used, why they were selected, any deviations between the original study design and the replication study design that these selections present, and the procedures used to access the data.*
- *Which of the variables from the original study are available in the replication data sources, including relevant details about each measure.*
- *The procedure for creating the replication dataset, in both narrative and script form.*
- *A data dictionary that documents each variable included in the replication dataset.*

In the sections below, please provide links to the original materials whenever possible -- including descriptions of the original datasets and corresponding codebooks. If materials can be shared on the OSF, please do so, and provide view-only links to those materials.

Specific points to keep in mind for reviewers:
- *Does the preregistration describe which data sources were selected for the replication study and why each is suitable?*
- *Does the preregistration make clear how the data sources were used to construct the replication dataset?*

(a) Data Needed

RR TEAM INSTRUCTIONS: *List below the datasets and variables the original author used to analyze the focal claim. Include details regarding the sample size, waves or years used, and other details pertinent to finding an existing dataset for replication. Please include page numbers when excerpting from the original article. If possible, categorize the list of variables as one of the following: dependent variable, focal independent variable, control variable, or sample parameters/clustering variable. Finally, include the sample size of the original study's focal analysis, if it is available.*

Fear of being infected by COVID-19

- Quantitative online survey; n = 1300 (overall); n = 1242 (t-test)
- Data from March, 2020
- Geographical location: All regions in Germany
- Measured on a 5-point likert scale, converted to a binary variable that takes a value of 1 if respondents answer either with “strongly agree” or “agree” on the 5-point Likert Scale to the question “The COVID-19 (Coronavirus SARS-CoV-2) worries me” (p.5, under table 4).

Gender
- Quantitative online survey; n = 1300 (overall); n = 1242 (t-test)
- Data from March, 2020
- Geographical location: All regions in Germany
- Can take two values: female and male.

Sample Parameters

Sample size of analysis has 1242 observations.

(b) Data Access

RR TEAM INSTRUCTIONS: *Describe below the data sources that will provide the replication variables. Include information such as the name of the data source (e.g., Indonesian Family Life Survey), the description and link of the data source, and the waves needed to create a final replication dataset.*

Also describe the process for accessing the data sources that will be used to create the final replication dataset; specify how long long it took for the registration to be approved and what information was required (e.g., writeup of the purpose of the project, email address from an IPCSR institution, etc.); and verify that the data can be opened as expected. If applicable, provide a link to the page where you registered to access the data.

Describe in detail any restrictions on data access and data-sharing, as well as any additional terms of data use that will be relevant for the replication study and final report (e.g. citations that will need to be made). If you were able to access the data because of special permissions that you have, but that you expect other researchers might not have, please document those as well.

The dataset from Fetzer et al. (2020) is openly available and accessed via OSF at this link. In order to the use the dataset, authors must agree to the following GDPR-related statements:
By downloading and using the dataset, you agree to comply with the European Union's GDPR requirements, as well as the following:
— My research team and I will only use the data for research, and not for commercial purposes
— My research team and I will not attempt to identify individuals using these data
— My research team and I will share the resulting pre-print/publication with the researchers
— My research team and I will cite the dataset using the citation below.

Additionally, the dataset must be cited if being used in the replication. Other than this, there are no restrictions on data access and sharing.

(c) Variable Availability

RR TEAM INSTRUCTIONS: *For each variable required for the replication analysis (listed above), describe the variables from the replication data that can be used to measure it (including which data files or sources each measure is found in),* ***any notes a data analyst should consider when using the measure in a replication analysis****, and any important differences between the original variable and the proposed replication variable.*

*If there are multiple variables in the replication data that correspond to a required variable (e.g. two different measures of education in the replication data), include all of those options below. If a variable from the original study **cannot** be measured using the replication data, please make that clear as well. **Finally, include a description of the identifiers used to merge multiple datasets, if applicable.***

Fear of being infected by COVID-19

- Dataset: "GlobalBehaviorsPerceptions_Data_May21_2020.dta"
- Variable name: mh_anxiety_1: "I am nervous when I think about current circumstances"
- Variable name: mh_anxiety_3: "I am worried about my health"
- Notes: These questions do not explicitly ask about worries related to COVID-19, but are positioned in a survey on COVID-19, so it may be assumed that respondents view this question about "current circumstances" to be the same as "COVID-19". The response is measured in a 5-point scale: "Does not apply at all" (value 1) to "Strongly applies" (value 5).

Gender

- Dataset: "GlobalBehaviorsPerceptions_Data_May21_2020.dta"
- Variable gender: "Selected choice"
- Takes three values: 1 "Female" 2 "Male" and 3 "Other"; For purposes of analysis (t-test), indicator variable for female can be created or "Other" values can be set to missing.

Sample weights:

- Dataset: "GlobalBehaviorsPerceptions_Data_May21_2020.dta"
- Variable: weight_new
- Variable: weight_sample
- Is a continuous variable.
- Notes: Fetzer et al. (2020) provide population weights by country as well as sample weights. See Section 3 of the data documentation file available here for details on how weights were computed.

(d) Data Creation

RR TEAM INSTRUCTIONS: *Create a dataset using the data sources and variables listed above. Provide a detailed narrative describing how the various datasets were cleaned and merged into a final replication dataset. Provide a view-only link to a clearly commented script on the OSF that produces the replication data as described in the narrative. Our preference is that this be either an R script or a script from another language that similarly allows for open and reproducible analyses. Please let the SCORE team know if this is not possible.*

- *If the data can be freely shared and posted to OSF, please post it in your OSF project and provide a link to the completed dataset below.*
- *If any part of the dataset cannot be shared between researchers or posted to the OSF, please leave the final dataset off the OSF. Instead, include either below or in your script (commented out at the bottom) two pieces of information that will help an independent team verify they have created the dataset according to your instructions:*

Commented [5]: If there are at least 20 cases in the third group, ANOVA should be used. If only a few cases describe themselves as the members of the third group, they should be mentioned aside from the main analysis.

Commented [6]: We agreed that it will be best to examine a female vs. male difference because that is what was done in the original study.

- *The dimensions of the final dataset(s) you've created (# of rows, # of columns)*
- *A summary of 8-10 variables in the replication dataset. For numeric variables, the summary should include the mean, standard deviation, and count of NAs. For categorical variables, the summary should include each level present in the data and its count, as well as a count of NAs. If multiple datasets are submitted as part of your work, at least one variable should be included from each dataset.*

The data from the replication sources should be preserved in as 'raw' a form as possible, in order to give the data analyst the most latitude to clean the variables as they see fit. Variables from the original source should be preserved in their original form (e.g. do not recode values of 99 to NA). New variables should only be created when they're needed to complete the merge or combine the datasets; in those cases, please preserve a version of the original, unaltered variable in the new dataset.

When combining multiple datasets by binding rows, please be sure that the data type and measurement units are equivalent across each dataset. If there is a discrepancy in how a variable is measured across datasets, rename the variable in each dataset to indicate the original dataset, and then carefully document the resulting measures below and in the data dictionary. See here for an example of how this should work.

Please also use this section to describe:

- *Any deviations between the original study design and the replication design that would result from using this replication dataset.*
- *Any notes about using these variables that you would like to pass along to the data analyst.*

To access the Fetzer et al. (2020) dataset via OSF, first visit https://osf.io/3sn2k/files/

Download all the files from this link, or to access the main dataset: https://osf.io/usq5b/

Following this, open the dataset in Stata.

Use the code provided (partially derived from code made available by the authors) to clean the data and generate weights

The code to generate the replication dataset is available on OSF here: https://osf.io/x3g6v/

The replication dataset is available on OSF here: https://osf.io/kzq7r/

The number of rows is 10,192, and the number of columns is 6.

The summary statistics of the numeric variables (there are no missing/NA values)

Variable	Obs	Mean	Std. Dev.	Min	Max
mh_anxiety_1	10,192	2.979297	1.129728	1	5
mh_anxiety_3	10,192	3.000098	1.126265	1	5
weight_sample	10,192	.0001312	.0004746	7.10e-06	.0132449
weight_new	10,192	1.11e+08	4.02e+08	6012257	1.12e+10

The summary statistics for gender, the categorical variable is as below:
Male (5,138); Female (4,933); and other (121). There are no missing/NA observations.

The replication attempt will require a new categorical variable for female, which takes a value of 1 if the respondent is female and zero if the respondent is male (in the original study, there was only a binary response for gender). This is included in the code. The original variable from the replication dataset is preserved as is. It will also require computation of the sample weights using the procedure laid out in Fetzer et al. (2020).

In the original study, the focal claim appears to have been tested using a binary variable for worry about COVID-19 [“62.1% agree (answering either with “strongly agree” or “agree” on the 5-point Likert Scale) that they are worried about COVID-19 in general (women=68.2%, men=55.7%, $p < .01$),” p.5]. However, details of the statistical test applied are not available in the paper or otherwise. Therefore, in the replication attempt, it may be useful to conduct an independent samples t-test between males and females using the “raw” versions of the variables (scales from 1 to 5) rather than creating a binary variable. Alternatively, if a binary variable is desirable, a chi-squared test is appropriate. The resulting binary variable can take a value of 1 if the worries variable takes a value of 4 [Somewhat applies] or 5 [Strongly applies], and a value of 0 if the worries variable takes a value of 1 [Does not apply at all] or 2 [Somewhat/Does not apply]. The details are provided in the code.

(e) Data Dictionary

RR TEAM INSTRUCTIONS: *Create a data dictionary following this template. Provide below a view-only link to the completed data dictionary included in the OSF project. If the Data Analyst will need to create new variables using the variables in the final replication dataset (e.g. recoding the provided education variable to be in a better format for analysis), please document below your recommendation on how the analyst should do so. Please also document any additional notes regarding the variables in the dataset that do not fit within the provided data dictionary template or the other sections above.*

The data dictionary can be found here: https://osf.io/b3qxz/

13. Sample size

RR TEAM INSTRUCTIONS: *Please report below the analytic sample size(s) in the replication dataset, with reference to however many units or levels are in the data. Please report as much information here as will be helpful for the review committee to be aware of, including differences in sample size resulting from various analytic decisions (e.g. listwise deletion vs multiple imputation).* ***Finally, when the replication combines observations from the original study with new observations, please estimate what proportion of the analytic sample's observations will be comprised of original vs. new observations.***

Data finders' response goes here:

Sample parameters (Germany):

- mh_anxiety_1 (Worry about current circumstances): 10,192 non-missing observations (10,192 total)
- mh_anxiety_3 (worry about my health): 10,192 non-missing observations (10,192 total)
- gender (female): 10,192 non-missing observations (10,192 total)
- gender (female indicator variable): 10,071 non-missing observations (10,192 total)
- weight_new (within country weights, based on share of popn): 10,192 non-missing observations (10,192 total)
- weight_sample (within country weights): 10,192 non-missing observations (10,192 total)

Required sample size [to be filled out by the SCORE team]: The primary unit of analysis is the respondent. An estimate of the minimum viable sample size for the data analytic replication is: 720. For comparison, the stage1 required sample size would be: 3488 and the stage2 sample size would be: 7844.

Notes on the power analysis: The original authors said they 'usually use t-Test and Anova for comparing mean values', but did not provide any specifics on the test statistic or df values when contacted by the SCORE team. This power analysis assumes a t-test here, and used the $p<.01$ from the claim as a p-value threshold and then back calculated a t-value and d value from there. There would be the 'worst case' values, meaning the sample size is very conservative and would likely 'overpower' the study based on its true t and d values, which are likely larger than those back-calculated here. The power analysis also assumes equal or very close to equal N for male and female respondents.

14. Sample size rationale

For data analytic replications in SCORE, three sample sizes are calculated:

- *A minimum threshold sample size, defined as the sample size required for 50% power of 100% of the original effect*
- *A stage 1 sample size, defined as the sample size needed to have 90% power to detect 75% of the original effect*
- *A stage 2 sample size, defined as the sample size needed to have 90% power to detect 50% of the original effect*

Details about how those sample sizes were calculated for this project are found here.

15. Stopping rule (provided by SCORE)

RR TEAM INSTRUCTIONS: *Because all existing data replications that clear SCORE's minimum power threshold will proceed to analysis, the stopping rule is not relevant for these kinds of projects.*

N/A -- all observations will be used in a single analysis.

Variables

RR TEAM INSTRUCTIONS: *The preregistration form divides variables across three questions: manipulated variables, measured variables, and indices (i.e. analytic variables derived from raw variables). For existing data replications, only fill out the "Measured variables' and 'Indices' sections. Please do not fill out anything in the 'Manipulated variables' section.*

The raw data of any transformed variable (e.g. reaction time → log reaction time) or any created index should be defined in the 'Measured variables' section. Details regarding the variable transformation should be specified in the 'Transformations' section. Details regarding the creation of an index should be specified in the 'Indices' section.

Across these questions, you should define all variables that will later be used during your analysis (including data preparation/processing). You can describe all variables in the preregistration and/or summarize and link to a data dictionary (codebook) in your repository to answer these questions.

If you will share data from your replication, this is also the place to state whether any variables will be removed prior to sharing the dataset (e.g. to reduce risk of participant identification or comply with copyright restrictions on scale items.)

16. Manipulated variables

RR TEAM INSTRUCTIONS: *Manipulated variables in this preregistration refer specifically to variables that have been randomly assigned in an experiment. The use of data from an experiment should be rare in existing data replications. If your existing data replication relies on experimental data, please document each manipulated variable as a measured variable, and use the codebook to indicate what each level of the variable corresponds to (e.g. participants assigned to the treatment condition = 1; participants assigned to the control condition = 0). The default language in bold below has been copied into all existing data replication preregistrations.*

N/A -- not documented for existing data replications.

17. Measured variables

RR TEAM INSTRUCTIONS: *Please use this section to document each variable that was used in the original study's analysis and the role it served (e.g. dependent variable, control variable, sample parameter, etc). For each variable, provide the description of the variable offered in the paper and/or codebook of the original study, the variable in the replication dataset that it corresponds to, and explain any deviations between the two. In cases where an equivalent replication variable was not found, explain how, if at all, you expect it will affect the replication*

attempt. In cases where you are adding a variable that was not present in the original study, please explicitly state that you are doing so, and explain how, if at all, you expect it will affect the replication attempt.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration surface all of the variables needed to replicate the focal analysis?*
- *Are deviations between the original variables and replication variables documented when needed?*

VARIABLE #1: GENDER (I.E., FEMALE)

- These data will be used to conduct a t-test to examine the replicability of the focal claim. Specifically, the "female" data from the Fetzer et al. (2020) data set will be used to ascertain if a female versus male mean difference exists with regard to scores on "mh_anxiety_1" ("I am nervous when I think about the current circumstances").
- A formal description of this variable was not provided in the original study. In addition, the survey item and the corresponding response options used to collect these data were not provided in the original study.
- A dummy vector where 0 = male and 1 = female has been provided in the replication data set (i.e.,Fetzer et al. [2020]). This categorical variable will be used to examine if group differences (i.e., female vs. male) exist.
- It appears has if the original study (Gerhold, 2020) asked individuals to indicate if they were female or male (i.e., two response options). In contrast, in the replication study (Fetzer et al., 2020) respondents were asked to indicate if they identified as female, male, or other (i.e., three response options). 121 individuals identified as other. However, given that the original study examined female versus male differences, data pertaining to these 121 individuals was marked as missing and, thus, removed from the replication analyses.

VARIABLE #2: mh_anxiety_1

- These data, along with the corresponding "female" data (see VARIABLE #1 above), will be used to examine if females and males differ with regard to worrying about their current circumstances.
- In the original study, a formal and clear definition of "worrying about current circumstances" was not provided. However, an inspection of Table 4, which was reported on page five of the original study, indicates that "worrying about current circumstances" was operationalized using two items: (1) The COVID-19 (Coronavirus SARS-CoV-2) worries me and (2) I am afraid of the being infected by COVID-19 (Coronavirus SARS-CoV-2).
- In the replication study (Fetzer et al., 2020), individuals responded to the following two items using a five-point Likert scale: (1) I am nervous when I think about current circumstances (mh_anxiety_1) and (2) I am worried about my health (mh_anxiety_3). These two items served as proxies for the two items that were measured in the original

study. However, given that these items are not complementary and, thus, would likely not form a consistent higher-order composite, it is suggested that the replicability of the focal claim be examined using “mh_anxiety_1.” Indeed, this suggest is supported by the fact that the replication data were collected in March and April 2020 (i.e.,during the midst of the COVID-19 pandemic in Europe), which likely means “mh_anxiety_1” will serve as a good proxy for what was tested in the original study.

18. Indices

RR TEAM INSTRUCTIONS: *If any of the measured variables described in Section 17 will be combined into a composite measure (including simply a mean), describe in detail what measures you will use and how they will be combined. Please be sure this preregistration includes a link to a clearly commented script that constructs the index according to the narrative.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify each of the composite measures (e.g. mean scores, factor scores) that are needed for the focal analysis, and which of the measured variables in Section 17 are used in each one (e.g. the happiness, joy, and satisfaction items will be used to create the ‘positive feelings’ measure)?*
- *Does the preregistration link to a clearly commented script that constructs the indices according to the narrative description?*

A composite measure for “mh_anxiety_1” will be calculated. Specifically, female and male group means for these constructs will be calculated so that an independent sample t-test can be conducted to ascertain if gender differences exist. A commented script, which can be found at the project website (see https://osf.io/a7h9n/), outlines how the analyses were conducted.

Analysis Plan

19. Statistical models

RR TEAM INSTRUCTIONS: *This section should describe in detail the analysis that will be performed to replicate the focal result. This analysis must align as closely as possible with the original study’s analysis, even if you have identified limitations in the original study. The level of detail should allow anyone to reproduce your analyses from your description below. Examples of what should be specified: the model; each variable; adjustments made to the standard errors and to case weighting; additional analyses that are required to set up the focal analysis; and the software used.*

Beyond the replication of the focal analysis from the original study, it is at your discretion to test the claim using other analytic approaches as a check of the robustness of the claim. The original test should be listed first and be clearly distinguished from any other tests. If you are testing additional confirmatory hypotheses, describe them in the same order as you numbered

them in the "Hypotheses" section above and make clear reference to the specific hypothesis being tested for each.

Please provide a link to a clearly commented script that performs the analysis described in the narrative provided below. Our preference is that this be either an R script or a script from another language that similarly allows for open and reproducible analyses. Please let the SCORE team know if this is not possible. Please also test that the code runs without error on a random subset of 5% of the replication dataset, and provide verification that the code has produced a sensible result below (a screenshot of the results is preferable). Finally, please confirm that you have only developed and tested your analysis plan and code using 5% of the data.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which statistical model will be used to provide the 'focal evidence' for the SCORE test (e.g. a regression coefficient in a larger multiple regression model), and does it correspond closely to the model and evidence from the original study?*
- *Does the preregistration describe each variable that will be included in the focal analysis, and what role each variable has (e.g. dependent variable, independent variable)?*
- *Does the preregistration include a detailed specification of the focal analysis, including interactions, lagged terms, controls, etc., in both narrative form and in a clearly commented script?*
- *Does the preregistration verify that the code runs without error on a random subset of the replication dataset?*

As noted on page 12 by the previous team (see "Data Creation" section), "details of the statistical test applied [in the original study] are not available in the paper or otherwise." As such, and aligned with their recommendation, an independent samples t-test will be performed to assess the replicability of the focal result. The t-test examines if "mh_anxiety_1" scores differ between females and males. In the following section, procedural details for the first t-test are outlined.

First, the Fetzer et al., (2020) data set (n = 10,192) is imported. Second, irrelevant/missing data (i.e., gender = "other") are removed from the data set. 121 individuals who identified as "other" are removed, which reduces the usable data set to n = 10,071. Third, female and male data are stratified into two subsets. The sample sizes for the female and male subsets are n = 4,933 and n = 5,138, respectively. Fourth, a test for homoscedasticity is conducted for each group's "mh_anxiety_1" scores. If the p-value from this test is $p > .05$ (greater than .05), then it can be assumed that the variances of both samples are homogenous. If this is the case, then the last step in the replication process is to conduct a classic Student's two-sample t-test to ascertain if a group difference potentially exists. The resulting p-value will be used to determine whether or not a female vs. male difference is observed. Specifically, the data suggest that a gender difference exists if the p-value is less than .05, In contrast, a gender difference is not observed if the resulting p-value is not less than .05.

The linked script (see https://osf.io/a7h9n/) performs the proceeding analyses.

This statement confirms that only 5% of the data have been randomly sampled in developing the analysis plan and code contained in this preregistration. [Note that this statement is bolded because the data analyst confirmed that it is true.]

20. Transformations

RR TEAM INSTRUCTIONS: *This section should describe how any of the measured variables or composite measures mentioned above will be transformed prior to the analyses listed in Section 19. These are adjustments made to variables* ***after*** *measurement or measure creation, and might include centering, logging, lagging, rescaling etc. Please provide enough detail such that anyone else could reproduce the transformations based on the description below. Please be sure this preregistration includes a link to a clearly commented script that performs the transformations described in the narrative provided below.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which of the measured variables or composite measures will need to be transformed prior to the focal analysis?*
- *For each variable needing transformation, does the preregistration adequately describe the transformations, including any centering, logging, lagging, recoding, or implementation of a coding scheme for categorical variables?*
- *Does the preregistration link to a clearly commented script that performs each transformation?*

Transformations are not necessary because the data set provided from the replication study (Fetzer, 2020) includes all required data in the appropriate form.

21. Inference criteria

RR TEAM INSTRUCTIONS: *This section describes the precise criteria that will be used to assess whether the hypotheses listed above were confirmed by the analyses in Section 19. The default language below only applies to the test of the SCORE claim,* ***H****. It is at your discretion to describe the inferential criteria you will use for any additional analyses. They need not rely on p-values and/or the same alpha level we have specified for* ***H****.*

If the additional analyses will use multiple comparisons, the inference criteria is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.

Criteria for a successful replication attempt for the SCORE project is a statistically significant effect (alpha = .05, two tailed) in the same pattern as the original study on the focal hypothesis test (**H***).

The hypothesis to be tested is: Women will express more concern about COVID-19 than men. As such, the inference criteria that will be used to assess whether the focal claim replicates will be the *p*-value derived from a t-test that examines if females and males differ with regard to "mh_anxiety_1" scores. Specifically, a p-value less than .05 will indicate that a gender difference exists.

22. Data exclusion

RR TEAM INSTRUCTIONS: *The section below should describe the rules you will follow to exclude collected cases from the analyses described in Section 19. Note that this refers to exclusions **after** the creation of the replication dataset; exclusion criteria that prevent a case from entering the replication dataset in the first place should be detailed in the 'Data Collection Procedure' section above. Please be as detailed as possible in describing the rules you will follow (e.g. What is the specific definition of outliers you will use? Exactly how many attention checks does a participant need to fail before their removal from the analytic sample?).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on whether any cases included in the replication dataset will be excluded prior to data analysis?*
- *If yes, does the preregistration provided detailed instructions on how the exclusions will be performed (e.g. Is the definition of outlier provided? Is the number of attention checks failed before a participant is excluded specified?)*

Given that the original study (Gerhold, 2020) focused on female versus male differences, only data from individuals who identified their gender as "other" will be excluded from the current analyses because they are not needed. The linked script (see https://osf.io/a7h9n/), which performs all replication analyses, removes the irrelevant data.

23. Missing data

RR TEAM INSTRUCTIONS: *The section below should describe how missing or incomplete data will be handled. Please be as detailed as possible in describing the exact procedures you will follow (e.g. last value carried forward; mean imputation) and any software required (e.g. We will use Amelia II in R to perform the imputation).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on how missing or incomplete data will be addressed (e.g. casewise removal, missing data imputation)?*
- *If applicable, does the preregistration specify how many missing variables will lead to a case's removal (e.g. If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.)?*
- *If applicable, does the preregistration describe how missing data imputation will be performed, including relevant software?*

A complete data set was provided and, thus, missing data was not an issue. However, as previously mentioned in Section 23, data from individuals who identified their gender as “other” were removed because they are irrelevant for the current replication analyses.

24. Exploratory analysis (Optional)

RR TEAM INSTRUCTIONS: *If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time. If any exploratory analyses involve additions to the data collection procedure beyond what was performed in the original study (e.g. additional items on the survey; running another condition in the experiment), please describe them below.*

In the interest of completeness, an exploratory analysis is recommended. Specifically, it is recommended that all procedural details reported in Sections 17-23 be repeated with one exception, “mh_anxiety_1” is replaced by “mh_anxiety_3.” Effectively, it is suggested that a second t-test be conducted that examines whether or not females and males differ with regard to scores on “mh_anxiety_3” (I am worried about my health).

25. Other

RR TEAM INSTRUCTIONS: *This section serves two purposes. First, please use this section to discuss any features of your replication plan that are not discussed elsewhere. Literature cited, disclosures of any related work such as replications or work that uses the same data, plans to make your data and materials public, or other context that will be helpful for future readers would be appropriate here. Second, please also re-surface any major deviations from earlier in the preregistration that you expect a reasonable reviewer could flag for concern. Give a summary of these deviations, focusing on larger changes and any possible challenges for comparing the results of the original and replication study.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration reference other sections of the preregistration where substantial deviations from the original study have been described (including deviations due to differences in location or time compared to the original study)?*
- *Does the preregistration comment on plans to make the data and materials from the replication study public?*

N/A.

Final review checklist

REVIEWER INSTRUCTIONS: *For the following questions, reviewers please indicate whether you can 'sign off' on the following items by adding a comment. You can update this response as the lab moves through revisions during the review period!*

- Included in this pre-registration are specific materials needed to create a replication dataset:
    - Is the final replication dataset that the research team constructed suitable for performing a high-quality, good-faith replication of the focal claim selected from the original study?
    - Is the procedure for constructing the final replication dataset sufficiently documented that an independent researcher could construct the same dataset following the procedures and code they lay out?
- Included with this pre-registration is a narrative description of how the replication dataset will be used to perform the focal replication analysis, as well as the specific analytic scripts/code/syntax that will be used:
    - Is the analysis plan (including code) that's documented in the preregistration consistent with a high-quality, good-faith replication of the focal claim selected from the original study?
    - Has the data analyst demonstrated that the analysis code works as expected on a random 5% of the final replication dataset?
- I have reviewed all sections of this pre-registration, and I believe it represents a good-faith replication attempt of the original focal claim.