ML4 Results Section in rMarkdown

|  |
| --- |
| Richard A. Klein1 |
| 1 Université Grenoble Alpes |
|  |

# Author note

Correspondence concerning this article should be addressed to Richard A. Klein, . E-mail: [raklein22@gmail.com](mailto:raklein22@gmail.com)

Abstract

*Keywords:* Terror Management Theory, mortality salience, replication, many labs

Word count: X

ML4 Results Section in rMarkdown

## Participants

21 labs participated and provided a total sample of 2,281 participants. In accordance with the pre-registration (<https://osf.io/4xx6w>), we immediately excluded from all analyses participants who either failed to complete all 6 ratings of the essay authors, or who failed to complete both writing prompts within the mortality salience or control conditions (e.g., the between-subjects manipulation).[[1]](#footnote-24) Thus, the usable sample included 2,220 participants (see Table 1 for a summary of sites). 1,157 participants (52.12%) reported being female and 708 participants (31.89%) reported being male; the remaining participants did not respond to the item, were asked about gender in a non-standard way, or chose a different response. The mean age was 19.87 years (*SD* = 2.79). Participant reported race was 910 (40.99%) White, 221 (9.95%) Asian, 120 (5.41%) Black or African American, 36 (1.62%) American Indian or Alaska Native, 20 (0.90%) Native Hawaiian or Pacific Islander, 114 (5.14%) Other. The remaining participants did not report their race, or responses were not easily recoded to match these categories.

## Analysis Plan

The primary finding of interest from Greenberg et al., (1994) was that participants who underwent the mortality salience treatment showed greater preference for the pro-US essay author over the anti-US essay author compared to the control condition. To assess whether the replication results support the original, we followed a similar analysis plan as in the original article. Scores from the three items evaluating the authors of the anti-American essays were averaged ( = 0.90) and then subtracted from the average of the three items evaluating authors of the pro-American essays ( = 0.89).[[2]](#footnote-26) An independent samples *t*-test was then conducted comparing those in the “subtle own death salient” (MS) condition with scores from the “TV salient” (control) condition. Some labs administered both Author Advised and In House protocols. To account for this nesting of effect sizes within labs, a three-level random-effects meta-analysis was conducted using the MetaSEM package (Cheung, 2014) in R (R Core Team, 2019).

Original authors were not entirely in agreement about what exclusions should be implemented. So, we repeated our analyses under different exclusion criteria suggested by original authors:

*Exclusion set 1:* Include all participants who completed the materials (e.g., wrote something for both writing prompts, and completed all six items evaluating the essay authors). Reduces the usable N from 2,281 to 2,220 participants. This sample size gives us 95% power to detect a condition effect of *d* = .15 in an independent samples *t*-test.

*Exclusion set 2:* All prior exclusions, and further exclude participants who did not identify as White or who indicated they were born outside the United States. Reduces *N* to 1,874. This sample size gives us 95% power to detect a condition effect of *d* = .16.

*Exclusion set 3:* All prior exclusions, and further exclude participants who responded lower than 7 on the American Identity item (“How important to you is your identity as an American?” 1 - not at all important; 9 - extremely important). Further reduces the usable *N* to 1,693 participants. This sample size gives us 95% power to detect a condition effect of *d* = .18.

Exclusion sets 2 and 3 were specifically recommended by original authors and these criteria were used to analyze the data from Author Advised labs. However, the data required to make these exclusions were often not collected at In House replication sites because they made independent decisions about design and demographic measures for potential exclusion, and these measures were not in the original article. Thus, for all analyses only Exclusion Set 1 was used for In House participants. All analysis plans and procedures were pre-registered on the OSF prior to data collection (<https://osf.io/4xx6w>).

# Results

## Researcher Expectations and Characteristics

A total of 28 researchers from 21 participating sites completed an experimenter survey about their motivations and expertise. This survey was administered during data collection, and although no researcher had access to overall project-wide results, ~⅓ of the researchers reported looking at or analyzing their own site’s data prior to completing the survey. Psychology research experience ranged from 0 to 28 years (*M* = 9.32, *SD* = 8.80). 1 (4%) researcher indicated they were an expert in TMT, 5 (18%) indicated they had “a lot” of TMT knowledge, 10 (36%) indicated “some” knowledge, 5 (18%) indicated little knowledge, 6 (21%) indicated zero knowledge, and 1 (4%) did not respond to the question.

When asked what outcome they wanted to happen, 13 (46%) indicated that they hoped for the project to successfully replicate the TMT effect, 10 (36%) indicated no preference, and 3 (11%) hoped the project would result in a failure to replicate, with 2 (7%) researchers leaving the question blank. On average, the teams estimated a 54% chance of successful replication with a wide range of estimates from 20% to 95% (*SD* = 22.14).[[3]](#footnote-30)

## Research Question 1: Meta-analytic results across all labs (random effects meta-analysis).

The most basic question is whether we observed the predicted effect of mortality salience on preference for pro- vs anti- American essay authors. To assess this we conducted a three-level random-effects meta-analysis.[[4]](#footnote-32) This analysis produces the grand mean effect size across all sites and versions. Regardless of which exclusion criteria were used, we did not observe the predicted effect and the confidence interval was quite narrow: Exclusion set 1: *Hedges’ g* = 0.03, 95% CI = [-0.05, 0.11], *SE* = 0.04, *Z* = 0.71, *p* = 0.48. Exclusion set 2: *Hedges’ g* = 0.07, 95% CI = [-0.04, 0.19], *SE* = 0.06, *Z* = 1.26, *p* = 0.21. Exclusion set 3: *Hedges’ g* = 0.04, 95% CI = [-0.09, 0.18], *SE* = 0.07, *Z* = 0.61, *p* = 0.54. Forest plots showing the effects for individual sites and the aggregate are available in Figure 1 for exclusion set 1 (see <https://osf.io/8ccnw/> for the other two exclusion sets).

There may have been a mortality salience effect at some sites and not others, so we next examined how much variation was observed among effect sizes (e.g., heterogeneity). For exclusion sets 1 and 3, this sort of variation did not exceed variation expected by chance (e.g., sampling variance): Exclusion set 1: *Q*(20) = 25.82, *p* = 0.17; Exclusion set 3: *Q*(20) = 29.61, *p* = 0.08. The amount of variation between sites did exceed chance for Exclusion set 2, *Q*(20) = 36.32, *p* = 0.01, however it was small in magnitude, Tau2within labs = 0.00, Tau2between labs = 0.02.

In sum, we observed little evidence for an overall effect of mortality salience in these replications. And, overall results suggest that there was minimal or no heterogeneity in effect sizes across sites. This lack of variation suggests that it is unlikely we will observe an effect of Author Advised versus In House protocols or other moderators such as differences in samples or TMT knowledge. Even so, the plausible moderation by Author Advised/In House protocol is examined in the following section.

## Research Question 2: Moderation by Author Advised/In House protocol

A covariate of protocol type was added to the random effects model to create a three-level mixed-effects meta-analysis. This was pre-registered as our primary analysis.[[5]](#footnote-35)

This analysis again produces an overall grand mean effect size, and those were again near zero and relatively precisely estimated across all three exclusion sets: Exclusion set 1: *Hedges’ g* = 0.02, 95% CI = [-0.10, 0.14], *SE* = 0.06, *Z* = 0.32, *p* = 0.75. Exclusion set 2: *Hedges’ g* = 0.04, 95% CI = [-0.10, 0.17], *SE* = 0.07, *Z* = 0.51, *p* = 0.61. Exclusion set 3: *Hedges’ g* = 0.02, 95% CI = [-0.11, 0.16], *SE* = 0.07, *Z* = 0.34, *p* = 0.73.

Variation among effect sizes also followed the previously observed pattern. Weak heterogeneity for exclusion set 2, *Q*(20) = 36.32, *p* = 0.01, Tau2within labs = 0.00, Tau2between labs = 0.02; while variation did not meet the statistical significance threshold for exclusion sets 1 *Q*(20) = 25.82, *p* = 0.17; or exclusion set 3: *Q*(20) = 29.61, *p* = 0.08.

Critically, protocol version did not significantly predict replication effect size regardless of which exclusion criteria were used. Exclusion set 1: *b* = 0.02, *Z* = 0.29, *p* = 0.77; Exclusion set 2: *b* = 0.11, *Z* = 0.92, *p* = 0.36; Exclusion set 3: *b* = 0.10, *Z* = 0.57, *p* = 0.57. The Author Advised version did not produce larger effect sizes when compared with the In House versions.

## Research Question 3: Effect of Standardization

Finally, we tested whether In House protocols displayed greater variability in effect size than Author Advised protocols. To test this hypothesis, we ran the mixed-effects models but constrained the variances at both level 2 and level 3 to 0, effectively creating fixed-effects models. These models were then compared with a chi-squared differences test to assess whether the fit significantly changed. In this case, none of the three models significantly decreased in fit: Exclusion set 1: *²* (2) = 0.16, *p* = 0.92; Exclusion set 2: *²* (2) = 0.85, *p* = 0.65; Exclusion set 3: *²* (2) = 0.27, *p* = 0.87. Overall, there was no evidence that In House protocols elicited greater variability than Author Advised protocols despite the fact that they were unambiguously more variable in their procedural implementation.

## Follow-Up Exploratory Analyses

**Results for TMT-knowledgeable sites.** One principal investigator reported being an expert in TMT, while 5 others indicated having “a lot” of knowledge about TMT. One might expect that these locations would have greater success at replicating the mortality salience effect. Aggregating across these sites, and using only the first exclusion rule, these sites did not elicit a larger difference between the mortality salience group (*M* = 1.02, *SD* = 2.30) and the control group (*M* = 0.93, *SD* = 2.30), *t*(520.81) = 0.43, *p* = 0.67, *Hedges’ g* = 0.04, 95% CI = [-0.13, 0.21].[[6]](#footnote-38)

**Results for participants who preferred the pro-US author** The present hypothesis that mortality salience would cause a participant to become more favorable to the pro-US author as compared to the anti-US author relies on the participant perceiving the pro-US stance as more similar to their own worldview (and/or the anti-US stance as threatening to their worldview). Original authors anticipated that the essays from the original study may not serve this function in the replication, run in 2016. For this reason, the anti-US essay from the original study was made more extreme in the Author Advised version of the replication. There was a particular concern that in the months leading up to and following the 2016 US Presidential Election of Donald Trump, the generally more liberal leaning student bodies on college campuses may feel less patriotic and not identify with the pro-US worldview. Indeed, analysis suggests the original authors anticipated and more successfully addressed this issue. Among In House replications, 49% of participants prefered the pro-US essay author, 40% preferred the anti-US essay author, and 11% had no preference. Among Author Advised replications, 68% of participants prefered the pro-US essay author, 22% preferred the anti-US essay author, and 10% had no preference.

However, the predicted mortality salience effect was not larger or detectable via statistical significance when subsetting to only participants at Author Advised sites who preferred the pro-US author. In all exclusion sets the mortality salience and control groups showed similar levels of preference for the pro-US author over the anti-US author: Exclusion set 1: mortality salience group (*M* = 1.23, *SD* = 2.06), control group (*M* = 1.15, *SD* = 1.83), *t*(796.99) = 0.58, *p* = 0.56, *Hedges’ g* = 0.04, 95% CI = [-0.10, 0.18]; Exclusion set 2: mortality salience group (*M* = 1.53, *SD* = 2.15), control group (*M* = 1.38, *SD* = 1.97), *t*(446.51) = 0.79, *p* = 0.43, *Hedges’ g* = 0.07, 95% CI = [-0.11, 0.26]; Exclusion set 3: mortality salience group (*M* = 1.96, *SD* = 2.15), control group (*M* = 1.83, *SD* = 2.05), *t*(264.51) = 0.49, *p* = 0.62, *Hedges’ g* = 0.06, 95% CI = [-0.18, 0.30]. The confidence intervals were wider because of the smaller total sample size, but this evidence is not consistent with the hypothesis that preference for the pro-US author would elicit an effect of mortality salience in this context.

1. The latter exclusion criteria applied only to participants from Author Advised sites, because the necessary data was not always available for In House sites. [↑](#footnote-ref-24)
2. Supplemental analyses treating these as two separate dependent variables are available in the online supplement (<https://osf.io/xtg4u/>), and those outcomes do not qualify the conclusions offered here. [↑](#footnote-ref-26)
3. Including only sites that had not looked at any data, researchers estimated a 56% chance of successful replication. [↑](#footnote-ref-30)
4. Sample code to run this analysis is: meta3(y=es, v=var, cluster=Location, data=dataset). In this sample code, “y=es” directs the program to the column of effect sizes, “v=var” indicates the variable to be used as the sampling variance for each effect size, and the “cluster=Location” command groups the effect sizes by a location variable in the dataset (in this case, a unique identifier assigned to each replication site). [↑](#footnote-ref-32)
5. The addition of the argument “x = version” to the prior metaSEM R code can be seen here: meta3(y=es, v=var, cluster=Location, x=version, data=dataset) [↑](#footnote-ref-35)
6. One site, UW Madison In House, used a 7-point scale. This has been rescaled to a 9-point scale for this analysis to approximately compare it with the others. [↑](#footnote-ref-38)