Social Media, Age, and Mental Well-Being

Replicating an Age-Moderation Claim Using the General Social Survey

Abstract

This project replicates Hardy and Castonguay's 2018 analysis of social media use, age, and mental well-being using the 2016 General Social Survey. The published study reported that additional social networking site use was associated with higher odds of reporting a past feeling of being close to a nervous breakdown, with the association varying across age groups. This replication reconstructed the descriptive benchmarks and reproduced three logistic regression models from the published article. The replicated regression sample was smaller than the published sample (714 versus 743), while model fit closely matched the source article: Nagelkerke R^2 values were 0.18, 0.19, and 0.18. The replicated odds-ratio pattern also matched the age-moderation claim: social media use was positive in Model 1 (OR = 1.22), weaker for ages 18–29 (OR = 0.59), and stronger for ages 30–49 (OR = 1.41).

1. Background and Significance

Social networking site use has become a major topic in mental well-being research. Prior studies have reported conflicting relationships between social media use and mental health, including positive, negative, and null associations. Hardy and Castonguay (2018) addressed this inconsistency by examining whether age moderates the relationship between social networking site use and mental well-being. Their study used the 2016 General Social Survey (GSS) and measured mental well-being through whether respondents reported having ever felt they were going to have a nervous breakdown.

This article is a strong replication target because its main claim depends on a public dataset, a binary outcome, and logistic regression models with interaction terms. The statistical question is direct, yet the workflow requires many analytic decisions: selecting the correct GSS year, handling split-ballot data, converting reserved missing-value codes, constructing a social media count, creating age categories, and applying listwise deletion. Each decision may affect the final analytic sample and coefficient estimates.

This replication asks whether the central Hardy and Castonguay (2018) result is reproducible from the same public data source under a documented workflow. The project replicates two parts of the source article: descriptive statistics from the Methods section and the three logistic regression models in Table 2. The hypothesis tested is that social media use is associated with nervous breakdown reports and that this association varies by age group, with adults ages 18–29 showing a weaker association and adults ages 30–49 showing a stronger association.

2. Methods

2.1 Data Collection

The data source was the 2016 General Social Survey, a nationally representative survey of adults in the United States conducted by NORC. The GSS uses a split-ballot design, meaning not every respondent receives every question. As a result, the original article relied on a subset of respondents with nonmissing values for the outcome, social media variables, age categories, and control variables. Hardy and Castonguay (2018) reported an analytic sample of 743 respondents. This replication generated a descriptive-statistics sample of 951 respondents and a regression sample of 714 respondents.

2.2 Variable Creation

The dependent variable was EVBRKDWN, coded as a binary indicator for whether a respondent reported having ever felt close to a nervous breakdown. The main predictor was a count of regular social media platform use, constructed from the GSS questions for Facebook, Twitter, Instagram, and Snapchat. Age was represented with indicator variables for ages 18–29 and ages 30–49, leaving ages 50 and older as the reference group.

The models included control variables from the published study: sex, race, Hispanic identity, rural residence, religiosity, education, income, general happiness, and number of bad mental health days in the past month. GSS reserved codes for missing, inapplicable, skipped, and refused responses were treated as missing prior to analysis. Logistic regression models were estimated using complete cases across all variables included in each model.

2.3 Analytic Methods

Three logistic regression models were estimated. Model 1 included main effects only. Model 2 added the interaction between social media use and ages 18–29. Model 3 added the

interaction between social media use and ages 30–49. Odds ratios were reported for interpretation, and Nagelkerke R^2 was used to compare model fit with the published article.

Data management and analysis followed a reproducible script workflow. Separate scripts imported the raw GSS file, restricted the data to 2016, selected variables, cleaned reserved missing values, verified variable ranges, constructed analysis variables, created analytic samples, produced descriptive statistics, and estimated the three logistic regression models. A manager script executed the full sequence.

3. Results

The descriptive statistics replicated several published benchmarks closely and differed more sharply on others. Hardy and Castonguay (2018) reported 37.3% “Yes” for EVBRKDWN, a social media mean of 1.48, a bad mental health days mean of 3.67, a happiness mean of 2.12, and an average age of 44.76 years. The replication produced 34.7% “Yes” for EVBRKDWN, a social media mean of 1.11, a bad mental health days mean of 3.82, a happiness mean of 1.90, and an average age of 48.27 years. Mental health days and race estimates were close to the published benchmarks. Age, education, happiness, and social media use showed larger differences.

Table 1 summarizes the main replication evidence. The regression sample was smaller than the published sample by 29 respondents. Despite this difference, the model-fit statistics nearly matched after rounding. The replicated Nagelkerke R^2 values were 0.18, 0.19, and 0.18, compared with published values of 0.18, 0.19, and 0.19.

Table 1. Main Replication Evidence

Replication Target	Published Result	Replication Result
Regression sample size	743	714
EVBRKDWN “Yes”	37.3%	34.7%
Social media mean	1.48	1.11
Bad mental health days mean	3.67	3.82
Age mean	44.76	48.27
Nagelkerke R^2, Model 1	0.18	0.18
Nagelkerke R^2, Model 2	0.19	0.19
Nagelkerke R^2, Model 3	0.19	0.18
Social media, Model 1	positive	OR = 1.22*
Social media × Age 18–29	negative	OR = 0.59**
Social media × Age 30–49	positive	OR = 1.41*

The logistic regression results directionally replicated the published age-moderation

claim. In Model 1, each additional social media platform was associated with higher odds of reporting a nervous breakdown (OR = 1.22, $p < 0.05$). In Model 2, the interaction between social media use and ages 18–29 was negative (OR = 0.59, $p < 0.01$), meaning the social media association was weaker for the youngest adults than for adults ages 50 and older. In Model 3, the interaction between social media use and ages 30–49 was positive (OR = 1.41, $p < 0.05$), meaning the association was stronger for adults ages 30–49 than for adults ages 50 and older.

Several control variables followed stable patterns across models. Bad mental health days was a consistent positive predictor (OR = 1.06, $p < 0.001$ in all three models). Income was negatively associated with nervous breakdown reports (OR = 0.96, $p < 0.05$ in all three models). Female respondents and white respondents had higher odds of reporting a nervous breakdown across the replicated models. These results indicate that the social media and age interaction findings were not artifacts of a single model specification.

4. Discussion/Conclusions

This replication substantially reproduces the central Hardy and Castonguay (2018) result: the association between social media use and nervous breakdown reports differs by age group. The replicated social media main effect was positive, the interaction for ages 18–29 was negative, and the interaction for ages 30–49 was positive. Model-fit values also closely matched the published article after rounding.

The main limitation is the regression sample mismatch. The published article reported N = 743, while this replication produced N = 714. This difference likely reflects missing-data treatment, GSS reserved codes, platform-count construction, listwise deletion, or a combination of these decisions. Since logistic regression requires complete cases across the outcome, predictor, controls, and interaction variables, small coding differences may shift the final sample. The descriptive differences in age, education, happiness, and social media use also suggest that the replicated sample composition was not identical to the published analytic sample.

The outcome measure also limits interpretation. EVBRKDWN records whether a respondent ever felt close to a nervous breakdown. It is not a clinical diagnosis, and it does not identify timing. The 2016 GSS is observational and cross-sectional, so the estimates describe associations rather than causal effects. The findings should therefore be interpreted as a replication of an association pattern, not evidence that social media use causes mental-health outcomes.

Future work should test whether the results remain stable under alternative missing-data rules and alternative social media count construction rules. A sensitivity analysis that moves the replicated sample closer to the published N = 743 would strengthen the comparison. Additional research might examine later GSS waves, although survey-mode changes and changes in social media measurement would require careful treatment. The main contribution of this project is that it verifies the direction of the published age-moderation pattern while showing that transparent documentation of sample construction is essential for public-data replication.

References

Hardy, B. W., & Castonguay, J. (2018). The moderating role of age in the relationship between social media use and mental well-being: An analysis of the 2016 General Social Survey. *Computers in Human Behavior, 85*, 282–290.

Mangiafico, S. S. (2026). *rcompanion*: Functions to support extension education program evaluation (Version 2.5.2) [R package]. Rutgers Cooperative Extension. https://CRAN.R-project.org/package=rcompanion

NORC at the University of Chicago. (2017). General Social Survey, 2016 [Data file and codebook]. NORC at the University of Chicago.

Generative AI Statement

Generative AI tools were used during the original course project as learning aids for understanding R coding errors, debugging strategies, and example code logic. Generative AI was not used to conduct analyses, generate results, write, condense, edit, or revise this competition submission. All statistical decisions, data management choices, analyses, interpretation, writing, and final submitted content were completed independently.

Appendix

Appendix Table A1. Descriptive Statistics for Study Variables

Variable	Mean or Percent	SD
Happiness	1.90	0.66
Bad mental health days	3.82	7.26
Social media count	1.11	1.17
Religious attendance, days/year	22.27	30.52
Age	48.27	17.04
Years of formal schooling	13.97	2.97
Income, 26-category scale	17.39	5.89
Ever felt breakdown, 1 = Yes	34.7%	—
Female	57.6%	—
Race, White	78.9%	—
Rural	10.7%	—

Data: 2016 General Social Survey, N = 951.

Appendix Table A2. Logistic Regression Results Predicting Nervous Breakdown Reports

Predictor	Model 1 OR	Model 2 OR	Model 3 OR
Female	1.69**	1.63**	1.68**
Age 18–29	1.63	4.73***	2.03*
Age 30–49	2.06***	1.90**	1.32
White	2.37***	2.34***	2.37***
Hispanic	0.87	0.92	0.90
Religiosity	1.00	1.00	1.00
Education	0.99	0.99	0.99
Income	0.96*	0.96*	0.96*
Rural	1.01	1.00	1.01
Generally happy	1.26	1.26	1.23
Bad mental health days	1.06***	1.06***	1.06***
Social media	1.22*	1.42***	1.04
Social media × Age 18–29	—	0.59**	—
Social media × Age 30–49	—	—	1.41*
Nagelkerke R^2	0.18	0.19	0.18

Data: 2016 General Social Survey, N = 714. ***p < 0.001; **p < 0.01; *p < 0.05.