

The Power of Rumor Mill: A replication Study of Berinsky's (2017) Research on Political Rumors concerning the 2010 ACA

By DAN XU*

This project is a replication of Berinsky's study of political rumors surrounding the 2010 U.S. health care reform act. I re-examined the structure of the experimental design and re-analyzed the results generated by the survey experiment. I conducted chi - square tests to investigate the potential relationships between treatment groups, and the results are generally consistent with Berinsky's findings. To conclude, my results show that the relationships between control and treatment groups, in both the first and second studies, are not statistically significant. However, Berinsky did identify one significant relationship, which is between the "rumor only" and "rumor + Republican correlation"; in other words, he found that Republican debunking rumors is statistically related to rumor acceptance. I further conducted t - tests to compare group means of the treatment groups and did not discover any statistically significant relationships.

JEL: A10, A11

Keywords: Rumors, Political decision making

I. Experimental design

Berinsky conducted the survey experimental using a two wave online survey. From 17 - 19 May 2010, Berinsky performed a between subjects design experiment, in which the outcome is only measured post - treatment, online with a national sample of 1701 American adults. The second wave was administered to only 699 of the initial respondents from 25 - 29 May 2010. The experiment was conducted by Survey Sample International (SSI) - a U.S. based digital research business that offers survey sampling and related services for market survey research. The survey was constructed according to the the US adult population on education, gender, age, geography, and income.

A. Between and Within Subject Design

Widely used in Psychological studies, between-subject experimental designs are an experimental approach in which each subject is tested under one condition and only exposed to a single treatment. In Berinsky's experimental design, the

* Xu: University of Toronto, zedan.xu@mail.utoronto.ca. Acknowledgements

survey respondents were randomly assigned to four different treatment groups and one control group and presented with the stories regarding the controversial 2010 ACA. The following details the conditions which the survey respondents received,

1. Rumor only. In this group, respondents were presented with a rumor surrounding the 2010 ACA.
2. Rumor and Correction. In this treatment group, respondents were presented with a rumor and a correction debunking the rumor.
3. Rumor and Republican correction. In this treatment group, respondents were presented with a rumor and correction, which was a quote from Republican senator.
4. Rumor and Democratic correction. In this condition, respondents were presented with a rumor and a correction, which was a quote from a Democratic representative.
5. Control. In the control group, no experimental conditions were applied to the respondents.

I would like to further clarify the differences between between-subject (between - group) experimental designs and within-subjects (or repeated-measures) experimental designs. In a between-subjects design, or a between-groups design, each subject is only exposed to one condition. In a within-subjects design, however, each participant experiences all conditions. The same participants are tested repeatedly in order to assess differences between conditions. Therefore, if Berinsky were to conduct a within - subject experiment with the sample, the survey respondents would not be assigned to five groups (four treatment and one control). Instead, they would be exposed to all the same treatments to examine their reactions to the rumors.

B. Structure of the experimental design - Wave 1

Even though there are 1701 respondents who were surveyed, only 1593 complete cases were documented. 99 respondents have not completed both of the screeners. 1043 of the survey respondents passed Screener1, while 559 of them failed. Likewise, 1098 of the respondents passed the second screener, while 504 failed. Approximately two thirds of the respondents passed each of the screened questions, which confirms the author's results. Furthermore, 879 of the respondents passed both of the screened questions. Therefore, 54.9% of the respondents whose responses were recorded (excluding the incomplete cases that contain missing values) passed both questions, which corresponds to the author's results. As he states in the paper, 55% of the sample passed both questions.

The following are the variables for the first-wave survey:

"ACA" refers to "Support for Health Care Reform." The question asks "Overall, given what you know about them, would you say you support or oppose the

changes to the health care system that have been enacted by Congress and the Obama administration?” The answer choices are support (1) and oppose(2) .

death refers to “Death Panel Rumor.” The question asks “Do you think the changes to the health care system have that have been enacted by Congress and the Obama administration create “death panels” which have the authority to determine whether or not a gravely ill or injured person should receive health care based on their “level of productivity in society?” The answer choices are ”yes (1), no (2), not sure(3).

euth refers to “Euthanasia Rumor.” The question asks “Do you think the changes to the health care system have that have been enacted by Congress and the Obama administration require elderly patients to meet with government officials to discuss “end of life” options including euthanasia?” The answer choices are ”yes (1), no (2), not sure(3).

Screener 1 is the first of the two screened question the author used to to measure respondent attention on the self-administered survey. For this question, if the respondent passed the attention check, their answer would be documented as “1”, otherwise, “0.”

Screener 2 is the second screened question for attention check. “pass” is coded as “1”, and “fail” is coded as “0.”

treat specifies the condition to which each group of respondents were assigned. From 1 - 5, the groups are labelled as “control” “rumor only” “rumor+correct” “rumor+rep”, “rumor+dem.”

PID is a variable that documents one’s party identification. From 0 - 6, , it corresponds to “str rep” “wk rep” “lean rep” “ind” respectively, which translates to “Strong Republican”, “Weak Republican”, “Lean Republican”, “Independent” , “Lean Democrat”, “Weak Democrat”, “Strong Democrat” respectively.

The following analysis will be focused on the “attentive sample,” where the respondents passed both of the attention checks, according to Berinsky.

II. Full sample

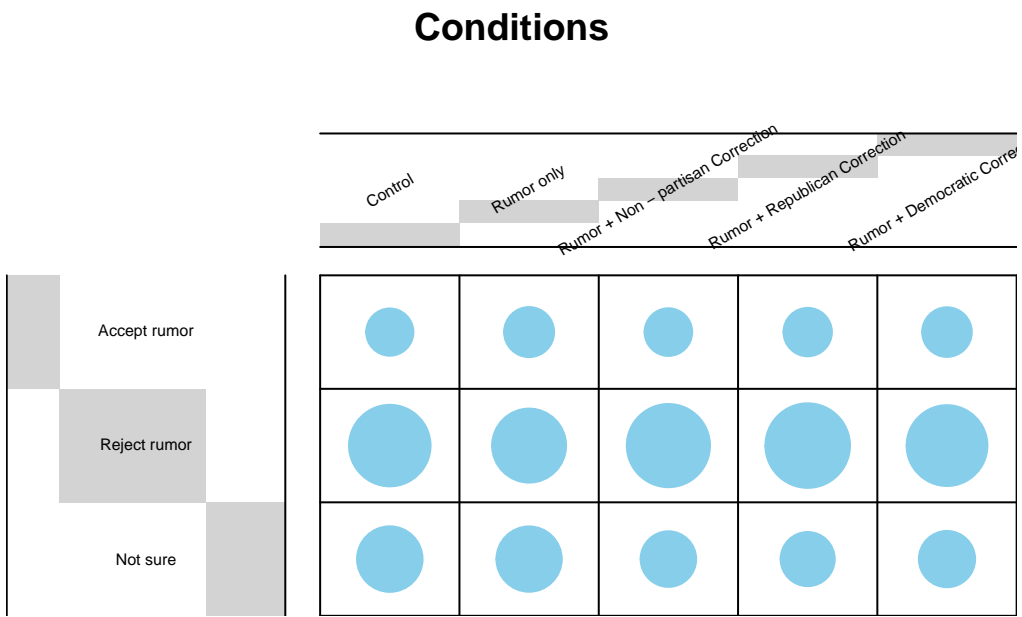
As stated earlier, the number of the complete cases for the dataset is 1593, instead of 1596. The researcher of the original study may need to clarify why 1596 cases were used for analysis in the study.

TABLE 1—TREATMENT GROUPS IN THE FULL SAMPLE

	Control	Rumor only	Rumor + Non - partisan Correction	Rumor + Republican Correction	Rumor + Democratic Correction
Accept rumor	57	64	58	60	63
Reject rumor	170	140	177	183	167
Not sure	110	109	79	75	81

Given the cross tabulation, I performed a Pearson chi-square test, which is

essentially a test to examine whether the results of a crosstab are statistically significant. That is, if the two categorical variables independent of one another. The statistics of the significance test are as follows: $N = 1593$, $\chi^2(8) = 18.819$, $Pr = 0.015$, which are s consistent with Berinsky’s finding.



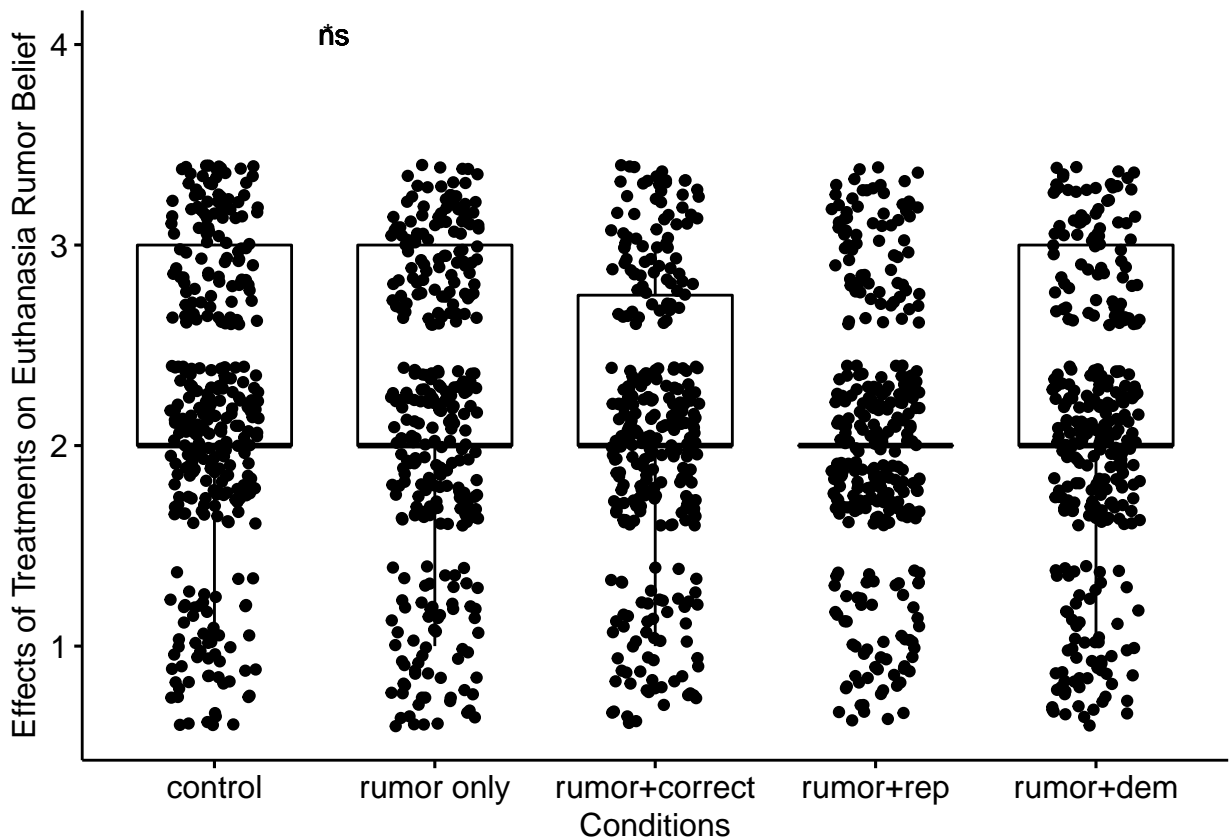
The graph above also gives us an intuitive impression of the potential relationship between the categorical variables.

A. Effects of treatments on Euthanasia rumor belief

To investigate the effects of treatments on Euthanasia rumor belief, I first presented a statistics summary table for each treatment group. As seen in the table, there seems to be relatively little variation across the five treatment groups. The means and standard deviations of the conditions are similar, suggesting that the effects are minimal across treatment groups.

TABLE 2—EFFECTS OF RUMOR ON TREATMENTS

treat	variable	n	mean	sd
control	euth	337	2.157	0.687
rumor only	euth	313	2.144	0.731
rumor+correct	euth	314	2.067	0.658
rumor+rep	euth	318	2.047	0.651
rumor+dem	euth	311	2.058	0.679



One of the most common methods for comparing means is t-test, which is also the method Berinsky used to compared treatment group means. As the graph

shows, there was not a significant difference between the control and treatment groups.

TABLE 3—SIGNIFICANCE TEST RESULTS BETWEEN CONDITIONS

.y.	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
euth	control	rumor only	337	313	0.2422141	636.4013	0.809	1.000	ns
euth	control	rumor+correct	337	314	1.7140886	648.5025	0.087	0.641	ns
euth	control	rumor+rep	337	318	2.1058857	652.9904	0.036	0.356	ns
euth	control	rumor+dem	337	311	1.8507636	642.9751	0.065	0.582	ns
euth	rumor only	rumor+correct	313	314	1.3843712	617.9056	0.167	0.835	ns
euth	rumor only	rumor+rep	313	318	1.7527597	618.4225	0.080	0.641	ns
euth	rumor only	rumor+dem	313	311	1.5212066	619.2563	0.129	0.774	ns
euth	rumor+correct	rumor+rep	314	318	0.3784739	629.6417	0.705	1.000	ns
euth	rumor+correct	rumor+dem	314	311	0.1682396	621.9619	0.866	1.000	ns
euth	rumor+rep	rumor+dem	318	311	-0.2018223	624.3881	0.840	1.000	ns

Berinsky conducted two sets of Chi-squared Tests of Independence, also referred to as a χ^2 tests, to examine the significance relationships across conditions. The first χ^2 test was to examine the overall significance in order to examine if any significant differences among any of the conditions exist. Likewise, I further conducted a significance test to investigate the relationships between groups. Given that there are five conditions, therefore, ten pairs of comparisons were calculated. By comparing one condition to another, as the table shows, the p values are greater than 0.05, and therefore, I conclude that there are no relationships between each groups. In other words, a heterogeneous treatment effect between the groups does not exist according to the statistical results. However, Berinsky identified one significant relationship amongst the ten pairs, which is that the “rumor only” condition is statistically significantly different from the “rumor + Republican correction”. My results suggest otherwise, partisanship is not a factor in rumor acceptance in this particular case. The Republican correction is not a powerful treatment for rumor belief.

III. Attentive sample

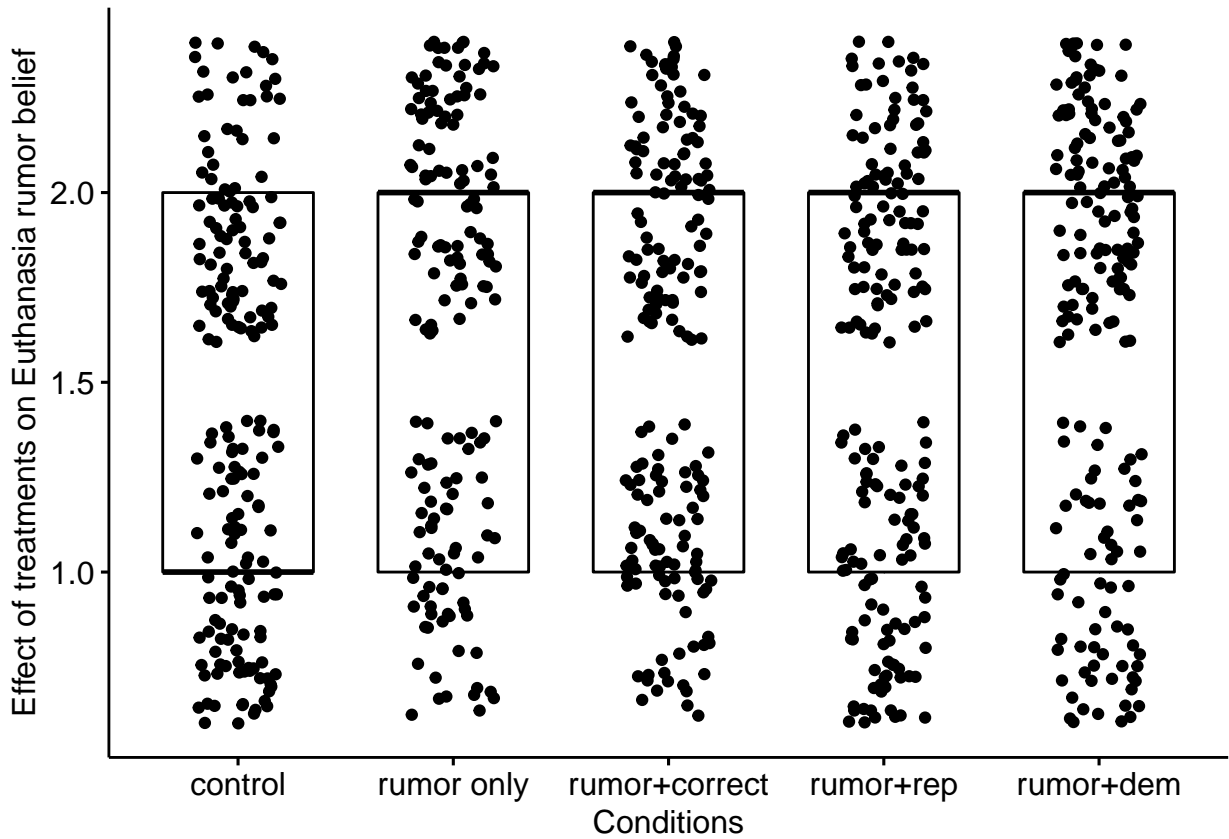
To ensure that the respondents paid close attention to the survey experiment, Berinsky applied two attention checks in this study. The attention check questions are designed to make sure that survey takers’ attention is at a high level throughout the entire survey. For those who have successfully and correctly answered both of the questions, Berinsky further categorized them into a subsample, which is called “attentive sample” accordingly.

Likewise, significant tests were performed on the subsample, which only includes complete cases where both of the attention checks are marked as “pass”. In total, there were 879 respondents who passed both of the attention checks.

TABLE 4—ATTENTIVE SAMPLE

	Control	Rumor only	Rumor + Non - partisan Correction	Rumor + Republican Correction	Rumor + Democratic Correction
Support	94	65	81	87	65
Oppose	91	90	96	94	113

By performing a Chi-square test, the following statistics were reported: N = 879. X-squared = 23.946, df = 8, p-value = 0.00234, which are consistent with Berinsky’s finding.



For the sub-sample, the attentive sample according to Berinsky, I performed significance tests amongst the ten pairs of treatment groups, in order to examine the relationship between the treatment groups and the effects of treatments on belief of the Euthanasia rumor. As the t - test and significance test results show, there is no statistically significant relationship between the two variables.

TABLE 5—SIGNIFICANCE TESTS OF ATTENTIVE SAMPLE

.y.	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
euth	control	rumor only	185	155	-1.6370006	329.0199	0.103	0.721	ns
euth	control	rumor+correct	185	177	-0.9593972	359.3963	0.338	1.000	ns
euth	control	rumor+rep	185	181	-0.5238201	363.8340	0.601	1.000	ns
euth	control	rumor+dem	185	178	-2.7672907	360.9995	0.006	0.059	ns
euth	rumor only	rumor+correct	155	177	0.6997569	324.9972	0.485	1.000	ns
euth	rumor only	rumor+rep	155	181	1.1253546	327.2372	0.261	1.000	ns
euth	rumor only	rumor+dem	155	178	-1.0078028	322.3540	0.314	1.000	ns
euth	rumor+correct	rumor+rep	177	181	0.4355653	355.8630	0.663	1.000	ns
euth	rumor+correct	rumor+dem	177	178	-1.7728210	352.4419	0.077	0.617	ns
euth	rumor+rep	rumor+dem	181	178	-2.2241205	356.8546	0.027	0.241	ns

A. Effect of treatments on health care policy opinion (attentive sample)

TABLE 6—HEALTH CARE POLICY OPINION (ATTENTIVE SAMPLE)

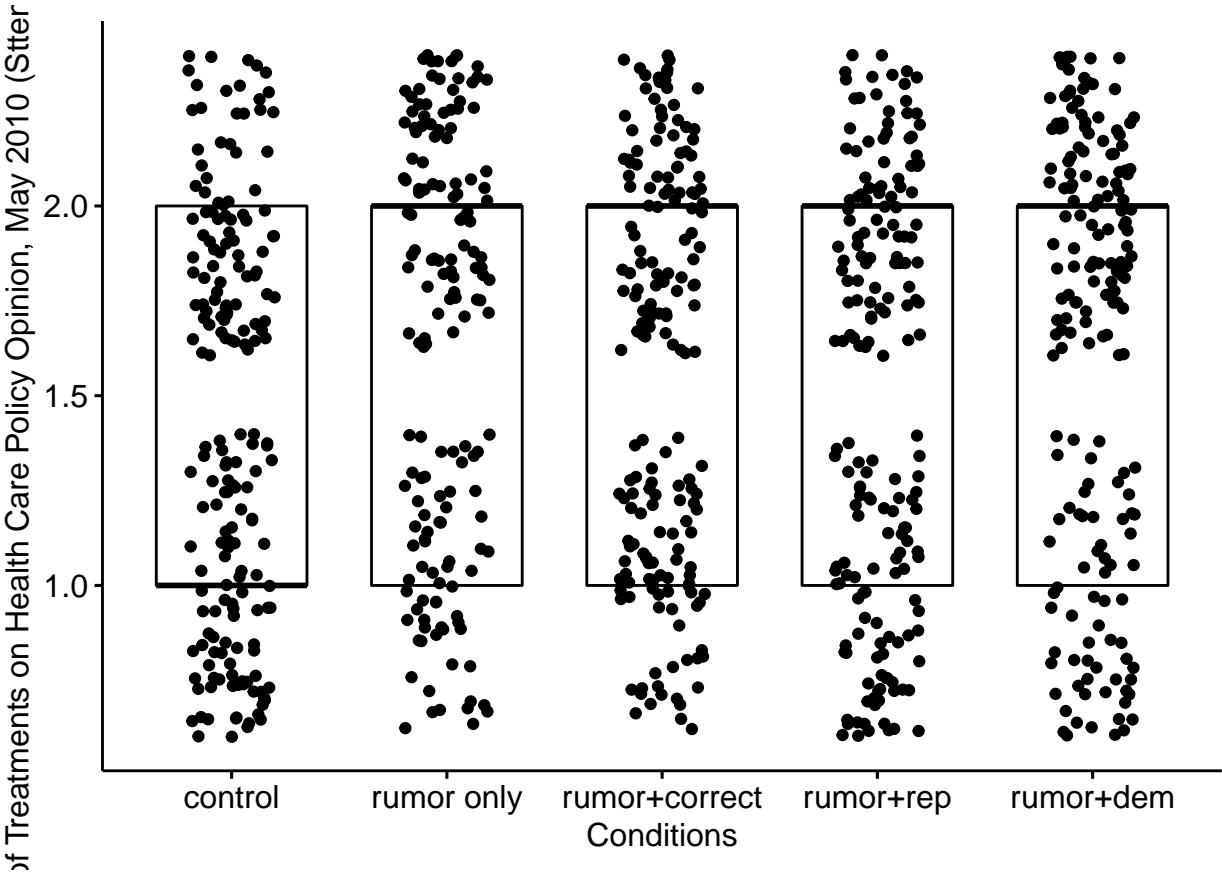
	Control	Rumor only	Rumor + Non - partisan Correction	Rumor + Republican Correction	Rumor + Democratic Correction
Support	94	65	81	87	65
Oppose	91	90	96	94	113

The above table shows the health care policy opinion by each treatment group for the attentive sample.

The statistics of the Chi-square test are reported as follows, $N = 879$, $\chi^2 = 9.0038$, $p\text{-value} = 0.061$, which is greater than 0.05. Such statisticals are consistent with Berinsky's finding.

TABLE 7—STATISTICS OF SUPPORT FOR HEALTH CARE POLICY REFORM (ATTENTIVE SAMPLE)

treat	variable	n	mean	sd
control	ACA	185	1.492	0.501
rumor only	ACA	155	1.581	0.495
rumor+correct	ACA	177	1.542	0.500
rumor+rep	ACA	181	1.519	0.501
rumor+dem	ACA	178	1.635	0.483



Again, none of the relationships are statitically significant. Overall the results are consistent with Berinsky’s findings. However, he did identified one statis-

TABLE 8—HEALTH CARE POLICY OPINION (ATTENTIVE SAMPLE)

.y.	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
ACA	control	rumor only	185	155	-1.6370006	329.0199	0.103	0.721	ns
ACA	control	rumor+correct	185	177	-0.9593972	359.3963	0.338	1.000	ns
ACA	control	rumor+rep	185	181	-0.5238201	363.8340	0.601	1.000	ns
ACA	control	rumor+dem	185	178	-2.7672907	360.9995	0.006	0.059	ns
ACA	rumor only	rumor+correct	155	177	0.6997569	324.9972	0.485	1.000	ns
ACA	rumor only	rumor+rep	155	181	1.1253546	327.2372	0.261	1.000	ns
ACA	rumor only	rumor+dem	155	178	-1.0078028	322.3540	0.314	1.000	ns
ACA	rumor+correct	rumor+rep	177	181	0.4355653	355.8630	0.663	1.000	ns
ACA	rumor+correct	rumor+dem	177	178	-1.7728210	352.4419	0.077	0.617	ns
ACA	rumor+rep	rumor+dem	181	178	-2.2241205	356.8546	0.027	0.241	ns

tically significant relationship between “rumor only” and “rumor + Republican correction.” He also considered the pairs with p values that are slightly greater than 0.05 marginally statistically significant. My statistical results suggest that none of the relationships are considered statistically significant in Study 1.

IV. Study 2

A few months after the first study, in which Berinsky surveyed 1701 American adults for their opinions on a political rumor surrounding the ACA 2010, Berinsky conducted a second experiment on October/November 2010. However, the experiment was administered by YouGov instead of Survey Sampling International this time. The subject of the second experiment was still rumors concerning the ACA 2010. According to Berinsky, the stories presented to the respondents were modeled on the stories from Study 1. However, the conditions were modified. The control condition was identical to the “rumor” condition from Study 1. The second “correction only” was altered and did not mention the rumor concerning the death penalty, but only with the description of the provisions in the 2010 ACA. The third condition that was identical to the “rumor and Republican correction” condition from Study 1 and presented both the rumor from the first condition and the correction from the second condition. For Study 2, Berinsky did not apply attention checks. In addition, no control group was included in this follow-up study.

In this study, the respondents were presented with two types of recall questions, where they were asked to identify the person who said the quote. The sample was split into two groups for comparison. Half of the survey takers were assigned to an “irrelevant” recall condition, in which the question asks what office was held by Besty McCaughey, while the other half was placed in a “long recall” condition, where they received two questions with identical answer choices. The first recall question asks “You have every right to fear... [You] should not have a

government-run plan to decide when to pull the plug on Grandma.” The second recall questions asks “The health care reform bill requires “people in Medicare have a required counseling session that will tell them how to end their life sooner.”

This dataset contains all the 1000 cases Berinsky collected for the second study, which is referred to as CCES. The following are the variables,

“randtreat” refers to the story condition, including the three conditions, namely ‘Rumor only’, ‘Correction only’, ‘Rumor + Correction’. Please note that no control group is included in this study.

“randtreat2” refers to the first recall condition, including ‘Short recall’ and ‘Long recall’ questions

” “ACA” refers to whether the respondent supports or opposes the act.

“euth_w1” refers to whether the Euthanasia rumor was believable in Wave 1.

“euth_w2” refers to whether the Euthanasia rumor was believable in Wave 2.

Please note that Berinsky conducted two waves of the study, in October and November, respectively, and the belief question was asked in both waves. Therefore, “euth_w1” and “euth_w2” ask the same question. However, support for the health care reform bill (“ACA”) was only asked in Wave 1.

TABLE 9—STUDY 2 (WAVE 1)

	Rumor only	Correction only	Rumor + Correction
Accept rumor	102	71	67
Reject rumor	168	167	206
Not sure	80	70	69

In general, we can observe from the table that respondents tended to reject the rumor. The statistics of the Chi - square test are consistent with Berinsky’s finding: $N = 1000$, $\chi^2(4) = 12.237$, $p\text{-value} = 0.01567$, which is smaller than 0.05, indicating that the relationship is significant.

The Chi square test statistics are $X\text{-squared} = 3.7756$, $df = 3$, $p\text{-value} = 0.2867$, suggesting that the relationship is not statistically significant. That is, there is no difference between the effects of irrelevant or long recall questions. Respondents exposed to either conditions showed no difference in accepting political rumors.

To gauge the number of respondents who completed both waves, I used the variables euth_w1 and euth_w2 in the dataset provided by Berinsky. Given that all the 1000 respondents have provided a valid response for the question concerning euth_w1, I can conclude that the number of respondents who completed the first wave is 1000, and the number of the respondents who completed both waves depends on the second wave. 165 Respondents have not provided a answer

TABLE 10—RECALL EFFECTS IN STUDY 2 (WAVE 2)

	Short recall	Long recall
Strongly support	105	94
Somewhat support	124	111
Somewhat oppose	80	72
Strongly oppose	192	221

to the question concerning euth_w2, therefore, the number of respondents who completed both waves should be 835 instead of 837.

TABLE 11—RECALL EFFECTS IN STUDY 2 (WAVE 1)

	Rumor only	Correction only	Rumor + Correction
Accept rumor	79	61	52
Reject rumor	137	144	181
Not sure	73	56	52

The statistics are: X-squared = 15.172, df = 4, p-value = 0.004357, which is less than 0.05, suggesting the overall significance between the three treatment conditions. This finding is consistent with Berinsky's.

The statistics are X-squared = 5.0398, df = 4, p-value = 0.2832. Which again confirms Berinsky's finding.

A. rehearsal effect

For "rumor only" condition, the statistics are N= 289, X-squared = 4.0712, df = 2, p-value = 0.1306.

The statistics for the "rumor + correction" condition are N = 285, X-squared = 2.4477, df = 2, p-value = 0.2941

Again it seems that there is no statistically significant relationship between the "recall" condition and rumor acceptance. My results are consistent with Berinsky's, merely increasing the fluency of the rumor increases its effectiveness.

TABLE 12—RECALL EFFECTS IN STUDY 2 (WAVE 2)

	Rumor only	Correction only	Rumor + Correction
Accept rumor	76	68	66
Reject rumor	135	138	156
Not sure	78	55	63

TABLE 13—RECALL EFFECTS ON RUMOR ONLY CONDITION IN STUDY 2 (WAVE 2)

	Short recall	Long recall
Accept rumor	35	41
Reject rumor	74	61
Not sure	32	46

V. Discussion and conclusions

Overall my results are consistent with Berinsky’s findings, suggesting that the studied Berinsky conducted are replicable.

For Study 1, Berinsky concludes that corrections from Republicans debunking health care rumors are the most effective way to counter misinformation. However, my replicated results show that none of the relationships amongst the control and treatment groups are significant. In other words, my results do not lend support to Berinsky’s conclusion. We did not identify other statistically significant relationships between the treatment groups. However, the results of the t - tests and significance tests I conducted lend strong support to the conclusion that there is no statistically significant relationship among the variables.

For Study 2, our results are consistent. My results show that Study 2 replicates the findings from Study 1, no statistically significant relationships existed among the treatment groups. The hypothesis that fluency of the rumor may increase effectiveness of rumor spreading does not gain sufficient support.

TABLE 14—RECALL EFFECTS ON RUMOR + CORRECTION CONDITION IN STUDY 2 (WAVE 2)

	Short recall	Long recall
Accept rumor	31	35
Reject rumor	84	72
Not sure	27	36