Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

PUBLICATION BIAS AND EDITORIAL STATEMENT ON NEGATIVE FINDINGS

Cristina Blanco-Perez, Abel Brodeur

Working Paper No: 190001

www.canadiancentreforhealtheconomics.ca

August, 2019

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 190001
August, 2019

Publication Bias and Editorial Statement on Negative Findings[*]

Cristina Blanco-Perez, Abel Brodeur

Abstract

In February 2015, the editors of eight health economics journals sent out an editorial statement which aims to reduce the extent of specification searching and reminds referees to accept studies that: "have potential scientific and publication merit regardless of whether such studies' empirical findings do or do not reject null hypotheses". Guided by a pre-analysis, we test whether the editorial statement decreased the extent of publication bias. Our differences-in-differences estimates suggest that the statement decreased the proportion of tests rejecting the null hypothesis by 18 percentage points. Our findings suggest that incentives may be aligned to promote more transparent research.

JEL Classification: A11; C13; C44; I10

Keywords: Publication bias; specification searching; pre-analysis plan; research in economics; incentives to publish

[*]Blanco-Perez, University of Ottawa, cblancop@uottawa.ca. Brodeur, University of Ottawa, abrodeur@uottawa.ca. We thank Pierre Brochu, Garret Christensen, Chishio Furukawa, Jason Garred, Fernando Hoces de la Guardia, Edward Leamer, Ted Miguel and seminar participants at the ACFAS meeting, the BITSS annual meeting and the WEAI annual conference for useful comments. Funding for this research was provided by the Berkeley Initiative for Transparency in the Social Sciences, a program of the Center for Effective Global Action (CEGA), with support from the Laura and John Arnold Foundation/William and Flora Hewlett Foundation. We thank Joanne Haddad and Yuvraj Pathak for outstanding research assistance. All errors are our own.

Our study deals with the selective reporting and abuse of statistical significance. Researchers are subject to an increasing pressure to publish their results in top journals and may thus select a subset of positive results from a larger set of possible specifications (Abadie (2018); Fanelli (2010); Gerber and Malhotra (2008a); Gerber and Malhotra (2008b); Gerber et al. (2010); Henry (2009); Leamer (1983); Ioannidis (2005); McCloskey (1985); Ridley et al. (2007); Simmons et al. (2011); Stanley (2005)). Many papers also point out that there is a selection bias in favor of positive results by editors and referees (Ashenfelter and Greenstone (2004); Fanelli (2009); Franco et al. (2014); Havránek (2015); Havránek and Sokolova (2016); McCloskey (1985); Stanley (2008)). These practices (i.e., specification searching and publication bias) may lead to a high percentage of false positives and to policy-makers seeing only a subset of the research (Brodeur et al. (2018); Card and Krueger (1995); De Long and Lang (1992); Doucouliagos and Stanley (2013); Furukawa (2017); Ioannidis et al. (2017)).

In this study, we test the impact of a simple, low-cost, new transparent practice that aims to reduce the extent of publication bias and the incentives to engage in specification searching. In February 2015, the editors of eight health economics journals sent out an editorial statement[1] reminding authors to submit and referees to not be biased against studies that: "have potential scientific and publication merit regardless of whether such studies' empirical findings do or do not reject null hypotheses that may be specified." This article assesses the extent to which the Editorial Statement on Negative Findings increased the proportion of papers with negative results in health economics journals.

Prior to data collection, the analysis presented here was pre-specified and publicly archived in a pre-analysis plan (PAP).[2] Our pre-specification

[1] The editorial statement is reproduced in the Appendix.

[2] Our pre-analysis plan was archived on November 28, 2016, at https://osf.io/mjbj2/ (Blanco-Perez and Brodeur (2017)). We had to register the pre-analysis plan before obtaining the funding from the Berkeley Initiative for Transparency in the Social

was designed to minimize issues of data and specification mining and to provide a record of the full set of planned analyzes. The use of PAPs is now common in field experiments, but it is much less common for "natural" experiments (see, for instance, Neumark (2001)).[3]

The hypotheses to be tested were specified in the PAP; (1) the Editorial Statement on Negative Findings increased the number of papers not rejecting the null hypothesis published in the health economics journals, and (2) the Editorial Statement on Negative Findings induced a change in the behavior of researchers, referees and editors.

To guide our empirical work, we wrote a conceptual framework in the PAP for understanding how the editorial statement might impact publication bias. The conceptual framework from the PAP is entirely reproduced in Section I. The editorial statement may affect the behavior of both authors and referees. For authors, the editorial statement could have decreased the extent of specification searching, but also increased the perceived likelihood of acceptance in the health economics journals for negative findings. The latter could lead to an increase in the number of submissions of papers not rejecting the null hypothesis to these journals. For referees, the statement may affect their behavior by changing their beliefs about editors' preferences for positive or negative findings, *ceteris paribus*. The editorial statement could also change their perceptions of the degree of selective reporting in health economics.

Following the PAP, we use a differences-in-differences approach for the evaluation of the pre-analysis plan. We first compare the distribution of tests before and after the editorial statement for health economics journals (i.e., treated journals) and then rely on two non-health economics journals as control journals. We collect z-statistics from five of the eight health

Sciences for the data collection.

[3]See Casey et al. (2012) and Olken (2015) for a discussion of the benefits and limitations of pre-analysis plans.

economics journals that sent out the editorial statement. We find that test statistics in papers submitted and published after the editors sent out the editorial statement are less likely to be statistically significant. Using a simple pre/post difference for treated journals, we find that the editorial statement decreased by about 12 percentage points the number of test statistics that are statistically significant at the 5% (and 10%) level. In other words, the editorial statement shifted the distribution of tests to the left and (arguably) decreased the extent of publication bias.

We then look at whether there was a similar shift in the distribution of z-statistics at the time of the editorial statement for two non-health economics journal. On the contrary, we find that the distribution of z-statistics shifted to the right after the editorial statement for our control journals, possibly due to the increasing difficulty to publish in top journals (Card and DellaVigna (2013)). Our differences-in-differences estimates suggest that the editorial statement decreased the proportion of tests rejecting the null hypothesis by about 18 percentage points. We also document that the impact of the statement "intensifies" over the time period studied. As a (pre-specified) robustness check, we test whether the decrease in the share of test statistics that are statistically significant after the editorial statement is driven by a change in the characteristics of the papers published. We find that the decrease in publication bias is not related to a change in the share of papers that are single-authored or have a theoretical model.

Our findings suggest that a simple transparent practice such as the editorial statement may decrease the extent of publication bias. But it is unclear whether the effect is arising from a displacement of positive results into non-health economic journals and the displacement of null results into the health economic journals. More generally, our pre-registered analysis cannot disentangle whether the editorial statement changed authors' research practices. In order to shed light on the mechanisms at play, we collect data on a large number of working papers in the field of health economics over the period 2014–2018. We find that the proportion of tests that are statistically significant at conventional levels decreased during this time period, suggesting that the behavior of health economists did change over time. But the lack of a comparison group and the fact that this exploratory analysis was not pre-registered prevent us from concluding that the editorial statement changed authors' behavior.

Last, we show that the impact factor of the health economics journals remained stable for the years before and after the editorial statement (Appendix Figure A1). This is suggestive evidence that the statement has not affected the yearly average number of citations to recent articles published in the health economics journals.

Our research question is directly related to research culture and adoption of new editorial policies (Dufwenberg and Martinsson (2014); Miguel et al. (2014); Christensen and Miguel (2018); Nosek et al. (2015)). Many outlets have implemented data and code availability policies to at least make replication theoretically possible (Blanco-Perez and Brodeur (2019); Camerer et al. (2016); Chang and Li (2015); Dewald et al. (1986); Maniadis et al. (2017); McCullough et al. (2008)).[4] Another new transparent practice was implemented by the journal *Psychological Science* in January 2014 (Kidwell et al. (2016)). This psychology journal offered badges to published articles if the authors reported open data and materials. This new method increased data reporting by about 35 percentage points. We contribute to this literature by providing empirical evidence that a simple editorial statement may decrease the extent of publication bias and that incentives may be aligned to promote a more transparent research.

Our findings relate to Andrews and Kasy (2017), Brodeur et al. (2016)

[4]Brodeur et al. (2016) provide suggestive evidence that data and code availability policies implemented in top economic journals did not change the extent of specification searching.

and Vivalt (2017), who document the extent of publication bias in economics. Andrews and Kasy (2017) develop two approaches for identifying and correcting publication bias in meta-analyses and systematic replications. Brodeur et al. (2016) collect p-values from three prestigious economics journals and document the extent of specification searching and publication bias. They show that the size of the bias is related to authors, and papers' characteristics. Vivalt (2017) confirms one of their results by providing evidence that the extent of publication bias is smaller for RCTs papers than for quasi-experimental studies.

Our research is also linked to meta-analyzes since we are interested in collecting data from scientific articles in economic journals. Many papers point out that there is a selection bias in favor of positive results in meta-analyses (Ashenfelter et al. (1999); Fanelli (2009); Franco et al. (2014); Havránek (2015); Havránek and Sokolova (2016)). Ioannidis et al. (2017) point out that altering incentives towards transparent research is important for economic science and that reducing publication bias and increasing statistical power are necessary steps to increasing credibility.

The rest of the paper is organized as follows. Section I describes the editorial statement and provide some predictions. Section II details the methodology used to construct the data set and provides descriptive statistics. We discuss the results and the mechanisms in sections III and IV. Section V concludes.

I Background and Conceptual Framework

In this section, we describe how the Editorial Statement on Negative Findings might have induced a change in the behavior of researchers, referees and editors. We reproduce the editorial statement in the Appendix (Section VI).

A Editorial Statement

In February 2015, the editors of eight health economics journals published on their journals' websites an Editorial Statement on Negative Findings. In this statement, the editors express that: "well-designed, well-executed empirical studies that address interesting and important problems in health economics, utilize appropriate data in a sound and creative manner, and deploy innovative conceptual and methodological approaches [...] have potential scientific and publication merit regardless of whether such studies' empirical findings do or do not reject null hypotheses that may be specified."

There was a great deal of support for this editorial statement on Twitter, and some blogs posted the statement or an abstract of the statement. As of May 2019, the statement was still highlighted on many of the health economics journals' websites.

The editors point out in the statement that it: "should reduce the incentives to engage in two forms of behavior that we feel ought to be discouraged in the spirit of scientific advancement: 1. Authors withholding from submission such studies that are otherwise meritorious but whose main empirical findings are highly likely 'negative' (e.g., null hypotheses not rejected). 2. Authors engaging in 'data mining,' 'specification searching,' and other such empirical strategies with the goal of producing results that are ostensibly 'positive' (e.g., null hypotheses reported as rejected)."

B Editorial Statement and Incentives

The editorial statement can affect researchers' behavior in many ways. Readers should keep in mind that the statement could affect both the pool of papers submitted and those accepted in health economics journals. For instance, researchers could now select specifications that produce negative results instead of positive results. The editorial statement could also decrease the extent of specification searching by researchers (e.g., decrease the number of regressions the authors run). This in turn would lead to a decrease in the proportion of test statistics rejecting the null hypothesis for papers submitted to the health economics journals.

Furthermore, the statement may have affected health economists' decision to submit to one of the health economics journals instead of general audience outlets (e.g., *Economic Journal*) or other field journals (e.g., *Journal of Public Economics*). Authors with a manuscript not rejecting the null hypothesis might have thought that they were suddenly more likely to get published in one of the health economics journals. It is thus plausible that the editorial statement affected researchers' behavior in the short run.

Since the editorial statement increased the expected returns of producing a paper not rejecting the null hypothesis, researchers working in the field of health economics may also have started to work on new projects that were more "risky," i.e., research projects that were less likely to reject null hypotheses. This could lead to a change in the distribution of published test statistics in the medium or long run.

It is unclear whether the editorial statement would lead to a simple substitution between publishing in a non-health economics outlet and publishing in one of the health economics journals. We tackle this issue in Section IV by collecting test statistics on a large sample of health economics working papers. This exercise allows us to check whether positive results are not simply displaced into non-health economics outlets.

The editorial statement may also have affected referees' behavior. The editorial statement could increase the likelihood that referees will advise the editors to accept papers not rejecting the null hypothesis. Papers that would typically end up unpublished or published in other outlets could now be published in one of the eight health economics journals. This is plausible for two reasons. First, the editors mention in the editorial statement that they "will remind our referees of this editorial philosophy at the time they are invited to review papers." The editorial statement indicates to referees that editors do not have preferences for positive or negative findings, *ceteris paribus*. A copy of the referee-invitation letter from *Health Economics* is available at the end of the Appendix. Second, the editorial statement might have changed referees' preferences for negative findings. For example, the editorial statement and the reminder might have changed their perception of the extent of publication bias in health economics.

A third group that could be affected by this editorial statement is comprised of the editors themselves. It remains unclear whether their preferences for negative findings have changed over time. Arguably, the editors may have changed their preferences for publishing studies not rejecting null hypotheses before February 2015. This would thus lead us to underestimate the impact of the editorial statement. We tackle this issue in Section III.

We asked some of the editors of the health economics journals whether more papers not rejecting the null hypothesis were now submitted to their journal. Two co-editors answered our questions and both speculated that the editorial statement led to a somewhat higher fraction of papers with negative findings submitted to their journal. One of the co-editors and an associate editor also mentioned that the editorial statement has not changed her/his propensity to desk reject. While these answers are not based on hard data, we believe they provide some insights on the effect of the statement on authors and editors' behavior.

To sum up, the Editorial Statement on Negative Findings might have increased the number of papers published in the health economics journals not rejecting the null hypothesis. We confirm in Section III that the share of test statistics not rejecting the null hypothesis increased after February 2015 and test whether this effect has intensified over time.

II Methods

In this section, we first discuss the data collection method and identification strategy. We then provide summary statistics and describe the raw distribution of tests before the editorial statement.

A Data Collection and Classification of Journal Articles

The objective of this paper is to investigate whether the editorial statement increased the number of studies with negative findings, i.e., not rejecting the null hypothesis.

One methodology used to measure the extent of specification searching in the literature is the "caliper test." This methodology is based on the rate of reported occurrence of test statistics within a narrow band (Brodeur et al. (2018); Gerber and Malhotra (2008a); Vivalt (2017)). The existing economic literature identifies an unusually large number of observations just over the critical value associated with 0.10, 0.05 and 0.01 p-values. In the present study, we are not analyzing the number of test statistics within a narrow band. We are analyzing the whole distribution of tests and interested in testing whether the editorial statement changed the share of tests that are statistically significant. In other words, our analysis is less local, and what matters is whether the number of observations over the critical value decreases after the introduction of the editorial statement. This decision was based on the fact that the editorial statement might increase the number of published studies with a high p-value. Further research could rely on the caliper test method to test whether the editorial statement specifically changed the extent of specification searching.

More precisely, we look at the whole distribution of p-values before the editorial statement and compare it with distributions of p-values in non-health economics journals. We then examine the ratio of reported results above and below the three significance thresholds before and after the editorial statement was issued. This methodology allows us to test explicitly whether the editorial statement induced a change in the ratio of test statistics not rejecting the null hypothesis.

The data collection follows a PAP that was laid out in three steps: (1) we extract journal articles from multiple volumes of health economics journals that sent out the editorial statement; (2) classify journal articles with respect to the dates of submission and publication; and (3) manually collect the z-statistics.

To gauge whether the editorial statement had an impact on the distribution of test statistics, we assemble a group of research publications in five health economics journals: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. We focus on these health economics journals because information on the submission and acceptance dates for each article is available.[5]

The data collection described above provides 230 articles, of which 112 are from the *Journal of Health Economics* and 71 from *Health Economics*. The other health economics outlets provide less journal articles. Our sample includes 11 papers from the *Health Economics Review*, 14 papers from the *International Journal of Health Economics and Management* and 22 journal articles from the *European Journal of Health Economics*. We also collect test statistics for two control journals not affected by the editorial statement. This allows us to check whether the distribution of tests in health economics journals is changing over time for reasons other than the editorial statement. In our PAP, we mentioned that we would be us-

[5]Submission and acceptance dates are not available for the *Forum for Health Economics & Policy* and the *Health Economics, Policy and Law*. One exception is the *American Journal of Health Economics*. As mentioned in our pre-analysis plan, we exclude this journal since it was founded in 2015.

ing the *Journal of Public Economics* as a control journal. Note that we pre-specified this control journal since it publishes articles relying on similar methods (e.g., differences-in-differences). This journal publishes few health economics papers per year. We omit these health economics papers throughout.[6] We also rely on *Labour Economics* as a second control journal. We did not pre-specify this journal. Note that relying solely on the *Journal of Public Economics* as a control journal has no effect on our main conclusions.

Following the PAP, we classify papers in three categories. The first category, "Before," includes papers that were submitted and published before the new editorial statement. The second category, "During," includes papers that were submitted before the statement on negative findings, but published after. The last group, "After," includes papers submitted and published after the editorial statement was sent to the referees and the academic community. We rely on the submission and acceptance dates to classify papers into one of the three categories.

All articles that include at least one empirical tests are included in the sample. We report solely tests from the main tables of the paper. We report all the main results, even if the authors state in the comments that it is a good thing that an estimate is not statistically significant at conventional levels. We define variables of interest (i.e., main results) by looking at the purpose of each table and at comments of the table in the text. We collect coefficients drawn from multiple specifications of the same hypothesis. We exclude descriptive statistics, group comparisons, and explicit placebo tests. We exclude tables that are labeled as robustness checks in the text (or title of the table). In cases of ambiguity we err on the side of exclusion.

We do not report first-stages in the case of instrumental variable (nor matching) unless the authors describe it as a major contribution to the

[6] Including the few health economics papers has no effect on the share of tests statistically significant at conventional levels for the control journal.

paper. For differences-in-differences, we collect only the interaction term, unless the non-interacted terms are described by the author(s) as coefficients of interest.

Authors report their estimates in different ways. Most authors present coefficients and standard errors, but some tables report p-values. Following our PAP, we transform these statistics to obtain a homogenous sample. We simply take the ratio of the coefficients and standard errors and transform the p-values into the equivalent z-statistics. Note that the distribution of a z-statistic is standard normal which means that the level of rejection is inadequate when the sample size is small. We collect all reported decimal places. We do not round up the coefficients and standard errors. We make sure that all test statistics are two-tailed tests.

Last, we collect additional information on each article. We collect the number of authors, JEL codes when available, the presence of a theoretical model and the use of stars or bold printing to indicate that the estimate is statistically significant. This list of covariates was pre-specified. We collect these articles' characteristics since they have been shown to be related to specification searching. For instance, Brodeur et al. (2016) provide evidence that there is less specification searching in top economics journals for articles with a theoretical model.

B Identification Strategy

The objective is to investigate the impact of the editorial statement on the number of test statistics that are statistically significant at conventional levels. To identify this effect, we use test statistics from the treated and control journals and rely on a differences-in-differences approach. The identification strategy (i.e., differences-in-differences) and the control variables were pre-specified. Our unit of observation is a test statistic.

In our main specification, we estimate:

$$Y_{iajt} = \alpha + \beta During/After_{iajt} + \delta Treated_{iajt} \\ + \gamma During/After_{iajt} \times Treated_{iajt} + X'_{iajt}\lambda + \varepsilon_{iajt}, \quad (1)$$

where Y_{iajt} is the outcome variable (significant at conventional levels) for test statistic i in article a, journal j and time t. $During/After_{iajt}$ is a dummy that equals one if test statistic i in article a was published "During" or "After" the editorial statement and zero otherwise. $Treated_{iajt}$ is a dummy that equals one if test statistic i in article a was published in one of the health economics journals and zero if it was published in one of the control journals. The interaction of $During/After_{iajt}$ and $Treated_{iajt}$ shows the effect of the editorial statement. The coefficient of interest here is thus γ. The interpretation relies on the identification condition that the distribution of test statistics in the treated and control journals would follow the same time trend in the absence of the treatment, i.e., editorial statement. We test this assumption explicitly in the next section.

X'_{iajt} is a vector containing regressors known to be predictors of specification searching. We include dummies for popular JEL codes (I10, I11, I13 and I18), the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing to indicate statistically significant estimates.

We rely throughout on OLS estimates and follow the recommendations of Bertrand et al. (2004) and compute standard errors clustered at the journal article-level. We also report marginal effects from logit regressions in Appendix Tables. The estimates are virtually the same as the OLS estimates.

C Descriptive Statistics

Table 1 provides descriptive statistics on the number of articles, tables and test statistics per journal and along articles' characteristics. The average number of tables per article is about 1.4 for both the health economics journals and the control journals. The average number of test statistics per article for health economics journals is about 54. The number of test statistics per article is lower for control journals with approximately 41 tests.

Approximately 24% of journal articles in our sample are single-authored and 19% have a theoretical model. About 89% of the articles report either stars or bold printing to indicate whether the estimate is statistically significant at conventional levels. The most popular JEL codes are I10 (Health General), I11 (Analysis of Health Care Markets), I13 (Health Insurance, Public and Private) and I18 (Health Government Policy, Regulation and Public Health).

Appendix Table A1 reports summary statistics on the number of test statistics for the categories "Before" (column 1), "During" (column 2) and "After" (column 3) the editorial statement. Most of the test statistics in our sample are in the category "After" the editorial statement. Respectively, 26% and 25% of tests are in the categories "Before" and "During" the editorial statement.

Appendix Table A1 also shows the number of tests for the three categories according to articles' characteristics. The percentage of papers using stars to report significance levels is quite similar for the three categories. The number of tests in papers with a theoretical model is somewhat similar across the time categories. Respectively, 17% and 19% of tests are in articles with a model for the categories "After" and "Before" the editorial statement. The share of single-authored papers has been decreasing over time. We will test explicitly in Section III whether these compositional changes affect our main results.

Figure 1, panel A, illustrates the distribution of z-statistics for the category "Before" in the health economics journals. We limit our study to the interval [0,10] to allow a direct comparison with the figures presented in Brodeur et al. (2016). We chose a bandwidth of 0.20 to create a total of 35 bins in each figure.[7] Approximately 61% of z-statistics are statistically significant at the 10% level and about 55% are statistically significant at the 5% level. The distribution of tests has a two-humped camel shape with "missing" p-values between 0.25 and 0.10. This is in line with Brodeur et al. (2016), who found that approximately 54% of test statistics in three prestigious economics journals were significant at the 5% level.

III Results

In this section, we first rely on a simple difference model to test whether the Editorial Statement of Negative Findings decreased the extent of publication bias in the health economics journals. We then use a differences-in-differences approach with two non-health economics journals as a comparison group. Last, we test the robustness of the results to changes in the characteristics of the papers published.

A Descriptive Analysis

Figures 1 plots the distribution of z-statistics for the health economics journals. Panel A restricts the sample to articles published before the editorial statement. Panel B restricts the sample to papers that were submitted before the statement on negative findings, but published after. In panel C, we include articles from the categories "During" or "After." Panel D restricts

[7] We rely on an Epachnikov kernel density to smooth jumps in the distributions. We use a bandwidth of 0.20. This explains why there is a sharp incline from [0,0.2].

the sample to papers submitted and published after the statement.

What emerges clearly is that the editorial statement on negative findings shifted the distribution of test statistics leftwards. About 44% of z-statistics are statistically significant at the 5% level after the editorial statement. In comparison, 55% of z-statistics were significant at the 5% level before the statement. A similar pattern emerges for the 10% and 1% thresholds.

We test explicitly whether the editorial statement on negative findings induced a change in the number of test statistics not rejecting the null hypothesis in Table 2. Columns 1, 2 and 3 report means and standard deviations for three subsamples. Columns 1, 2 and 3 restrict the sample to papers published before, during and after the editorial statement, respectively. Columns 4, 5 and 6 present t-test values for the equality of the means between columns 1 and 2, columns 1 and 3 and columns 2 and 3. We combine all the health economics journals for this exercise.

The estimates presented in Table 2 suggest that the editorial statement significantly reduced the number of tests that are significant at the 10%, 5% and 1% levels. The difference between columns 1 and 3 is statistically significant at the 1% level for the three significance levels.[8] The difference between "Before" and "During" is also statistically significant for the three significance thresholds. About 43% of z-statistics are respectively statistically significant at the 5% level for the category "During."

These results suggest that a part of the decrease in the percentage of test statistics significant at conventional levels is not due to a change in the behavior of authors. Authors who submitted a paper before the statement on negative findings, but who ended up publishing their paper after the

[8] As a robustness check, we check whether the leftward shift after the editorial statement is driven by a change in the characteristics of the papers published (not shown for space consideration). We restrict the sample to three subsamples: papers without a theoretical model, multiple-authored papers and papers with at least one of the four most popular JEL codes in our study. The difference between columns 1 and 3 ("Before" and "After") remains statistically significant at the 5% level for the three significance levels.

statement were not "affected" by the statement. Nonetheless, we observe an effect on the share of significant tests for this set of papers. This is suggestive evidence that editors' preferences for null results changed from "Before" to "During." We further examine the role played by authors in Section IV.

We also investigate whether the distribution of tests changed from "Before" to "After" for our control journals. This is an important robustness check since the distribution of tests in the health economics journals could change for reasons other than the editorial statement. On the one hand, Miguel et al. (2014) and Nosek et al. (2015) point out that researchers have now access (and are now constrained) to more transparent practices and tools such as pre-analysis plans and open data. On the other hand, there is more competition to publish in top journals, which could provide incentive to data mine (Card and DellaVigna (2013)). It thus remains unclear whether the distribution of tests in non-health economics journals changed over time.

Figure 2 plots the distribution of z-statistics for our control journal. We find that the distribution of tests shifted *rightwards* after the editorial statement. This pattern is confirmed in Table 2, panel B. The estimates show that from "Before" to "After," the share of tests that are significant at the 10%, 5% and 1% levels significantly *increased*. This result suggests that publication bias for our control journal increased over the period studied. Furthermore, the share of tests significant at the 10% and 5% level did not changed from "During" to "After."

The rightward shift of the distribution of tests for the control journals possibly reflects a change in the preferences for negative findings from the editors and referees and/or more specification searching from authors. The comparison with the control journals thus offers evidence that we may have underestimated the impact of the editorial statement in Table 2, panel A.

Note that we omit journal articles in the field of health economics in the control journal. Hence, the rightward shift of the distribution of tests is not related to a substitution of papers rejecting the null hypothesis from health economics journals to the control journals.

B Simple Difference

Before turning to our differences-in-differences approach, we formally test using a simple difference whether the proportion of statistically significant tests decreased in health economics journals since February 2015. Table 3 contains OLS estimates of equation (1) for the health economics journals over the "Before," "During" and "After" periods. (See Appendix Table A2 for the logit estimates.) The variable $During/After_{iajt}$ captures the simple difference between the periods "Before" and "During" and "After" combined. The dependent variables are dummies for whether the test statistic is significant at the 10% level (columns 1–3) and at the 5% level (columns 4–6), respectively. (See Appendix Table A3 for the other conventional significance level.) The sample size is 12,751 observations (i.e., test statistics). We report standard errors clustered by journal article in parentheses.

In columns 1 and 4, we only include the variable $During/After_{iajt}$. The identification assumption is that the proportion of reported significant findings would have been the same in the absence of the editorial statement. We relax this assumption by controlling for articles' characteristics and JEL codes in columns 2–3 and 5–6. Our estimates suggest that the editorial statement decreased by about 12 percentage points the number of test statistics that are statistically significant at the 10% and 5% level. The estimates are statistically significant at the 1% level in columns 1–3 and at the 2% level in columns 4–6.

Appendix Table A4 shows the estimates for our control variables. We find a positive association between the likelihood to reject the null hypothesis and the presence of a theoretical model and single-authored articles, but the estimates are not statistically significant at conventional levels. The use of stars or bold printing is negatively related to the likelihood to reject the null hypothesis in our sample, but the estimates are also statistically insignificant at conventional levels.

Overall, our findings provide suggestive evidence that the editorial statement increased the proportion of papers whose main empirical findings are negative (e.g., null hypotheses not rejected). We turn in the next subsection to our pre-specified differences-in-differences approach.

C Differences-in-Differences

Table 4 contains OLS estimates of equation (1) for both treated and control journals. (See Appendix Table A5 for the logit estimates.) The sample size is 17,653 observations. The dependent variable is a dummy for whether the test statistic is significant at the 5% level. In column 1, we only include the variables $During/After_{iajt}$ and $Treated_{iajt}$. The variable $Treated_{iajt}$ compares the significance of test statistics in the health economics journals in comparison to the control journal. The estimate is small, negative and statistically insignificant. The estimate for the variable $During/After_{iajt}$ is also small, negative and statistically insignificant. These results suggest that there were no differences between control and treated journals, nor from the period "Before" to "During-After."

In column 2, we add the interaction term necessary in our differences-in-differences model. Columns 3–4 augment the model progressively with more controls. Coefficients in columns 2–4 suggest that the editorial statement decreased the proportion of test statistics rejecting the null hypothesis by about 17 percentage points. The estimates are statistically significant at the 2% level in all columns. Adding articles' characteristics and JEL codes has no effect on the size and significance of the results.

Appendix Tables A6 and A7 replicate Table 4 for the other conventional significance levels. The estimates suggest that the editorial statement reduced the number of tests significant at the 1% and 10% levels by approximately 22 and 16 percentage points. The estimates are statistically significant at the 2% level and confirm the findings presented in Table 4.

To sum up, our findings provide suggestive evidence that the editorial statement significantly increased the number of test statistics not rejecting the null hypothesis. In contrast, the proportion of significant tests seem to have increased in control journals over the same time period. We check whether the effect of the editorial statement is intensifying over time in Figure 3. This figure provides a visual summary of the evolution in the percentage of tests significant at the 5% level for the treated and control journals. Appendix Figures A2 and A3 provide similar summaries for the other thresholds. Figure 3 plots the percentage of tests significant at the 5% level for the categories "Pre Before," "Before," "During," "After 2015 – 16" and "After 2017." The last two categories include papers submitted after the statement on negative finding and accepted in the years 2015–2018. In other words, we split the category "After" and compare multiple "After" time periods. The year of submission determines whether the journal article is in the category "After 2015 – 16" or "After 2017." Similarly, we split the category "Before" and compare the "Pre Before" and "Before" time periods

In the treated journals, approximately 43% (48%) of test statistics were significant at the 5% (10%) level for the category "During" in comparison to 44% (51%) for the category "After 2015 – 16" and 36% (43%) for the category "After 2017." On the other hand, the fraction of tests significant at the 5% level remained somewhat stable (with an upward trend) over the time periods "During," "After 2015–16" and "After 2017" for the control journals. These results provide suggestive evidence that the effect of the editorial statement is intensifying over time. Of note, though, our research focuses on a time period in which the editorial statement was relatively new. Future research should investigate further the long run effects of the editorial statement.

D Robustness Checks

In Table 5, we present estimates from a number of robustness checks.[9] The number of test statistics per journal varies quite a lot in our sample. For instance, 112 are from the *Journal of Health Economics* and 11 from the *Health Economics Review*. We thus check whether our findings are driven by one of the health economics journals. In column 1, we omit the *European Journal of Health Economics*. Column 2 omits *Health Economics*. In column 3, we omit *Health Economics Review*. In column 4, we omit the *International Journal of Health Economics and Management*. Column 5 omits the *Journal of Health Economics*. Our estimates for the interaction term are all large (range from -17 to -20 percentage points), negative and statistically significant at the 4% level.

Furthermore, we test whether the proportion of test statistics not rejecting the null hypothesis in treated journals was already increasing prior to the editorial statement. In other words, we worry about the time trend for the treated journals. As a robustness check, we split the category "Before" in two separate categories: "Before" and "Pre Before." The latter category contains research articles published exactly one year prior to the category "Before." Figure 3 and Appendix Figures A2 and A3 plot the percentage of tests significant at conventional levels for the different time categories. The percentage of tests significant at the 10%, 5% and 1% level remained somewhat stable from the category "Pre Before" to the category "Before"

[9] Note that our results are not driven by research articles from the coeditors and associate editors of the health economics journals themselves. Few research articles were published by an associate editor or a coeditor and excluding them has no effect on our conclusions. This means that the impact of the editorial statement is not driven by research articles written by the editors of these two journals.

for the treated and control journals.

We test explicitly whether the trends are parallel in Table 6. We restrict the sample to the time periods "Before" and "Pre Before." More specifically, we estimate:

$$Y_{iajt} = \alpha + \eta Before_{iajt} + \delta Treated_{iajt} + \theta Before_{iajt} \times Treated_{iajt} + X'_{iajt}\lambda + \varepsilon_{iajt}, \quad (2)$$

where Y_{iajt} is the outcome variable. $Before_{iajt}$ is a dummy that equals one if test statistic i in article a was published "Before" the editorial statement and zero if it was published prior to the category "Before." The interaction of $Before_{iajt}$ and $Treated_{iajt}$ shows the effect of the editorial statement. The same controls as in equation 1 are included. The coefficient of interest here is thus θ. Our estimates are very small, positive and statistically insignificant confirming that the trends are parallel.

Overall, we find evidence that the editorial statement reduced publication bias in the health economics journals. These results provide suggestive evidence that editors' preferences for null findings changed over time. We test in the next section whether the statement induced a change in the behavior of authors.

IV Mechanisms

A Authors' Behavior

We now check whether the statement has affected the distribution of tests in the health economics journals through a change in authors' behavior. We collect data on a large number of working papers in the field of health economics. This allows us to have a better sense of the potential pool of publications (i.e., not just the ones that were accepted) and to better disentangle whether the effect is arising from: (1) displacement of positive results into other journals and the displacement of null results into the treated journals; (2) changing standards by editors and reviewers; (3) changing research practices by authors.

Note that this exercise was not pre-registered. We did not write in the PAP that we would collect test statistics from working papers. This is an exploratory analysis and there is thus a need for greater caution in interpreting the results and drawing conclusions.

For this exercise, we select working papers using the IDEAS/RePEc NEP reports on Health Economics (https://ideas.repec.org/n/nep-hea/). Given the very large number of papers reported in the NEP reports on Health Economics (over 1,000 working papers per year), we randomly selected two NEP reports for the year 2014 (NEP-HEA-2014-06-07 and NEP-HEA-2014-11-01). We then selected two NEP reports per year for the time period 2015–2018. We selected the corresponding NEP reports, i.e., published during the same month/week (NEP-HEA-2015-06-05, NEP-HEA-2015-10-25, NEP-HEA-2016-06-09, NEP-HEA-2016-11-06, NEP-HEA-2017-06-11, NEP-HEA-2017-10-29, NEP-HEA-2018-06-11 and NEP-HEA-2018-10-29), to avoid seasonality issues.[10] This data collection provides 94 working papers. A number of working papers were excluded; we excluded papers not written in English and PhD thesis.

We cannot classify working papers into the categories "Before," "During" and "After" since we do not have information on the date of submission to economics journals. Instead, we plot the share of test statistics that are statistically significant at conventional levels per year in Figure 4. We find that the proportion of tests that are statistically significant at the 5% level decreased during this time period. Approximately 44% of z-statistics in the health economics working papers are statistically significant at the 5% level

[10]The number of working papers and the proportion of NBER and IZA working papers is quite similar across years/NEP reports.

in 2014 in comparison to 38% in 2018. Similarly, the proportion of tests rejecting the null hypothesis decreased for the other conventional levels over the years 2014–2018 with the largest decline for the 10% significance level (see Appendix Figures A4 and A5).

Overall, we find that health economists are increasingly disseminating new working papers not rejecting the null hypothesis. This set of results provides suggestive evidence that the behavior of health economists did change over time and that positive results are not simply displaced into other journals. Note that we do not know whether non-health economists also changed their behavior over this time period. The lack of a control group for this exercise prevents us from concluding that the editorial statement changed health economists' research practices.

B Subgroup Analysis

We now test whether the statement has a larger effect on papers with specific characteristics. More precisely, we check whether the leftward shift in the distribution of tests after the editorial statement is related to the number of authors and the presence of a theoretical model. Table 7 provides estimates of equation (1). Column 1 restricts the sample to journal articles without a theoretical model. Column 2 keeps only articles with a theoretical model. The estimates are negative in both columns, but larger in column 1 than in column 2. In other words, the statement seems to have a larger effect on journal articles without a theoretical model.

Columns 3 and 4 restrict the sample to multi-authored research articles and single-authored research articles, respectively. The estimates suggest that the decrease in the proportion of tests rejecting the null hypothesis during and after the statement is driven entirely by multi-authored research articles. The estimate for single-authored research articles is positive and statistically insignificant.

To sum up, the editorial statement shifted leftward the distribution of tests and this shift is mostly driven by multi-authored research articles and journal articles without a theoretical model. These results are interesting since there is a positive association between single-authored research articles (and the presence of a theoretical model) and the likelihood to reject the null hypothesis (Appendix Table A4). These findings provide suggestive evidence that the statement has affected mostly researchers who have characteristics that are negatively related to the likelihood to reject the null hypothesis.

C External Validity

Finally, we provide suggestive evidence that our results can be generalized to non-health economics journals in Appendix Figure A6. We rely on data from Brodeur et al. (2016) and check whether the distribution of test statistics in top economics journals (2005–2011) is similar for health economics and non-health economics published articles. Appendix Figure A6 plots the distributions of z-statistics for journal articles reporting at least one I1 JEL code and for journal articles not reporting I1 JEL codes.

The distributions of tests for health and non-health economics articles have a similar shape. For example, about 60% of test statistics are statistically significant at the 10% level in both subsamples. This finding suggests that our results seem to have external validity and that the editorial statement could have a similar effect if implemented by non-health economics journals.

V Conclusion

In this paper, we documented how a new transparent practice may reduce the extent of selective reporting. We test whether the Editorial Statement on Negative Findings which was sent out by the editors of eight health economics journals decreased the extent of publication bias. Our estimates suggest that the editorial statement significantly decreased the number of test statistics that are statistically significant at conventional levels in five health economics journals. We conclude that the editorial statement reduced the extent of publication bias. Interestingly, we found that the impact of the statement seemed to "intensify" over the period studied, since the decrease in the share of significant tests was larger for papers recently submitted than for papers submitted before the statement, but accepted after.

Analyzing the mechanisms, we provide evidence that a part of the decrease in the share of significant tests is due to a change in editors' preferences for negative findings. We also show that health economists are increasingly disseminating new working papers not rejecting the null hypothesis. This result provides suggestive evidence that the editorial statement induced authors to change their behavior. But while our results are suggestive and indicative of a change in behavior, they are not definitive. More research is needed to test the robustness of our findings and to better disentangle the mechanisms through which the editorial statement decreased the extent of publication bias.

Our results have interesting implications for editors and the academic community (Nosek et al. (2015)). They suggest that incentives may be aligned to promote a more transparent research. The decrease in publication bias should be of interest to policy-makers who use empirical findings to inform policies. These findings should also be of interest to editors of other journals who could make similar statements and perhaps take even stronger measures. Further research could then ensure that what we have found here is not merely some idiosyncratic trend or pattern of editorial, review or submission practices at these specific journals and that our findings are truly representative for economic research and journals, in general. Further work could also test the effectiveness of other transparent practices using the methods described in this research.

References

Abadie, A.: 2018, Statistical Non-Significance in Empirical Economics. National Bureau of Economic Research.

Andrews, I. and Kasy, M.: 2017, Identification of and Correction for Publication Bias. National Bureau of Economic Research.

Ashenfelter, O. and Greenstone, M.: 2004, Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias, *American Economic Review* **94**(2), 454–460.

Ashenfelter, O., Harmon, C. and Oosterbeek, H.: 1999, A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias, *Labour Economics* **6**(4), 453–470.

Bertrand, M., Duflo, E. and Mullainathan, S.: 2004, How Much Should we Trust Differences-in-Differences Estimates?, *Quarterly Journal of Economics* **119**(1), 249–275.

Blanco-Perez, C. and Brodeur, A.: 2017, Publication Bias and Editorial Statement on Negative Findings. Pre-Analysis Plan: https://osf.io/mjbj2/.

Blanco-Perez, C. and Brodeur, A.: 2019, Transparency in Empirical Economic Research, *IZA World of Labor* .

Brodeur, A., Cook, N. and Heyes, A.: 2018, Methods Matter: P-Hacking and Causal Inference in Economics. IZA Discussion Papers 11796.

Brodeur, A., Lé, M., Sangnier, M. and Zylberberg, Y.: 2016, Star Wars: The Empirics Strike Back, *American Economic Journal: Applied Economics* **8**(1), 1–32.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T. et al.: 2016, Evaluating Replicability of Laboratory Experiments in Economics, *Science* **351**(6280), 1433–1436.

Card, D. and DellaVigna, S.: 2013, Nine Facts about Top Journals in Economics, *Journal of Economic Literature* **51**(1), 144–61.

Card, D. and Krueger, A. B.: 1995, Time-Series Minimum-Wage Studies: A Meta-analysis, *American Economic Review* **85**(2), 238–243.

Casey, K., Glennerster, R. and Miguel, E.: 2012, Reshaping Institutions: Evidence on Aid Impacts Using a Pre-Analysis Plan, *Quarterly Journal of Economics* **127**(4).

Chang, A. C. and Li, P.: 2015, Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not". Board of Governors of the Federal Reserve System (US).

Christensen, G. S. and Miguel, E.: 2018, Transparency, Reproducibility, and the Credibility of Economics Research, *Journal of Economic Literature* **56**(3), 920–80.

De Long, J. B. and Lang, K.: 1992, Are all Economic Hypotheses False?, *Journal of Political Economy* **100**, 1257–1257.

Dewald, W. G., Thursby, J. G. and Anderson, R. G.: 1986, Replication in Empirical Economics: The Journal of Money, Credit and Banking Project, *American Economic Review* **76**(4), 587–603.

Doucouliagos, C. and Stanley, T. D.: 2013, Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity, *Journal of Economic Surveys* **27**(2), 316–339.

Dufwenberg, M. and Martinsson, P.: 2014, Keeping Researchers Honest: The Case for Sealed-Envelope-Submissions. IGIER (Innocenzo Gasparini Institute for Economic Research), number 533.

Fanelli, D.: 2009, How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data, *PLoS ONE* **4**(5).

Fanelli, D.: 2010, "Positive" Results Increase Down the Hierarchy of the Sciences, *PLoS ONE* **5**(4).

Franco, A., Malhotra, N. and Simonovits, G.: 2014, Publication Bias in the Social Sciences: Unlocking the File Drawer, *Science* **345**(6203), 1502–1505.

Furukawa, C.: 2017, Unbiased Publication Bias. MIT Mimeo.

Gerber, A. and Malhotra, N.: 2008a, Do Statistical Reporting Standards Affect what is Published? Publication Bias in two Leading Political Science Journals, *Quarterly Journal of Political Science* **3**(3), 313–326.

Gerber, A. S. and Malhotra, N.: 2008b, Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results?, *Sociological Methods & Research* **37**(1), 3–30.

Gerber, A. S., Malhotra, N., Dowling, C. M. and Doherty, D.: 2010, Publication Bias in Two Political Behavior Literatures, *American Politics Research* **38**(4), 591–613.

Havránek, T.: 2015, Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting, *Journal of the European Economic Association* **13**(6), 1180–1204.

Havránek, T. and Sokolova, A.: 2016, Do Consumers Really Follow a Rule of Thumb? Three Thousand Estimates from 130 Studies Say "Probably Not". Czech National Bank, Research Department, number 2016/08.

Henry, E.: 2009, Strategic Disclosure of Research Results: The Cost of Proving your Honesty, *Economic Journal* **119**(539), 1036–1064.

Ioannidis, J. P.: 2005, Why Most Published Research Findings Are False, *PLoS medecine* **2**(8), e124.

Ioannidis, J., Stanley, T. D. and Doucouliagos, H.: 2017, The Power of Bias in Economics Research, *Economic Journal* **127**(605).

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C. et al.: 2016, Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency, *PLoS Biology* **14**(5), e1002456.

Leamer, E. E.: 1983, Let's Take the Con Out of Econometrics, *American Economic Review* **73**(1), pp. 31–43.

Maniadis, Z., Tufano, F. and List, J. A.: 2017, To Replicate or not to Replicate? Exploring Reproducibility in Economics Through the Lens of a Model and a Pilot Study, *Economic Journal* **127**(605).

McCloskey, D. N.: 1985, The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests, *American Economic Review: Papers and Proceedings* **75**(2), 201–205.

McCullough, B., McGeary, K. A. and Harrison, T. D.: 2008, Do Economics Journal Archives Promote Replicable Research?, *Canadian Journal of Economics* **41**(4), 1406–1420.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D., Humphreys, M., Imbens, G. et al.: 2014, Promoting Transparency in Social Science Research, *Science* **343**(6166), 30–31.

Neumark, D.: 2001, Evidence on Employment Effects of Recent Minimum Wage Increases from a Pre-Specified Research Design, *Industrial Relations* **40**(1), 121–144.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., Buck, S., Chambers, C. D., Chin, G., Christensen, G. et al.: 2015, Promoting an Open Research Culture, *Science* **348**(6242), 1422–1425.

Olken, B. A.: 2015, Promises and Perils of Pre-Analysis Plans, *Journal of Economic Perspectives* **29**(3), 61–80.

Ridley, J., Kolm, N., Freckelton, R. P. and Gage, M. J. G.: 2007, An Unexpected Influence of Widely Used Significance Thresholds on the Distribution of Reported P-Values, *Journal of Evolutionary Biology* **20**(3), 1082–1089.

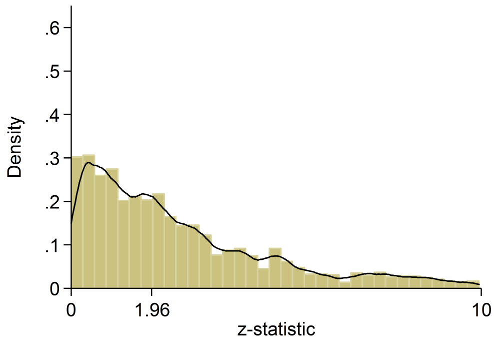
Simmons, J. P., Nelson, L. D. and Simonsohn, U.: 2011, False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant, *Psychological Science* **22**, 1359–1366.

Stanley, T. D.: 2005, Beyond Publication Bias, *Journal of Economic Surveys* **19**(3), 309–345.

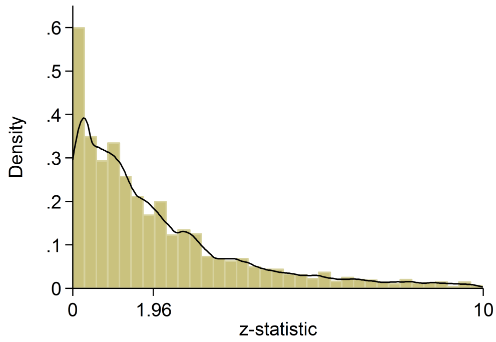
Stanley, T. D.: 2008, Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection, *Oxford Bulletin of Economics and Statistics* **70**(1), 103–127.

Vivalt, E.: 2017, The Trajectory of Specification Searching and Publication Bias Across Methods and Disciplines. mimeo: Australian National University.

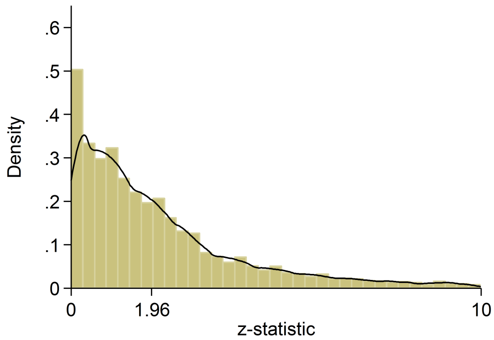
Figure 1: Distributions of z-statistics before, during and after the editorial.



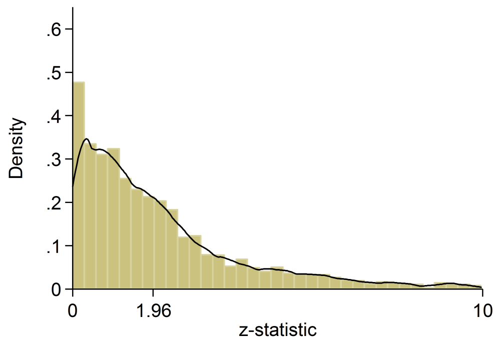
(a) Raw distribution of z-statistics. Before the editorial.



(b) Raw distribution of z-statistics. During the editorial.



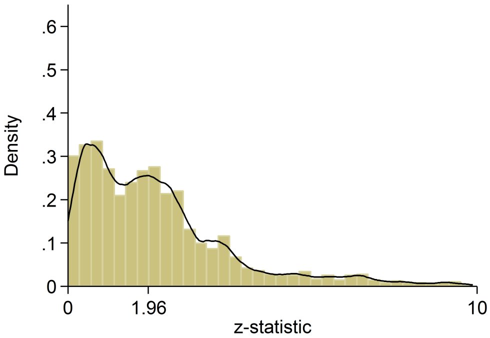
(c) Raw distribution of z-statistics. During and after the editorial.



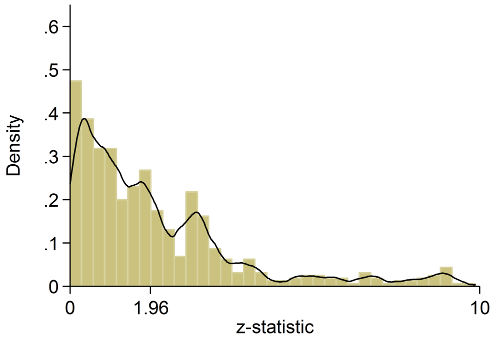
(d) Raw distribution of z-statistics. After the editorial.

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. Before the editorial category includes papers that were submitted and published before the statement on negative findings. During the editorial category includes papers that were submitted before the statement on negative findings, but published after. After the editorial category includes papers submitted and published after the statement on negative findings. Lines correspond to kernel density estimates.

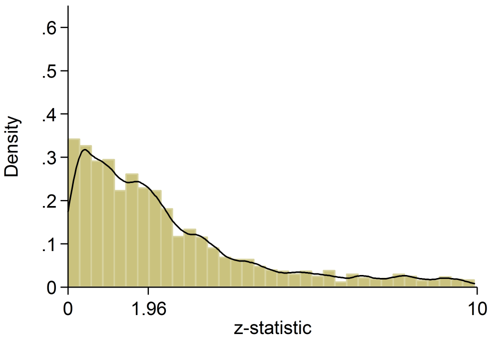
Figure 2: Distributions of z-statistics for the control journals.



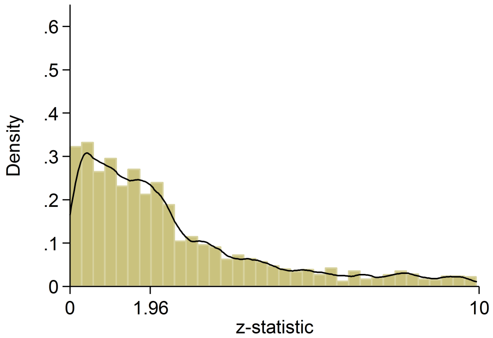
(a) Raw distribution of z-statistics. Before the editorial.



(b) Raw distribution of z-statistics. During the editorial.



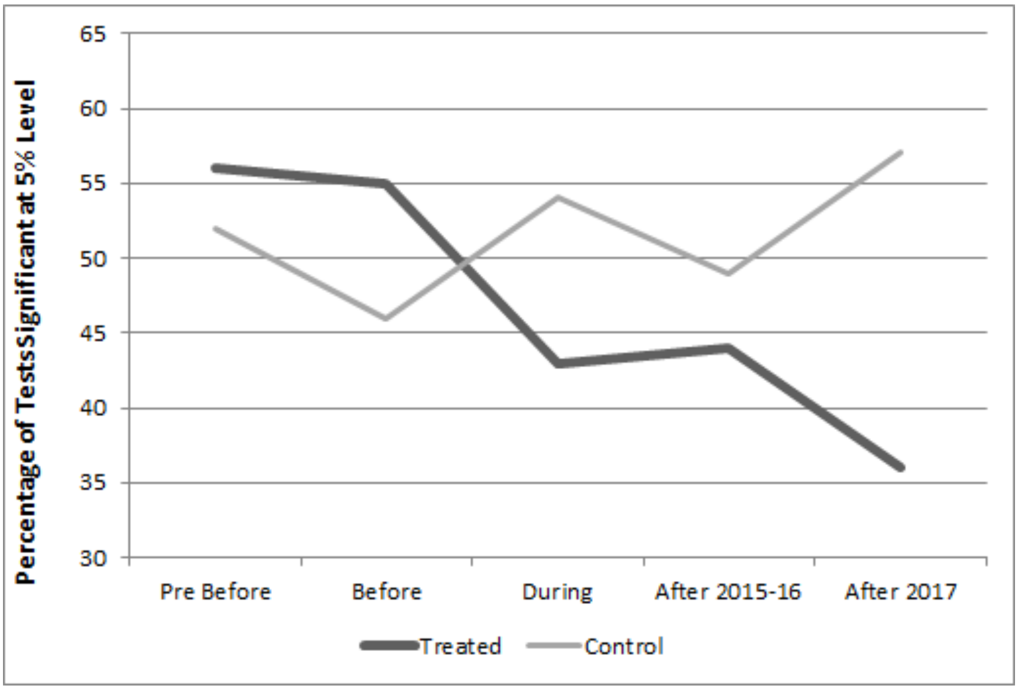
(c) Raw distribution of z-statistics. During and after the editorial.



(d) Raw distribution of z-statistics. After the editorial.

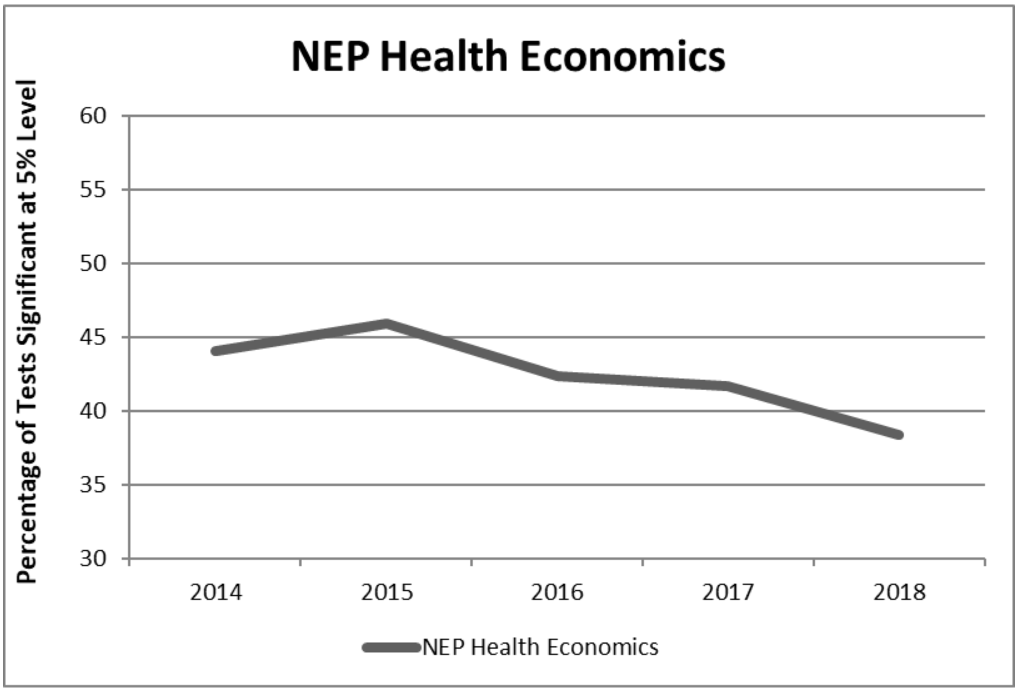
Sources: *Labour Economics* and the *Journal of Public Economics*. Before the editorial category includes papers that were submitted and published before the statement on negative findings. During the editorial category includes papers that were submitted before the statement on negative findings, but published after. After the editorial category includes papers submitted and published after the statement on negative findings. Lines correspond to kernel density estimates.

Figure 3: Percentage of tests significant at the 5% level.



Sources: treated journals include the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. Control journals include *Labour Economics* and the *Journal of Public Economics*. Percentage of tests significant at the 5% level by categories. Pre Before the editorial category includes papers that were published one year before the category Before. Before the editorial category includes papers that were submitted and published before the statement on negative findings. During the editorial category includes papers that were submitted before the statement on negative findings, but published after. After the editorial categories include papers submitted and published (respectively in 2015–16 and 2017) after the statement on negative findings.

Figure 4: Working papers: Percentage of tests significant at the 5% level.



Sources: NEP-HEA-2014-06-07, NEP-HEA-2014-11-01, NEP-HEA-2015-06-05, NEP-HEA-2015-10-25, NEP-HEA-2016-06-09, NEP-HEA-2016-11-06, NEP-HEA-2017-06-11, NEP-HEA-2017-10-29, NEP-HEA-2018-06-11 and NEP-HEA-2018-10-29. Percentage of tests significant at the 5% level by year.

Table 1: Descriptive Statistics

	Proportion of ...		
	Articles	Tables	Tests
European Journal of Health Economics	7	5	2
Health Economics	20	21	35
Health Economics Review	3	3	2
Intl. J. Health Economics and Management	4	4	2
Journal of Health Economics	32	32	32
Journal of Public Economics	12	16	19
Labour Economics	22	20	8
Using Stars	89	93	93
With Model	19	17	19
Single-Authored	24	26	26
JEL I10	10	10	11
JEL I11	8	7	9
JEL I13	8	9	7
JEL I18	18	16	17

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table reports the proportion of articles, tables and test statistics for each variable. *Using Stars* corresponds to articles or tables using stars or bold printing to highlight statistical significance. Note that many papers do not report JEL codes.

Table 2: Percentage of Test Statistics Statistically Significant at Conventional Levels

	Treated Journals					
	Before	During	After	Diff (1)-(2)	Diff (1)-(3)	Diff (2)-(3)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Significant at 1%	0.453	0.341	0.328	-0.111	-0.132	-0.013
	(0.498)	(0.474)	(0.469)	(0.012)	(0.013)	(0.010)
Significant at 5%	0.555	0.432	0.439	-0.122	-0.116	0.007
	(0.497)	(0.495)	(0.596)	(0.012)	(0.013)	(0.010)
Significant at 10%	0.615	0.481	0.504	-0.132	-0.109	0.023
	(0.487)	(0.500)	(0.500)	(0.012)	(0.013)	(0.010)

	Control Journals					
	Before	During	After	Diff (1)-(2)	Diff (1)-(3)	Diff (2)-(3)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B						
Significant at 1%	0.322	0.470	0.418	0.148	0.106	-0.053
	(0.467)	(0.499)	(0.493)	(0.021)	(0.016)	(0.021)
Significant at 5%	0.471	0.538	0.530	0.067	0.069	-0.008
	(0.499)	(0.499)	(0.499)	(0.022)	(0.017)	(0.021)
Significant at 10%	0.549	0.597	0.596	0.047	0.053	-0.001
	(0.498)	(0.491)	(0.491)	(0.022)	(0.017)	(0.020)

Sources: treated journals include the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. Control journals include *Labour Economics* and the *Journal of Public Economics*. This table reports the percentage of test statistics statistically significant at conventional levels for three categories of articles: 1) Before the editorial category includes papers that were submitted and published before the statement on negative findings. 2) During the editorial category includes papers that were submitted before the statement on negative findings, but published after. 3) After the editorial category includes papers submitted and published after the statement on negative findings. Standard deviations are in parentheses (standard errors for the last two columns).

Table 3: Editorial Statement and Test Statistics: Simple Difference

	Significant at the 10% Level			Significant at the 5% Level		
	(1)	(2)	(3)	(4)	(5)	(6)
During/After	-0.119	-0.113	-0.120	-0.119	-0.113	-0.120
	(0.046)	(0.044)	(0.046)	(0.050)	(0.048)	(0.050)
Constant	0.615	0.579	0.589	0.555	0.530	0.541
	(0.037)	(0.094)	(0.095)	(0.041)	(0.089)	(0.094)
Article Characteristics		✓	✓		✓	✓
JEL Codes			✓			✓
R-squared	0.010	0.011	0.013	0.010	0.012	0.014
Observations	12,751	12,751	12,751	12,751	12,751	12,751

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. This table shows OLS estimates of equation (1). In columns 1–3, the dependent variable is a dummy for whether the test statistic is significant at the 10% level. In columns 4–6, the dependent variable is a dummy for whether the test statistic is significant at the 5% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 4: Editorial Statement and Test Statistics: Main Estimates

	Significant at the 5% Level			
	(1)	(2)	(3)	(4)
During/After	-0.059	0.061	0.052	0.052
	(0.039)	(0.055)	(0.052)	(0.052)
Treated Journals	-0.040	0.084	0.080	0.085
	(0.039)	(0.056)	(0.053)	(0.060)
During/After *		-0.180	-0.164	-0.174
Treated Journals		(0.075)	(0.071)	(0.072)
Constant	0.548	0.471	0.494	0.499
	(0.038)	(0.038)	(0.078)	(0.077)
Article Characteristics			✓	✓
JEL Codes				✓
R-squared	0.004	0.010	0.012	0.014
Observations	17,653	17,653	17,653	17,653

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows OLS estimates of equation (1). The dependent variable is a dummy for whether the test statistic is significant at the 5% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 5: Editorial Statement and Test Statistics: Robustness Checks

	Significant at the 5% Level				
	Omit EJHE (1)	Omit HE (2)	Omit HER (3)	Omit IJHEM (4)	Omit JHE (5)
During/After	0.053	0.049	0.052	0.052	0.047
	(0.052)	(0.051)	(0.052)	(0.052)	(0.052)
Treated Journals	0.075	0.088	0.084	0.087	0.118
	(0.061)	(0.058)	(0.062)	(0.061)	(0.088)
During/After * Treated Journals	-0.167	-0.185	-0.171	-0.178	-0.200
	(0.073)	(0.072)	(0.074)	(0.074)	(0.099)
Constant	0.485	0.487	0.496	0.499	0.567
	(0.079)	(0.084)	(0.078)	(0.079)	(0.093)
Article Characteristics	✓	✓	✓	✓	✓
JEL Codes	✓	✓	✓	✓	✓
R-squared	0.013	0.020	0.013	0.013	0.019
Observations	17,328	11,460	17,358	17,328	12,040

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows OLS estimates of equation (1). The dependent variable is a dummy for whether the test statistic is significant at the 5% level. Robust standard errors are in parentheses, adjusted for clustering by article. In column 1, we omit the *European Journal of Health Economics*. Column 2 omits *Health Economics*. In column 3, we omit *Health Economics Review*. In column 4, we omit the *International Journal of Health Economics and Management*. Column 5 omits the *Journal of Health Economics*. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 6: Editorial Statement and Test Statistics: Before the Editorial Statement

	Significant 1% (1)	Significant 5% (2)	Significant 10% (3)
Before	-0.031	-0.022	-0.010
	(0.093)	(0.186)	(0.080)
Treated Journals	0.070	0.036	0.020
	(0.119)	(0.107)	(0.097)
Before * Treated Journals	0.045	0.009	0.008
	(0.147)	(0.132)	(0.121)
Constant	0.393	0.476	0.553
	(0.126)	(0.121)	(0.114)
Article Characteristics	✓	✓	✓
JEL Codes	✓	✓	✓
R-squared	0.049	0.036	0.030
Observations	4,580	4,580	4,580

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows OLS estimates of equation (2). The sample is restricted to the time period "Before" and prior to the category "Before." In column 1, the dependent variable is a dummy for whether the test statistic is significant at the 1% level. In column 2, the dependent variable is a dummy for whether the test statistic is significant at the 5% level. In column 3, the dependent variable is a dummy for whether the test statistic is significant at the 10% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 7: Editorial Statement and Test Statistics: Subgroups

	Significant at the 5% Level			
	No Theoretical Model	Theoretical Model	Not Single-Authored	Single-Authored
	(1)	(2)	(3)	(4)
During/After	0.090	-0.101	0.105	-0.156
	(0.058)	(0.097)	(0.066)	(0.085)
Treated Journals	0.111	-0.038	0.105	-0.060
	(0.068)	(0.106)	(0.062)	(0.139)
During/After *	-0.190	-0.127	-0.221	0.089
Treated Journals	(0.082)	(0.127)	(0.082)	(0.146)
Constant	0.385	0.877	0.463	0.823
	(0.094)	(0.116)	(0.086)	(0.135)
Article Characteristics	✓	✓	✓	✓
JEL Codes	✓	✓	✓	✓
R-squared	0.008	0.091	0.017	0.058
Observations	14,360	3,293	13,011	4,642

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows OLS estimates of equation (1). The dependent variable is a dummy for whether the test statistic is significant at the 5% level. Robust standard errors are in parentheses, adjusted for clustering by article. Column 1 restricts the sample to journal articles with no theoretical model. Column 2 keeps only articles with a theoretical model. Columns 3 and 4 restrict the sample respectively to single-authored research articles and to multi-authored research articles, respectively. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

VI Appendix

Editorial Statement on Negative Findings

The Editors of the health economics journals named below believe that well-designed, well-executed empirical studies that address interesting and important problems in health economics, utilize appropriate data in a sound and creative manner, and deploy innovative conceptual and methodological approaches compatible with each journal's distinctive emphasis and scope have potential scientific and publication merit regardless of whether such studies' empirical findings do or do not reject null hypotheses that may be specified. As such, the Editors wish to articulate clearly that the submission to our journals of studies that meet these standards is encouraged.

We believe that publication of such studies provides properly balanced perspectives on the empirical issues at hand. Moreover, we believe that this should reduce the incentives to engage in two forms of behavior that we feel ought to be discouraged in the spirit of scientific advancement:

1. Authors withholding from submission such studies that are otherwise meritorious but whose main empirical findings are highly likely "negative" (e.g. null hypotheses not rejected).

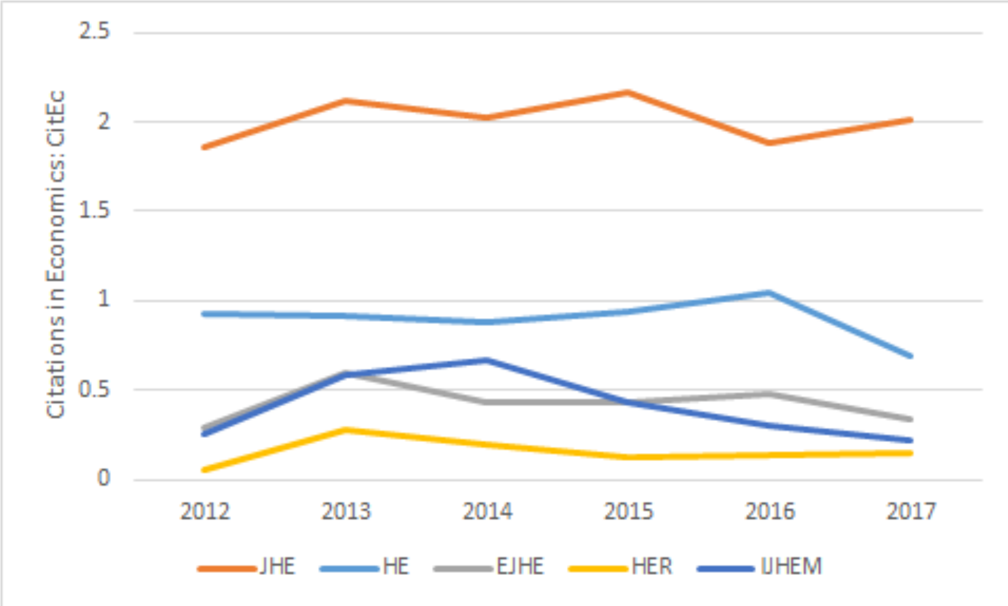
2. Authors engaging in "data mining," "specification searching," and other such empirical strategies with the goal of producing results that are ostensibly "positive" (e.g. null hypotheses reported as rejected).

Henceforth we will remind our referees of this editorial philosophy at the time they are invited to review papers. As always, the ultimate responsibility for acceptance or rejection of a submission rests with each journal's Editors.

American Journal of Health Economics
European Journal of Health Economics
Forum for Health Economics & Policy
Health Economics Policy and Law
Health Economics Review
Health Economics
International Journal of Health Economics and Management
Journal of Health Economics

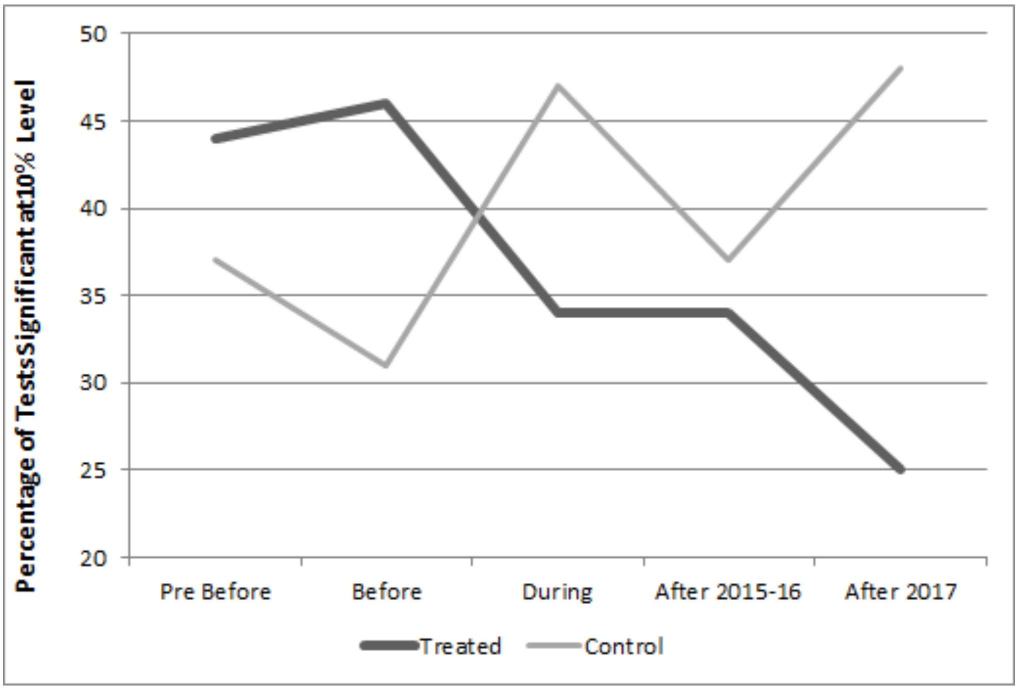
ONLINE APPENDIX: NOT FOR PUBLICATION

Figure A1: Impact factor of five health economics journals.



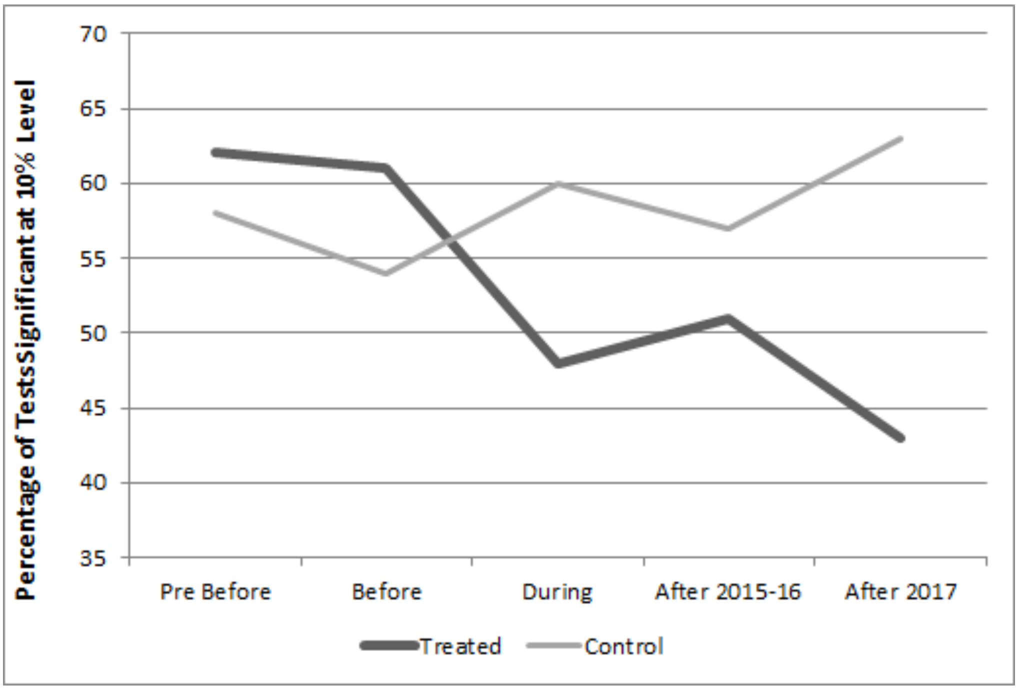
Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. CitEc is a RePEc service, providing citation data for economics journals and working papers. Data available here: http://citec.repec.org/search.html.

Figure A2: Percentage of tests significant at the 1% level.



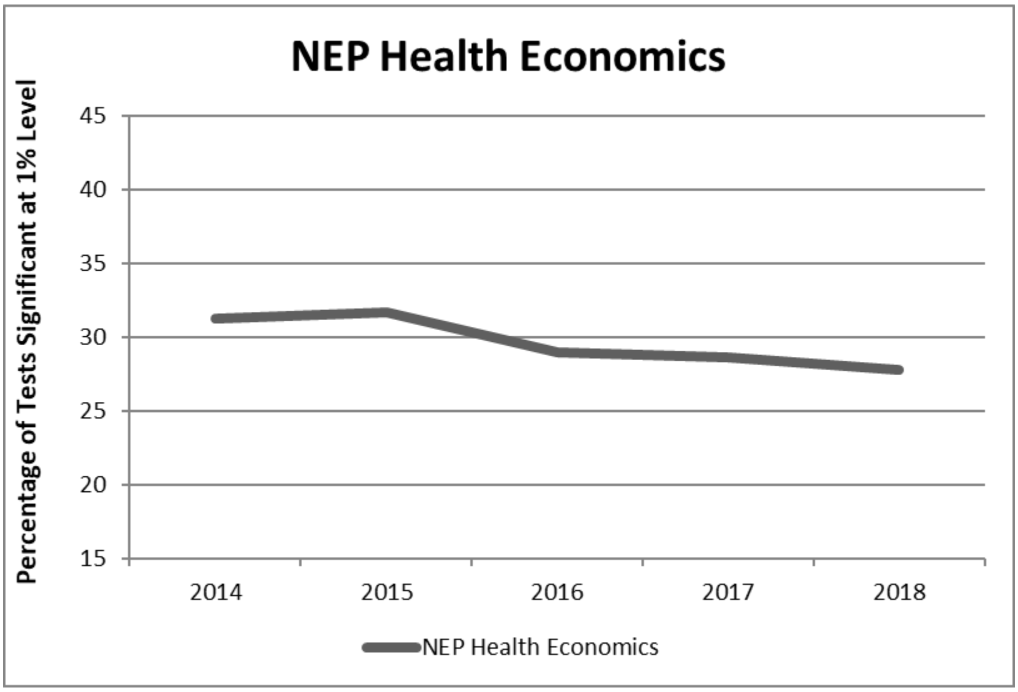
Sources: *Health Economics*, *Journal of Health Economics* and control journal. Percentage of tests significant at the 1% level by categories. Pre Before the editorial category includes papers that were published one year before the category Before. Before the editorial category includes papers that were submitted and published before the statement on negative findings. During the editorial category includes papers that were submitted before the statement on negative findings, but published after. After the editorial categories include papers submitted and published (respectively in 2015–16 and 2017) after the statement on negative findings.

Figure A3: Percentage of tests significant at the 10% level.



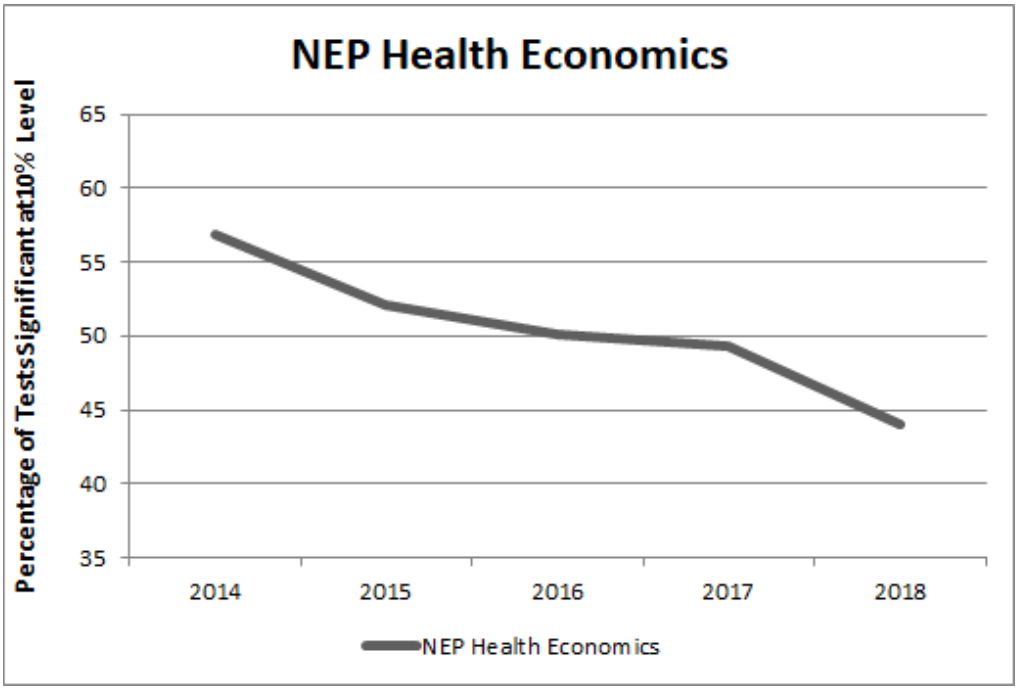
Sources: *Health Economics*, *Journal of Health Economics* and control journal. Percentage of tests significant at the 10% level by categories. Pre Before the editorial category includes papers that were published one year before the category Before. Before the editorial category includes papers that were submitted and published before the statement on negative findings. During the editorial category includes papers that were submitted before the statement on negative findings, but published after. After the editorial categories include papers submitted and published (respectively in 2015–16 and 2017) after the statement on negative findings.

Figure A4: Working papers: Percentage of tests significant at the 1% level.



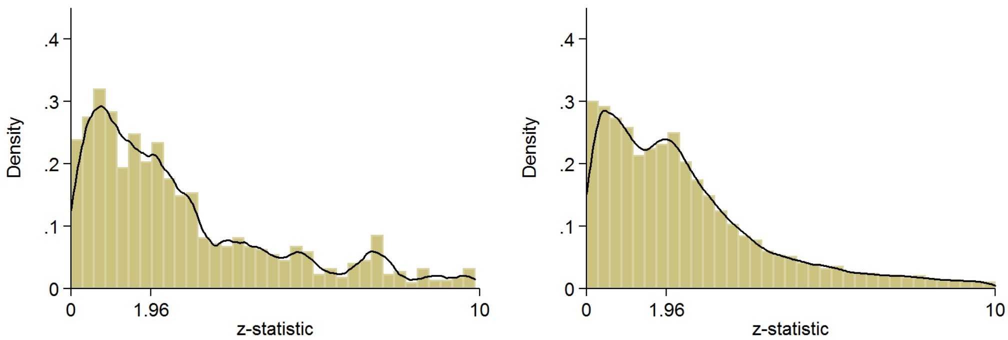
Sources: NEP-HEA-2014-06-07, NEP-HEA-2014-11-01, NEP-HEA-2015-06-05, NEP-HEA-2015-10-25, NEP-HEA-2016-06-09, NEP-HEA-2016-11-06, NEP-HEA-2017-06-11, NEP-HEA-2017-10-29, NEP-HEA-2018-06-11 and NEP-HEA-2018-10-29. Percentage of tests significant at the 1% level by year.

Figure A5: Working papers: Percentage of tests significant at the 10% level.



Sources: NEP-HEA-2014-06-07, NEP-HEA-2014-11-01, NEP-HEA-2015-06-05, NEP-HEA-2015-10-25, NEP-HEA-2016-06-09, NEP-HEA-2016-11-06, NEP-HEA-2017-06-11, NEP-HEA-2017-10-29, NEP-HEA-2018-06-11 and NEP-HEA-2018-10-29. Percentage of tests significant at the 10% level by year.

Figure A6: Distributions of z-statistics for health and non-health articles.



(a) Raw distribution of z-statistics. Health (I1 JEL codes).

(b) Raw distribution of z-statistics. All Journal Articles (exclude I1 JEL codes).

Sources: Data from Brodeur et al. (2016). AER, JPE, and QJE (2005–2011). Lines correspond to kernel density estimates.

Table A1: Descriptive Statistics: Before, During and After the Editorial

	Proportion of Tests ...		
	Before	During	After
	(1)	(2)	(3)
European Journal of Health Economics	3	2	2
Health Economics	21	46	36
Health Economics Review	2	1	2
Intl. J. Health Economics and Management	3	1	2
Journal of Health Economics	33	33	30
Labour Economics	8	2	12
Journal of Public Economics	29	15	16
Using Stars	94	95	92
With Model	19	21	17
Single-Authored	33	23	24
JEL I10	20	9	6
JEL I11	9	17	4
JEL I13	3	8	9
JEL I18	8	25	18

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table reports the proportion of test statistics for three categories of articles: 1) Before the editorial category includes papers that were submitted and published before the statement on negative findings. 2) During the editorial category includes papers that were submitted before the statement on negative findings, but published after. 3) After the editorial category includes papers submitted and published after the statement on negative findings. *Using stars* corresponds to articles or tables using stars or bold printing to highlight statistical significance. Note that many papers do not report JEL codes.

Table A2: Editorial Statement and Test Statistics: Logit Estimates for the Simple Difference

	Significant at the 10% Level			Significant at the 5% Level		
	(1)	(2)	(3)	(4)	(5)	(6)
During/After	-0.120	-0.114	-0.121	-0.118	-0.110	-0.119
	(0.046)	(0.045)	(0.047)	(0.049)	(0.047)	(0.049)
Article Characteristics		✓	✓		✓	✓
JEL Codes			✓			✓
Pseudo R-squared	0.007	0.008	0.010	0.007	0.008	0.010
Observations	12,751	12,751	12,751	12,751	12,751	12,751

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. This table shows logit estimates of equation (1). We report marginal effects. In columns 1–3, the dependent variable is a dummy for whether the test statistic is significant at the 10% level. In columns 4–6, the dependent variable is a dummy for whether the test statistic is significant at the 5% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table A3: Editorial Statement and Test Statistics: Simple Difference

	Significant at the 1% Level		
	(1)	(2)	(3)
During/After	-0.120	-0.112	-0.126
	(0.053)	(0.050)	(0.051)
Constant	0.453	0.460	0.487
	(0.044)	(0.083)	(0.090)
Article Characteristics		✓	✓
JEL Codes			✓
R-squared	0.011	0.013	0.017
Observations	12,751	12,751	12,751

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. This table shows OLS estimates of equation (1). The dependent variable is a dummy for whether the test statistic is significant at the 1% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table A4: Editorial Statement and Test Statistics: Covariates

	Significant 1% (1)	Significant 5% (2)	Significant 10% (3)
Using Stars	-0.098	-0.064	-0.053
	(0.077)	(0.072)	(0.069)
With Model	0.072	0.060	0.050
	(0.044)	(0.044)	(0.043)
Single-Authored	0.023	0.034	0.035
	(0.044)	(0.043)	(0.041)
JEL I10	-0.015	0.001	-0.002
	(0.065)	(0.061)	(0.058)
JEL I11	-0.003	-0.014	-0.006
	(0.053)	(0.055)	(0.054)
JEL I13	-0.063	-0.059	-0.060
	(0.062)	(0.064)	(0.063)
JEL I18	0.041	0.029	0.019
	(0.052)	(0.054)	(0.512)
R-squared	0.009	0.005	0.004
Observations	17,653	17,653	17,653

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management*, the *Journal of Health Economics*, the *Journal of Public Economics* and *Labour Economics*. This table shows OLS estimates of equation (1). In column 1, the dependent variable is a dummy for whether the test statistic is significant at the 1% level. In column 2, the dependent variable is a dummy for whether the test statistic is significant at the 5% level. In column 3, the dependent variable is a dummy for whether the test statistic is significant at the 10% level. Robust standard errors are in parentheses.

Table A5: Editorial Statement and Test Statistics: Logit Estimates

	Significant at the 5% Level			
	(1)	(2)	(3)	(4)
During/After	-0.058	0.060	0.051	0.051
	(0.039)	(0.055)	(0.052)	(0.051)
Treated Journals	-0.040	0.083	0.079	0.084
	(0.039)	(0.056)	(0.053)	(0.060)
During/After *		-0.178	-0.163	-0.173
Treated Journals		(0.073)	(0.070)	(0.071)
Article Characteristics			✓	✓
JEL Codes				✓
R-squared	0.003	0.007	0.009	0.010
Observations	17,653	17,653	17,653	17,653

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows logit estimates of equation (1). Marginal effects are reported. The dependent variable is a dummy for whether the test statistic is significant at the 5% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table A6: Editorial Statement and Test Statistics: Significant at 1%

	Significant at the 1% Level			
	(1)	(2)	(3)	(4)
During/After	-0.044	0.108	0.095	0.095
	(0.042)	(0.059)	(0.055)	(0.055)
Treated Journals	-0.027	0.131	0.128	0.139
	(0.042)	(0.060)	(0.056)	(0.062)
During/After *		-0.229	-0.209	-0.225
Treated Journals		(0.079)	(0.075)	(0.075)
Constant	0.420	0.322	0.373	0.384
	(0.041)	(0.040)	(0.081)	(0.082)
Article Characteristics			✓	✓
JEL Codes				✓
R-squared	0.002	0.012	0.017	0.019
Observations	17,653	17,653	17,653	17,653

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows OLS estimates of equation (1). The dependent variable is a dummy for whether the test statistic is significant at the 1% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table A7: Editorial Statement and Test Statistics: Significant at 10%

	Significant at the 10% Level			
	(1)	(2)	(3)	(4)
During/After	-0.063	0.047	0.040	0.040
	(0.036)	(0.052)	(0.049)	(0.049)
Treated Journals	-0.049	0.066	0.062	0.066
	(0.036)	(0.052)	(0.049)	(0.056)
During/After *		-0.166	-0.153	-0.162
Treated Journals		(0.069)	(0.066)	(0.067)
Constant	0.621	0.549	0.561	0.567
	(0.035)	(0.037)	(0.075)	(0.074)
Article Characteristics			✓	✓
JEL Codes				✓
R-squared	0.006	0.010	0.012	0.013
Observations	17,653	17,653	17,653	17,653

Sources: the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics* and two control journals. This table shows OLS estimates of equation (1). The dependent variable is a dummy for whether the test statistic is significant at the 10% level. Robust standard errors are in parentheses, adjusted for clustering by article. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.