Publication Bias and Editorial Statement on Negative Findings*

Cristina Blanco-Perez^a Abel Brodeur^b

November 2017

Abstract

In February 2015, the editors of eight health economics journals sent out an editorial statement which aims to reduce the incentives to engage in 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 that may be specified.” In this study, we collect z -statistics from two health economics journals and compare the distribution of tests before and after the editorial statement. Our results suggest that the editorial statement decreased the proportion of test statistics rejecting the null hypothesis and that incentives may be aligned to promote more transparent research.

KEYWORDS: Publication bias, specification searching, p -hacking, research in economics, incentives to publish.

JEL CODES: A11, C13, C44, I10.

*We thank Pierre Brochu, Garret Christensen, Chishio Furukawa, Jason Garred, Fernando Hoces de la Guardia and seminar participants at 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 Yuvraj Pathak for outstanding research assistance. All errors are our own.

^aUniversity of Ottawa, 120 University, Social Sciences Building, Ottawa, ON, Canada K1N 6N5; cblancop@uottawa.ca.

^bUniversity of Ottawa, 120 University, Social Sciences Building, Ottawa, ON, Canada K1N 6N5; abrodeur@uottawa.ca.

Our study deals with research transparency. 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 to publish their results (Brodeur et al. (2017); Fanelli (2010); Gerber and Malhotra (2008a); Gerber and Malhotra (2008b); Gerber et al. (2010); Henry (2009); Leamer (1983); Ioannidis (2005); 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); Simonsohn et al. (2014); Stanley (2008)). These practices (i.e., publication bias) may lead policymakers and the academic community to believe more in studies rejecting the null hypothesis than in studies presenting negative results (Card and Krueger (1995); De Long and Lang (1992); Doucouliagos and Stanley (2013); Furukawa (2017); Ioannidis et al. (2017)).

Fortunately, innovations in social sciences are under way to improve research transparency. Miguel et al. (2014) classify those practices in three categories: disclosure, registration and open data. Disclosure requirements are not as sophisticated in economics as they are in medical journals (Leamer and Leonard (1983); Leamer (1985)). But many applied economists report data exclusions and explain how they reached their final sample size. Registration and pre-analysis plans are more and more popular for randomized controlled trials (RCTs) and experiments (Casey et al. (2012); Olken (2015)). Whether data and codes are available on the website of the journal for replication purposes has attracted a great deal of attention lately.¹ Many economists put their programs (e.g., do-file) on their websites and many outlets require that the programs be available for replication (Camerer et al. (2016); Chang and Li (2015); Maniadis et al. (2017); McCullough et al. (2008)).

In this study, we test the impact of a simple, low-cost, new transparent practice that aims to reduce the incentives to engage in data mining and specification searching.² In February 2015, the editors of eight health economics journals sent out an editorial statement³ reminding authors to

¹This idea was already discussed in the early 1980s, see Dewald et al. (1986). Brodeur et al. (2016) provide evidence that data and code availability in top economic journals did not change the extent of publication bias.

²Another 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.

³See Appendix or [http : //ashecon.org/american - journal - of - health - economics/editorial - statement - on - negative - findings/](http://ashecon.org/american-journal-of-health-economics/editorial-statement-on-negative-findings/).

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.” We test whether this Editorial Statement on Negative Findings induced a change in the behavior of researchers, referees and editors.

Prior to data collection and analysis, the analysis presented here was pre-specified and publicly archived in a pre-analysis plan.⁴ Our pre-specification was designed to minimize issues of data and specification mining and to provide a record of the full set of planned analyzes.

We collect z -statistics from two of the eight health economics journals that sent out the editorial statement and compare the distribution of tests before and after the editorial statement. We show that the distribution of test statistics before the editorial statement had a two-humped camel shape with a first hump with low z -statistics and a second hump around 2 (p -values around 0.05). In this subsample of tests, approximately 61, 55 and 49% of test statistics were respectively significant at the 10%, 5% and 1% levels.

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. More precisely, the editorial statement decreased by about 6% the number of test statistics that are statistically significant at the 5% level. In other words, our results suggest that the editorial statement shifted the distribution of tests to the left and thus decreased the extent of publication bias. This result is consistent with anecdotal evidence that there is a larger fraction of submissions with negative findings.

As a robustness check, we look at whether there was a similar shift in the distribution of z -statistics at the time of the editorial statement for a 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 journal, possibly due to the increasing pressure to publish (Card and DellaVigna (2013)). This result may indicate that we underestimate the impact of the editorial statement for health economics journals. As an additional 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

⁴Our 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 Sciences for the data collection.

of papers that are single-authored or have a theoretical model. Note that the control journal and the robustness checks discussed in this paragraph were pre-specified.

In a set of regressions that were not pre-registered, we show that our results are robust to the inclusion of many additional controls such as authors' characteristics (e.g., tenured) and Google Scholar citations. Furthermore, we find that the impact of the editorial statement is mostly driven by multi-authored research articles and researchers from non-top institutions.

Last, we document that the impact of the statement "intensifies" over the time period studied. This result provides suggestive evidence that the decrease in the share of tests significant at conventional levels is due to both a change in editors' preferences for negative findings and a change in authors and/or referees' behavior. Overall, our findings suggest that a simple transparent practice such as the editorial statement may lead to more transparent research.

Our research question is directly related to research culture and adoption of new methods, and particularly to the question of whether incentives or norms may be structured to encourage transparent, reproducible research (Dufwenberg and Martinsson (2014); Nosek et al. (2015)). Our findings relate to Brodeur et al. (2016) and Vivalt (2017), who document the extent of publication bias in economics. 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-analyses 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 and Greenstone (2004); Fanelli (2009); Franco et al. (2014)). Our findings enhance our understanding of the external validity of economic research (Ashenfelter et al. (1999); Ellison (2002); Havránek (2015); Havránek and Sokolova (2016); McCloskey (1985)).

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. (See Appendix 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 November 2017, the statement is still highlighted on the News section of the *Journal of Health Economics* and in the Author Guidelines of *Health Economics*.

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 or other field journals. 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.⁵ 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.

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 two health economics journals

⁵It 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.

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 Methodology and Data Collection

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 decreased the extent of publication bias, i.e., specification searching and selection bias in favor of positive results.

One methodology used to measure the extent of publication bias in the literature is the “caliper test.” This methodology is based on the rate of reported occurrence of test statistics within a narrow band (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 particularly interested in analyzing the number of test statistics within a narrow band. We are more interested in the whole distribution of tests and in 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 is due to the fact that the editorial statement might increase the number of published studies with a high p -value.

More precisely, we look at the whole distribution of p -values before the editorial statement and compare it with other distributions of p -values (Brodeur et al. (2016)). 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.

We follow three steps to collect and prepare the data set. We first extract journal articles from multiple volumes of two health economics journals that sent out the editorial statement, then we classify journal articles with respect to the dates of submission and publication, and finally we (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 the *Journal of Health Economics* (JHE), *Health Economics* (HE) and a control journal. We focus primarily on the JHE and HE throughout for two reasons. First, some of the other health economics journals publish a very small number of papers per year and do not post Online the papers that are forthcoming. Second, we focus on the JHE and HE because of their general scope, i.e, they cover all aspects of health economics.⁶

We examine several volumes in both HE and the JHE⁷ and 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. Note that papers published in both the JHE and HE include the date the paper was first received, the date it was revised and the date it was accepted. We rely on those dates to classify papers into one of the three categories.

The data collection described above provides 185 articles, of which 94 are from the JHE and 60 from HE. We report solely tests from the main tables of the paper. That is, we report the main results, but do not report descriptive statistics, group comparisons, and explicit placebo tests or ro-

⁶One exception is the *American Journal of Health Economics*. We exclude this journal since it was founded in 2015.

⁷For the JHE, we collect data for two volumes (38-39) before the editorial statement and eight volumes (44-51) for papers submitted during or after the editorial statement. For HE, we collect test statistics from volume 25, Issue 4, and early view articles for papers classified in the categories during or after. Most papers analyzed in HE were thus available Online but not yet published in a volume.

bustness checks. 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 paper. The reporting guideline is available in the Online Appendix. A pre-analysis plan describing the hypothesis to be examined in the research study and the data collection was archived on November 28, 2016.

Authors report their estimates in different ways. Most authors present coefficients and standard errors, but some tables report p -values. We follow Brodeur et al. (2016) and 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 also collect test statistics for a control journal not affected by the editorial statement. This allow us to check whether the distribution of tests in health economics journals is changing over time for reasons other than the editorial statement. Note that we pre-specified this journal and chose it since it has an impact factor similar to the two health journals in our study. The control journal publishes a small number of papers in the field of health economics, and the methods used (e.g., differences-in-differences) are very similar. We follow the procedure described above and classify papers in three categories: “Before,” “During” and “After” the editorial statement. We collect data for two volumes before the editorial statement and three volumes for papers submitted either during or after the editorial statement.

Last, we collect additional information on each article, including the issue of the journal, the name and 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. We collect these characteristics since they have been shown to be related to publication bias. For instance, Brodeur et al. (2016) provide evidence that there is less specification searching in top economics journals for articles with a theoretical model and multiple-authored articles.

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 JHE, HE and a control journal. Our unit of observation is a test statistic.

In our main specification, we estimate:

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

where Y_{iajt} is the outcome variable (significant at conventional levels) for test statistic i in article a , journal j and time t . $\textit{During/After}_{iajt}$ is a dummy that equals one if test statistic i in article a was published “During” or “After” the editorial statement and zero otherwise. $\textit{Treated}_{iajt}$ is a dummy that equals one if test statistic i in article a was published in the JHE or HE and zero if it was published in the control journal. The interaction of $\textit{During/After}_{iajt}$ and $\textit{Treated}_{iajt}$ 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 two treated journals and the control group would follow the same time trend in the absence of the treatment, i.e., editorial statement.

X'_{iajt} is a vector containing regressors known to be predictors of publication bias. Following our pre-analysis plan, 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. X'_{iajt} also contains additional regressors that were not pre-specified. Following comments from colleagues, seminar participants and discussants, we include in the empirical model the following four variables: ratio of female authors, ratio of authors at top institutions,⁸ ratio of tenured (or associate professor) authors at the time of the publication and the ratio of authors who published in a top economics journal⁹ at the time of the publication.

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 probit regressions in the Online Appendix. The estimates are virtually the same as the OLS estimates.

⁸We consider the following universities as top institutions: Boston U, Brown, Chicago, Columbia, Dartmouth, Harvard, LSE, Michigan, MIT, Northwestern, NYU, Princeton, UC Berkeley, UCLA, UCSD, UPenn, Stanford and Yale. Adding or removing some universities to this list has no effect on our main results (results available upon request).

⁹We consider the following five outlets as top economics journals: the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics* and the *Review of Economic Studies*.

C Descriptive Statistics

Table 1 provides descriptive statistics on the number of articles (column 1), tables (column 2) and test statistics (column 3) per journal and along articles and authors' characteristics. In HE, the average number of tables per article is slightly higher than 2 while the average number of test statistics per article is about 90. The average number of tables per article is slightly lower than 2 for the JHE, and the number of test statistics per article (50) is much lower than for HE. There are approximately 80 test statistics per article in the control journal. Fifty-one percent of the articles in our sample were published in the JHE against 32% for HE and 17 for the control journal. But there are more test statistics for HE (43%) than for JHE (37%).

Approximately 30% of papers in our sample are single-authored and 18% have a theoretical model. About 93% 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). More than half of the research articles have at least one female author. Two-thirds of articles have at least one tenured author. About 20% of articles have at least one author from a top institution or with at least one publication in a top economics journal.

Table 2 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, 21% and 32% of tests are in the categories “Before” and “During” the editorial statement.

Table 2 also shows the number of tests for the three categories according to article and authors' 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 smaller for the category “After” than for the other categories. Respectively, 14% and 21% of tests are in articles with a model for the categories “After” and “Before” the editorial statement. On the other hand, 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 illustrates the distribution of z -statistics for the category “Before” in HE and the JHE. 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.¹⁰ Figure 1, panel A, reports the raw distribution. 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 differences-in-differences approach with a non-health economics journal as a comparison group to test whether the Editorial Statement of Negative Findings decreased the extent of publication bias in the health economics journals. Next, we test the robustness of the results to changes in the characteristics of the papers published and to changes in authors’ characteristics.

A Descriptive Analysis

Figure 2 plots the distribution of z -statistics for HE and the JHE combined. (See Appendix Figures 5 and 6 for the distributions of z -statistics of each journal.) 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 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. For the weighted distribution,¹¹ about 41% of z -statistics are statistically significant at the 1% level after the editorial statement. In comparison, 49% of z -statistics were significant at the 1% level before the statement. Approximately 44%

¹⁰We rely on an Epanechnikov kernel density to smooth jumps in the distributions. We use a bandwidth of 0.20. This explains why there is a sharp incline from $[0, 0.2]$.

¹¹In Figure 3 and Appendix Figure 7, we check whether our results are sensitive to the overweight attributed to articles with many test statistics. These figures suggest that putting the same weight for each article or table has no effect on our finding that the distribution of tests during and after the statement shifted leftwards.

of z -statistics are significant at the 1% level for the category “During,” suggesting that the effect of the editorial statement is possibly intensifying.

While we do not know what the distribution of test statistics in the health economics journals would be in the absence of publication bias, we think it is useful to use the distribution of z -statistics for RCT papers published in the *American Economics Review*, the *Journal of Political Economy* and the *Quarterly Journal of Economics* as a benchmark (Brodeur et al. (2016)). Arguably, articles relying on a RCT suffer from less publication bias than non-RCT articles (Blanco-Perez and Brodeur (2017)). Approximately 37% of z -statistics are statistically significant at the 5% level with RCTs in comparison to 49% in our sample after the editorial statement. This comparison suggest that the editorial statement might have reduced selection bias in favor of positive results and specification searching, but that there is still room for more transparent research.

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 3. (See Appendix Table 8 for the unweighted distribution.) Columns 1, 2 and 3 report means and standard deviations for three subsamples. Columns 1, 2 and 3 restrict the sample respectively to papers published before, during and after the editorial statement. 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 both health journals for this exercise.

The estimates presented in Table 3 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. Approximately 56%, 49% and 41% of z -statistics, respectively, are statistically significant at the 10%, 5% and 1% levels after the editorial in comparison with 61%, 55% and 49% of z -statistics before the editorial statement. The difference between “Before” and “During” is also significant (at the 5% level) for the three significance thresholds. About 57%, 52% and 44% of z -statistics are respectively statistically significant 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 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.”

If the editors’ preferences for negative findings did not change from “During” to “After,” this would be indicative that the editorial statement affected referees, and/or authors’ behavior. We assume that it is the case and check whether there is a decrease in the share of tests that are significant between “During” and “After.” We report t -test values for the equality of the means between columns 2 (“During”) and 3 (“After”) in column 6 of Table 3. The estimates are statistically significant at the 1% level for two of the significance levels (5% and 1%), suggesting that publication bias decreased between these two periods. Of note, though, our research focuses on a time period in which the editorial statement was relatively new. The impact of the statement may thus “intensify” in the coming years.

Thus far, we looked at whether the editorial statement changed the whole distribution of z -statistics. 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. Appendix Table 9 replicates Table 3, restricting the sample to three subsamples. Panel A restricts the sample to papers without a theoretical model while panel B restricts the sample to multiple-authored papers. Panel C keeps only papers with at least one of the four most popular JEL codes in our study (I10, I11, I13 and I18).¹²

For the three panels, the difference between columns 1 and 3 (“Before” and “After”) remains statistically significant at the 5% level for the three significance levels. This is suggestive evidence that the estimated effect of the editorial statement is not driven by compositional changes.

We further investigate whether the distribution of tests also changed from “Before” to “After” for a non-health economics journal. This is an important robustness check since the distribution of tests in HE and the JHE 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

¹²Appendix Figure 10 shows the distribution of z -statistics across the most popular JEL codes in our sample (I10, I11, I13 and I18). This bird’s-eye view across JEL codes indicates that there is some heterogeneity in the shape of the distribution. Appendix Figure 12 illustrates the tests distributions for single-authored and multi-authored papers and articles with and without a theoretical model.

journals also changed over time.

Figure 4 plots the weighted distribution of z -statistics for our control journal. (See Appendix Figure 8 for the raw distribution.) We find that the distribution of tests shifted *rightwards* after the editorial statement. This pattern is confirmed in Table 3, panel B. The estimates show that from “Before” to “After,” the number of tests that are significant at the 10%, 5% and 1% levels significantly *increased*. Moreover, the share of tests significant at the 10% and 5% level also significantly increased from “During” to “After.” These results suggest that publication bias for our control journal increased over the period studied.

The rightward shift of the distribution of tests for the control journal possibly reflects a change in the preferences for negative findings from the editors and referees or simply more specification searching from authors. Note that this journal publishes few health economics papers per year. We omit these health economics papers throughout.¹³ The comparison with the control journal thus offers evidence that we may have underestimated the impact of the editorial statement in Table 3, panel A.

B Main Results

Table 4 contains (unweighted) OLS estimates of equation (1) for the JHE, HE and the control journal over the “Before,” “During” and “After” periods. The sample size is 12,695 observations (i.e., test statistics).¹⁴ 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 variable $During/After_{iajt}$ captures the simple difference between the periods “Before” and “During” and “After” combined. The estimate is also small and statistically insignificant.

In column 2, we add the interaction term necessary in our differences-in-differences model. Columns 3–5 augment the model progressively with more controls. Coefficients in columns 2–5 suggest that the editorial statement decreased the proportion of test statistics rejecting the null hypothesis

¹³Including the few health economics papers has no effect on the share of tests statistically significant at conventional levels for the control journal.

¹⁴Appendix Table 12 shows the estimates for our control variables. We find a positive association between the likelihood to reject the null hypothesis and the proportion of male authors, the proportion of authors with publication in top journals and single-authored articles.

by about 20%. The estimates are statistically significant at the 5% level in all columns. Adding authors’ characteristics (not pre-registered) and journal articles’ characteristics (pre-registered) has no effect on the size and significance of the results. In column 6, we also include JEL codes and the natural log of the number of Google Scholar citations. Arguably, Google Scholar citations help to capture the quality of research articles. Our coefficient of interest remains negative and significant at the 5% level.

Appendix Tables 10 and 11 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 24% and 16%. The estimates are statistically significant and confirm the findings presented in Table 4.

C Robustness Checks

In Table 5, we present estimates from a number of robustness checks. The number of test statistics per article varies quite a lot in our sample. For instance, eight articles have over 200 test statistics while 32 have less than 20. This may be an issue if papers not rejecting the null hypothesis in the category “After” have more test statistics than papers not rejecting the null hypothesis published before the editorial statement.¹⁵ We thus check in columns 1 and 2 whether the share of z -statistics that are significant at the 5% or 1% level is similar when we weight test statistics using either the inverse of the number of tests presented in the same article (Table 5, column 1 and Figure 1, panel B) or the inverse of the number of tests presented in the same table (or result) multiplied by the inverse of the number of tables in the article to weight observations (Table 5, column 2 and Figure 1, panel C). The estimates are significant at the 5% level in both columns and suggest that the editorial statement reduced the proportion of tests significant at the 5% level by 21%, confirming that our results are not driven by few articles with many tests.

Columns 3 and 4 omit from the sample one of the two treated journals. The estimates remain negative and significant suggesting that our main findings are not driven by either HE or the JHE. Column 5 reports a robustness check that was not pre-registered. Column 5 restricts the sample to the first table of each paper with reported estimates, i.e., the first table

¹⁵The average number of tables per paper is slightly lower for the category “Before” than “During” and “After” the editorial statement with, respectively, 1.7 and 1.9 tables per paper. Articles published before the editorial statement also have less test statistics per paper (60) than the categories “During” and “After” (70).

from which we collected test statistics. This exercise allows us to omit robustness checks that were erroneously included in the data collection and to make sure that we only include the main results from each article. The estimated coefficient is slightly larger than the estimates presented in Table 4 and statistically significant at the 1% suggesting that our results are robust to table ordering.

We now check whether the proportion of test statistics not rejecting the null hypothesis in the 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 collect additional volumes for both the JHE and HE prior to the category “Before.” We collect test statistics from one volume for both the JHE (volume 33) and HE (volume 24). These volumes include research articles published exactly one year prior to the category “Before.” (Note that this data collection was not pre-registered.) In this subsample of tests, about 63, 54 and 44% of test statistics were respectively significant at the 10%, 5% and 1% levels in comparison to 61, 55 and 49% for the category “Before.”

We test explicitly whether there was an increase or a decrease in the proportion of tests rejecting the null hypothesis prior to the editorial statement in Appendix Table 14. We restrict the sample to treated journals and to the time periods “Before” and prior to the category “Before.” When the full set of controls are included, our estimates suggest that there is no significant difference between the number of tests rejecting the null hypothesis for the time periods “Before” and prior to “Before.” The estimates are small, positive and statistically insignificant.

Overall, we find evidence that the editorial statement reduced publication bias in two 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 referees and authors and document the characteristics of authors who are now less likely to reject the null hypothesis.

IV Mechanisms

A Is the Effect Intensifying?

In this subsection, we check whether the statement has affected the distribution of tests in health economics journals through a change in referees and authors’ behavior. We rely on an empirical model focusing exclusively

on the time categories “During” and “After,” and explore for the presence of an effect over time:

$$Y_{iajt} = \alpha + \mu After_{iajt} + X'_{ajt}\lambda + \varepsilon_{iajt}, \quad (2)$$

where Y_{iajt} is an outcome variable and $After_{iajt}$ is a dummy that assumes the value of one if the time category is “After” and zero if the time period is “During.” The sample is restricted to the two health economics journals. The coefficient of interest, μ , captures the impact of the statement from the time period “During” to the time period “After.” μ will be negative if the effect of the editorial statement is intensifying over time.

Table 6 presents the OLS estimates. We include the full set of controls in the model and the dependent variables are dummies for whether the test statistic is significant at the 5% level in columns 1–3 and significant at the 1% level in columns 4–6. Columns 1 and 4 include the two health economics journals in the sample. The estimates are small and statistically insignificant.

The other columns test whether the editorial statement has heterogeneous effects on the two health economics journals. Columns 2 and 5 restrict the sample to one of the health economics journals. Columns 3 and 6 restrict the sample to the other health economics journal. The estimates are small, positive and statistically insignificant for one of the two health economics journals. On the other hand, the estimates are negative and significant for the other journal and suggest that the proportion of tests rejecting the null hypothesis decreased by about 11% between the time periods “During” and “After.” This finding provides suggestive evidence that the statement changed authors and referees’ behavior for one of the health economics journals.

B Subgroup Analysis

We first check whether our results are driven by research articles from the coeditors and associate editors of the health economics journals themselves. 16 research articles were published by an associate editor of the two journals and 2 articles were published by a coeditor. We thus exclude from the sample publications from HE and JHE coeditors and coeditors and associate editors. The estimates (not shown) range from -0.23 to -0.24 and are significant at the 5% level. This means that the impact of editorial statement is not driven by research articles written by the editors of these two journals.

We now test whether the statement has a larger effect on authors and papers with specific characteristics. More precisely, we check whether the leftward shift in the distribution of tests after the statement is partly due to researchers from top institutions, to senior authors and to single-authored research articles. Table 7 provides estimates of equation (1). Column 1 restricts the sample to journal articles with at least one tenured author. Column 2 keeps only articles without a tenured author. The estimates are negative in both columns, but larger in column 2 than in column 1. In other words, the statement seems to have a larger effect on non-tenured authors' behavior than on tenured authors' behavior.

In columns 3 and 4, we check whether the leftward shift is driven by researchers from top institutions. Column 3 restricts the sample to research articles without an author from a top institution and without a publication in a top economics journal. In column 4, we keep only research articles published by at least one researcher from a top institution or by an author with at least one publication in a top economics journal. The estimates for the interaction term are negative in both columns, but larger in column 3.

Columns 5 and 6 restrict the sample to single-authored research articles and to multi-authored research articles. 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, non-tenured researchers and researchers from non-top institutions. These results are interesting since there is a positive association between single-authored research articles and the likelihood to reject the null hypothesis (Appendix Table 12). Researchers with at least one top publication are also more likely to reject the null hypothesis in our sample. These findings suggest that the statement has affected mostly researchers who were already less likely 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 13. 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 13 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 incentives to engage in specification searching and data mining. There are two main empirical results. First, 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 two health economics journals. We performed several robustness checks and concluded that the editorial statement reduced the extent of publication bias.

Second, we documented whether the editorial statement affected editors, referees, and/or researchers' behavior. 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. 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 submitted after the editorial ("After") than for papers submitted before the statement, but accepted after ("During"). This result provides some evidence that the editorial statement induced referees and/or authors to change their behavior.

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 and that editors may reduce the extent of publication bias quite easily.¹⁶ Ioannidis et al. (2017)

¹⁶The impact factor of HE and JHE remained stable for the years before and after the editorial statement (Appendix Figure 14). This is suggestive evidence that the statement did not affect the quality of the papers submitted to the eight health economics journals.

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.

Finally, we believe further work could test the effectiveness of other transparent practices using the methods described in this research.

References

- 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., Mullainathan, S. et al.: 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/>.
- Brodeur, A., Cook, N. and Heyes, A.: 2017, Methods Matter. University of Ottawa Mimeo.
- 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).

- 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.
- Doucouliaagos, 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.
- Ellison, G.: 2002, The Slowdown of the Economics Publishing Process, *Journal of Political Economy* **110**(5), 947–993.
- 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 medicine* **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.
- Leamer, E. E.: 1985, Sensitivity Analyses Would Help, *American Economic Review* **75**(3), 308–313.
- Leamer, E. E. and Leonard, H.: 1983, Reporting the Fragility of Regression Estimates, *Review of Economics and Statistics* **65**(2), pp. 306–317.
- 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.

- 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.
- Simonsohn, U., Nelson, L. and Simmons, J. P.: 2014, P-Curve: A Key to the File Drawer, *Journal of Experimental Psychology: General* **143**, 534–547.
- 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 the editorial.

(a) Raw distribution of z-statistics.

(b) Distribution of z-statistics, weighted by articles.

(c) Distribution of z-statistics, weighted by articles and tables.

Sources: *Health Economics* and *Journal of Health Economics* before the editorial, i.e., papers that were submitted and published before the statement on negative findings. The distribution presented in sub-figure (b) uses the inverse of the number of tests presented in the same article to weight observations. The distribution presented in sub-figure (c) uses the inverse of the number of tests presented in the same table (or result) multiplied by the inverse of the number of tables in the article to weight observations. Lines correspond to kernel density estimates.

Figure 2: 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: *Health Economics* and *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 3: Weighted distributions of z-statistics before, during and after the editorial.

(a) Raw distribution of z-statistics, weighted by articles. Before the editorial. (b) Raw distribution of z-statistics, weighted by articles. During the editorial.

(c) Raw distribution of z-statistics, weighted by articles. During and after the editorial. (d) Raw distribution of z-statistics, weighted by articles. After the editorial.

Sources: *Health Economics* and *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. The distributions use the inverse of the number of tests presented in the same article to weight observations. Lines correspond to kernel density estimates.

Figure 4: Weighted distributions of z-statistics for the control journal.

(a) Raw distribution of z-statistics, weighted by articles. Before the editorial. (b) Raw distribution of z-statistics, weighted by articles. During the editorial.

(c) Raw distribution of z-statistics, weighted by articles. During and after the editorial. (d) Raw distribution of z-statistics, weighted by articles. After the editorial.

Sources: Control journal. 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. The distributions use the inverse of the number of tests presented in the same article to weight observations. Lines correspond to kernel density estimates.

Table 1: Descriptive Statistics: *Health Economics*, *Journal of Health Economics* and Control Journal.

	Articles	Number of ... Tables	Tests
<i>Health Economics</i>	60 [32]	153 [35]	5,391 [43]
<i>Journal Health Economics</i>	94 [51]	169 [39]	4,752 [37]
Control Journal	31 [17]	113 [26]	2,552 [20]
Using Stars	170 [93]	362 [96]	12,238 [96]
With Model	33 [18]	57 [15]	2,151 [17]
Single-Authored	42 [23]	98 [26]	3,107 [24]
JEL I10	16 [9]	34 [9]	1,189 [9]
JEL I11	17 [9]	27 [7]	1,121 [9]
JEL I13	20 [11]	39 [10]	1,004 [8]
JEL I18	44 [24]	77 [20]	2,511 [20]
Top Institutions (ratio > 0)	40 [22]	84 [22]	2,289 [18]
Tenured (ratio > 0)	131 [70]	250 [66]	8,570 [68]
Top 5 (ratio > 0)	43 [23]	97 [26]	2,605 [21]
Female (ratio > 0)	109 [59]	209 [55]	7,220 [57]

Sources: *Health Economics*, *Journal of Health Economics* and control journal. This table reports the number of articles, tables and test statistics for each variable. Proportions relatively to the total population are indicated between brackets. *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: Descriptive Statistics: Before, During and After the Editorial.

	Number of Tests ...		
	Before (1)	During (2)	After (3)
<i>Health Economics</i>	631 [23]	2,017 [50]	2,742 [46]
<i>Journal Health Economics</i>	986 [37]	1,369 [34]	2,397 [40]
Control Journal	1,077 [40]	683 [17]	792 [13]
Using Stars	2,636 [98]	3,865 [95]	5,737 [97]
With Model	553 [21]	761 [19]	837 [14]
Single-Authored	823 [30]	939 [22]	1,375 [22]
JEL I10	396 [15]	394 [10]	399 [7]
JEL I11	210 [8]	691 [17]	340 [6]
JEL I13	53 [2]	348 [9]	603 [10]
JEL I18	246 [9]	1,009 [25]	1,376 [23]
Top Institutions (ratio > 0)	713 [26]	800 [20]	776 [13]
Tenured (ratio > 0)	1,852 [69]	2,647 [65]	4,071 [68]
Top 5 (ratio > 0)	836 [31]	599 [15]	1,170 [20]
Female (ratio > 0)	1,586 [59]	2,600 [64]	3,034 [51]

Sources: *Health Economics*, *Journal of Health Economics* and control journal. This table reports the number 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. Proportions relatively to the total population are indicated between brackets. *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 3: Weighted Percentage of Test Statistics Statistically Significant at Conventional Levels.

<i>Health Economics and Journal of Health Economics</i>						
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel A</i>						
Significant at 1%	0.494 (0.500)	0.441 (0.497)	0.410 (0.492)	-0.054 (0.015)	-0.084 (0.013)	-0.031 (0.011)
Significant at 5%	0.551 (0.497)	0.518 (0.500)	0.488 (0.500)	-0.034 (0.015)	-0.064 (0.013)	-0.030 (0.011)
Significant at 10%	0.605 (0.489)	0.572 (0.495)	0.559 (0.497)	-0.033 (0.015)	-0.046 (0.013)	-0.014 (0.011)
<i>Control Journal</i>						
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel B</i>						
Significant at 1%	0.372 (0.484)	0.536 (0.499)	0.585 (0.493)	0.164 (0.032)	0.213 (0.025)	0.049 (0.033)
Significant at 5%	0.454 (0.498)	0.567 (0.496)	0.654 (0.476)	0.113 (0.033)	0.200 (0.025)	0.087 (0.032)
Significant at 10%	0.544 (0.498)	0.599 (0.491)	0.704 (0.457)	0.055 (0.033)	0.160 (0.025)	0.105 (0.031)

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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). We use the inverse of the number of tests presented in the same article to weight observations.

Table 4: Editorial Statement and Test Statistics: Main Estimates

	Significant at the 5% Level					
	(1)	(2)	(3)	(4)	(5)	(6)
During/After	-0.012 (0.056)	0.148 (0.075)	0.139 (0.073)	0.145 (0.081)	0.137 (0.081)	0.144 (0.081)
Treated Journals	-0.006 (0.052)	0.148 (0.079)	0.130 (0.079)	0.174 (0.076)	0.170 (0.081)	0.176 (0.084)
During/After * Treated Journals		-0.234 (0.103)	-0.214 (0.101)	-0.227 (0.103)	-0.208 (0.101)	-0.202 (0.103)
Authors Characteristics			✓		✓	✓
Article Characteristics				✓	✓	✓
JEL Codes					✓	✓
Ln (Citations)						✓
R-squared	0.000	0.008	0.010	0.014	0.022	0.022
Observations	12,695	12,695	12,695	12,695	12,695	12,695

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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. Authors' characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. 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				
	Weight Article (1)	Weight Table (2)	Omit Journal (3)	Omit Other Journal (4)	Only 1st Table (5)
During/After	0.181 (0.089)	0.188 (0.088)	0.167 (0.077)	0.146 (0.075)	0.311 (0.095)
Treated Journals	0.130 (0.077)	0.150 (0.072)	0.206 (0.132)	0.153 (0.082)	0.140 (0.097)
During/After * Treated Journals	-0.209 (0.103)	-0.208 (0.100)	-0.229 (0.137)	-0.253 (0.095)	-0.331 (0.115)
Authors Characteristics	✓	✓	✓	✓	✓
Article Characteristics	✓	✓	✓	✓	✓
JEL Codes & Citations	✓	✓	✓	✓	✓
R-squared	0.026	0.026	0.050	0.033	0.044
Observations	12,695	12,695	7,943	7,304	6,944

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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 use the inverse of the number of tests presented in the same article to weight observations. In column 2, we use the inverse of the number of tests presented in the same table multiplied by the inverse of the number of tables in the article to weight observations. In column 3, we omit one of the treated journals. In column 4, we omit the other treated journal. Column 5 restricts the sample to tests from the first table of each paper. Authors' characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. 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: After in Comparison to During the Statement

	Significant at 5%			Significant at 1%		
	(1)	(2)	(3)	(4)	(5)	(6)
After	-0.004 (0.054)	-0.109 (0.059)	0.008 (0.079)	-0.012 (0.055)	-0.116 (0.061)	0.009 (0.083)
Authors Characteristics	✓	✓	✓	✓	✓	✓
Article Characteristics	✓	✓	✓	✓	✓	✓
JEL Codes & Citations	✓	✓	✓	✓	✓	✓
R-squared	0.019	0.102	0.058	0.024	0.110	0.066
Observations	10,001	3,766	4,760	10,001	3,766	4,760

Sources: *Health Economics* and *Journal of Health Economics*. This table shows OLS estimates of equation (2). We restrict the sample to the time periods “During” and “After.” In columns 1–3, the dependent variable is a dummy for whether the test statistic is significant at the 5% level. In columns 4–6, 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. Authors’ characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. 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					
	At Least 1 Tenured (1)	No Tenured (2)	No Top Pub. or Inst. (3)	Top Pub. or Inst. (4)	Not Single- Authored (5)	Single- Authored (6)
During/After	0.106 (0.103)	0.262 (0.135)	0.106 (0.098)	0.197 (0.101)	0.219 (0.088)	-0.236 (0.183)
Treated Journals	0.131 (0.092)	0.195 (0.158)	0.195 (0.114)	0.097 (0.106)	0.154 (0.081)	-0.105 (0.201)
During/After * Treated Journals	-0.145 (0.130)	-0.295 (0.153)	-0.208 (0.128)	-0.164 (0.141)	-0.270 (0.113)	0.124 (0.197)
Authors Characs	✓	✓	✓	✓	✓	✓
Article Characs	✓	✓	✓	✓	✓	✓
JEL Codes & Cites	✓	✓	✓	✓	✓	✓
R-squared	0.026	0.107	0.026	0.045	0.035	0.107
Observations	8,570	4,125	9,279	3,416	9,588	3,083

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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 at least one tenured author. Column 2 keeps only articles without a tenured author. Column 3 restricts the sample to research articles without an author from a top institution and without a publication in a top economics journal. In column 4, we keep only research articles published by at least one researcher from a top institution or at least one publication in a top economics journal. Columns 5 and 6 restrict the sample respectively to single-authored research articles and to multi-authored research articles. Authors’ characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. 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 5: Distributions of z-statistics before, during and after the editorial for Journal I.

Sources: *Health Economics* or 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 6: Distributions of z-statistics before, during and after the editorial for Journal II.

(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: *Health Economics* or 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 7: Weighted distributions of z-statistics.

Sources: *Health Economics* and *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. The distributions use the inverse of the number of tests presented in the same table multiplied by the inverse of the number of tables in the article to weight observations. Lines correspond to kernel density estimates.

Figure 8: Distributions of z-statistics for the control journal.

Sources: Control journal. 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 9: Weighted distributions of z-statistics for the control journal.

Sources: Control journal. 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. The distributions use the inverse of the number of tests presented in the same table multiplied by the inverse of the number of tables in the article to weight observations. Lines correspond to kernel density estimates.

Figure 10: Distributions of z-statistics for different sub-samples: model and single-authored.

(a) Raw distribution of z-statistics. Model. (b) Raw distribution of z-statistics. No model.

(c) Raw distribution of z-statistics. Single-authored. (d) Raw distribution of z-statistics. Multiple authors.

Sources: *Health Economics* and *Journal of Health Economics* before, during and after the editorial statement on negative findings. Lines correspond to kernel density estimates.

Figure 11: Distributions of z-statistics by JEL code.

(a) Raw distribution of z-statistics. I10: Health General. (b) Raw distribution of z-statistics. I11: Analysis of Health Care Markets.

(c) Raw distribution of z-statistics. I13: Health Insurance, Public and Private. (d) Raw distribution of z-statistics. I18: Health Government Policy, Regulation and Public Health.

Sources: *Health Economics* and *Journal of Health Economics* before, during and after the editorial statement on negative findings. Lines correspond to kernel density estimates.

Figure 12: Distributions of z-statistics by JEL code.

(a) Raw distribution of z-statistics. I10: Health General. (b) Raw distribution of z-statistics. I11: Analysis of Health Care Markets.

(c) Raw distribution of z-statistics. I13: Health Insurance, Public and Private. (d) Raw distribution of z-statistics. I18: Health Government Policy, Regulation and Public Health.

Sources: *Health Economics* and *Journal of Health Economics* before, during and after the editorial statement on negative findings. Lines correspond to kernel density estimates.

Figure 13: 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.

Figure 14: Impact factor of four health economics journals.

Sources: *Journal of Health Economics* (JHE); *Health Economics* (HE); *European Journal of Health Economics* (EJHE); *Health Economics, Policy and Law* (HEPL). CitEc is a RePEc service, providing citation data for economics journals and working papers. Data available here: [http : //citec.repec.org/search.html](http://citec.repec.org/search.html)

Table 8: Percentage of Test Statistics Statistically Significant at Conventional Levels.

<i>Health Economics and Journal of Health Economics</i>						
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel A</i>						
Significant at 1%	0.487 (0.500)	0.390 (0.488)	0.396 (0.489)	-0.098 (0.015)	-0.091 (0.014)	0.006 (0.011)
Significant at 5%	0.549 (0.498)	0.455 (0.498)	0.467 (0.499)	-0.094 (0.015)	-0.082 (0.014)	0.012 (0.011)
Significant at 10%	0.606 (0.489)	0.506 (0.500)	0.536 (0.499)	-0.100 (0.015)	-0.071 (0.014)	0.029 (0.011)
<i>Control Journal</i>						
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel B</i>						
Significant at 1%	0.316 (0.465)	0.534 (0.500)	0.499 (0.500)	0.217 (0.030)	0.183 (0.025)	-0.034 (0.033)
Significant at 5%	0.407 (0.492)	0.570 (0.496)	0.566 (0.496)	0.163 (0.031)	0.159 (0.025)	-0.004 (0.032)
Significant at 10%	0.510 (0.500)	0.604 (0.490)	0.618 (0.486)	0.094 (0.031)	0.109 (0.025)	0.014 (0.032)

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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 9: Weighted Percentage of Test Statistics Statistically Significant at Conventional Levels: *Health Economics and Journal of Health Economics*.

	No Theoretical Model					
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel A</i>						
Significant at 1%	0.467 (0.499)	0.427 (0.495)	0.402 (0.490)	-0.041 (0.018)	-0.066 (0.015)	-0.026 (0.012)
Significant at 5%	0.522 (0.500)	0.506 (0.500)	0.480 (0.500)	-0.017 (0.018)	-0.043 (0.015)	-0.026 (0.012)
Significant at 10%	0.578 (0.494)	0.560 (0.497)	0.551 (0.497)	-0.019 (0.018)	-0.028 (0.015)	-0.009 (0.012)
	Multiple Authors					
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel B</i>						
Significant at 1%	0.496 (0.500)	0.422 (0.494)	0.408 (0.492)	-0.075 (0.018)	-0.089 (0.016)	-0.015 (0.012)
Significant at 5%	0.549 (0.498)	0.502 (0.500)	0.484 (0.500)	-0.047 (0.019)	-0.065 (0.016)	-0.018 (0.012)
Significant at 10%	0.596 (0.497)	0.560 (0.496)	0.552 (0.491)	-0.036 (0.018)	-0.045 (0.016)	-0.009 (0.012)
	Popular JEL Codes					
	Before (1)	During (2)	After (3)	Diff (1)-(2) (4)	Diff (1)-(3) (5)	Diff (2)-(3) (6)
<i>Panel C</i>						
Significant at 1%	0.478 (0.500)	0.458 (0.498)	0.379 (0.485)	-0.020 (0.022)	-0.099 (0.021)	-0.079 (0.016)
Significant at 5%	0.540 (0.498)	0.523 (0.500)	0.465 (0.499)	-0.017 (0.022)	-0.076 (0.021)	-0.058 (0.017)
Significant at 10%	0.594 (0.492)	0.572 (0.495)	0.548 (0.498)	-0.021 (0.021)	-0.045 (0.021)	-0.024 (0.017)

Sources: *Health Economics* and *Journal of Health 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). We use the inverse of the number of tests presented in the same article to weight observations. Panel A restricts the sample to papers without a theoretical model. Panel B restricts the sample to papers with at least two authors. Panel C restricts the sample to papers with at least one of the following JEL codes: I10, I11, I13 and I18.

Table 10: Editorial Statement and Test Statistics: Significant at 1%

	Significant at the 1% Level					
	(1)	(2)	(3)	(4)	(5)	(6)
During/After	-0.006 (0.059)	0.185 (0.078)	0.174 (0.076)	0.182 (0.084)	0.172 (0.082)	0.182 (0.083)
Treated Journals	-0.005 (0.056)	0.179 (0.084)	0.161 (0.084)	0.201 (0.081)	0.199 (0.086)	0.208 (0.090)
During/After * Treated Journals		-0.279 (0.109)	-0.257 (0.106)	-0.271 (0.108)	-0.253 (0.107)	-0.245 (0.109)
Authors Characteristics			✓		✓	✓
Article Characteristics				✓	✓	✓
JEL Codes & Citations					✓	✓
R-squared	0.000	0.011	0.013	0.017	0.026	0.027
Observations	12,695	12,695	12,695	12,695	12,695	12,695

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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. Authors' characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 11: Editorial Statement and Test Statistics: Significant at 10%

	Significant at the 10% Level					
	(1)	(2)	(3)	(4)	(5)	(6)
During/After	-0.024 (0.050)	0.101 (0.073)	0.095 (0.072)	0.097 (0.079)	0.090 (0.079)	0.093 (0.078)
Treated Journals	-0.011 (0.048)	0.110 (0.073)	0.093 (0.074)	0.141 (0.071)	0.137 (0.075)	0.139 (0.078)
During/After * Treated Journals		-0.184 (0.096)	-0.166 (0.094)	-0.176 (0.096)	-0.159 (0.094)	-0.157 (0.096)
Authors Characteristics			✓		✓	✓
Article Characteristics				✓	✓	✓
JEL Codes & Citations					✓	✓
R-squared	0.001	0.005	0.007	0.012	0.019	0.019
Observations	12,695	12,695	12,695	12,695	12,695	12,695

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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. Authors' characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 12: Editorial Statement and Test Statistics: Covariates

	Significant 1% (1)	Significant 5% (2)	Significant 10% (3)
Using Stars	-0.043 (0.024)	0.001 (0.024)	0.028 (0.024)
With Model	0.082 (0.012)	0.078 (0.012)	0.074 (0.012)
Single-Authored	0.082 (0.013)	0.084 (0.013)	0.075 (0.013)
JEL I10	-0.055 (0.015)	-0.047 (0.016)	-0.050 (0.016)
JEL I11	-0.006 (0.017)	-0.009 (0.017)	0.001 (0.017)
JEL I13	-0.032 (0.018)	-0.028 (0.018)	-0.028 (0.018)
JEL I18	0.054 (0.012)	0.051 (0.012)	0.041 (0.012)
Top Institutions (ratio > 0)	0.025 (0.020)	0.027 (0.020)	0.027 (0.020)
Tenured (ratio > 0)	0.115 (0.016)	0.111 (0.016)	0.097 (0.016)
Top 5 (ratio > 0)	0.058 (0.026)	0.081 (0.026)	0.101 (0.025)
Female (ratio > 0)	-0.074 (0.012)	-0.076 (0.012)	-0.084 (0.013)
R-squared	0.018	0.016	0.015
Observations	12,695	12,695	12,695

Sources: *Health Economics*, *Journal of Health Economics* and control journal. 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 13: Editorial Statement and Test Statistics: Probit Regressions

	Significant at the 5% Level					
	(1)	(2)	(3)	(4)	(5)	(6)
During/After	-0.012 (0.055)	0.148 (0.075)	0.139 (0.074)	0.145 (0.081)	0.135 (0.080)	0.142 (0.081)
Treated Journals	-0.006 (0.052)	0.148 (0.079)	0.130 (0.080)	0.174 (0.077)	0.170 (0.081)	0.175 (0.084)
During/After * Treated Journals		-0.234 (0.102)	-0.214 (0.100)	-0.226 (0.102)	-0.207 (0.100)	-0.201 (0.102)
Authors Characteristics			✓		✓	✓
Article Characteristics				✓	✓	✓
JEL Codes & Citations					✓	✓
R-squared	0.000	0.008	0.007	0.010	0.016	0.016
Observations	12,695	12,695	12,695	12,695	12,695	12,695

Sources: *Health Economics*, *Journal of Health Economics* and control journal. This table reports marginal effects from probit regressions (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. Authors' characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

Table 14: Editorial Statement and Test Statistics: Before the Editorial Statement

	Significant 1%	Significant 5%	Significant 10%
	(1)	(2)	(3)
Before	0.087 (0.103)	0.050 (0.100)	0.044 (0.090)
Authors Characteristics	✓	✓	✓
Article Characteristics	✓	✓	✓
Citations	✓	✓	✓
R-squared	0.080	0.069	0.061
Observations	2,482	2,482	2,482

Sources: *Health Economics* and *Journal of Health Economics*. This table shows OLS estimates. The sample is restricted to treated journals and 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. Authors' characteristics include the ratio of female authors, the ratio of authors at top institutions, the ratio of tenured (or associate professor) authors and the ratio of authors who published in a top economics journal. Article characteristics include dummies for the presence of a theoretical model, single-authored journal articles and the use of stars or bold printing.

From: hejournal@york.ac.uk

To:

CC:

BCC:

Subject: Manuscript ID XXXXXXXXX now in your Reviewer Center - Health Economics

Body:

Dear XXXXXXXX,

Thank you for using the email link to agree to review Manuscript ID XXXXXXXXX entitled "XXXXXXX" for Health Economics. I ask that you please try your best to complete your review within the next 4 weeks.

In your review, please discuss the originality, accuracy and completeness of the work. I also invite your suggestions for condensing or amplifying the text. If the manuscript involves empirical analysis, we would ask you to become familiar with the Health Economics policy on so-called negative findings <http://onlinelibrary.wiley.com/enhanced/doi/10.1002/hec.3172> On the review page, there is a space for "Comments to Editor" and a space for "Comments to the Author". Please be sure to put your comments to the author in the appropriate space.

For fast-track access to the manuscript, you may click on the link below (which will take you right to the manuscript and review scoresheet).

https://mc.manuscriptcentral.com/hec?URL_MASK=XXXXXXX

Please note: It is essential that you click the "Save" button if you wish to exit the scoresheet before you submit it to the Editor. Otherwise, none of the information that you have entered will be saved in the system. When you have completed your review and are ready to submit it to the Editor, click on "Submit".

All communications regarding this manuscript are privileged. Any conflict of interest, suspicion of duplicate publication, fabrication of data or plagiarism must immediately be reported to me.

Thank you for evaluating this manuscript.

Sincerely,