Discussion Paper Series

IZA DP No. 18534 April 2026

Fact-Checking Politicians

Andrea Mattozzi
University of Bologna, CEPR and RCEA

Samuel Nocito
Sapienza University of Rome, CESifo
and IZA@LISER

Francesco Sobbrio
Tor Vergata University of Rome
and CESifo

IZA@LISER NETWORK
ISSN: 2365-9793

Fact-Checking Politicians*

Abstract

We study how politicians respond to the fact-checking of their public statements. Our research design employs a difference-in-differences approach, complemented by a randomized field intervention conducted in collaboration with a leading fact-checking organization. We find that fact-checking discourages politicians from making factually incorrect statements, with effects lasting several weeks. At the same time, we show that fact-checking neither increases nor displaces correct statements. Politicians who are fact-checked tend to substitute incorrect statements with either no statements or unverifiable ones, suggesting that they may also respond by increasing the "ambiguity" of their language to avoid public scrutiny.

JEL classification

D72, D78, D8, D91

Keywords

fact-checking, politicians, accountability, verifiability

Corresponding author

Samuel Nocito
samuel.nocito@uniroma1.it

* We thank Pietro Biroli, Giulia Caprini, Sergei Guriev, Nicola Mastrorocco, Pietro Ortoleva, Franco Scoglio, Erik Snowberg, Francesco Trebbi, Alex Yarkin, Ekaterina Zhuravskaya and audience seminars at the 2024 CESifo Economics of Social Media Workshop, 7th ETH Workshop on Democracy, 2023 PSE-CEPR Policy Forum, 2022 Barcelona Summer Forum (PE), 2022 Petralia Workshop, Poleconuk Webinar, Catholic University of Milan, Nazarbayev University, Tilburg University, EIEF, Collegio Carlo Alberto, University of Padova, Universitat Complutense de Madrid, Tor Vergata University of Rome, and Sapienza University of Rome for useful comments. We are grateful to Carlo Canepa, Tommaso Canetta and Giovanni Zagni of Pagella Politica, Valerio Riavez of Electica, and Daria Moretti and Bartolomeo Sciaccaluga. Simona Mandile and Giuseppe Spataro provided excellent research assistance. The study was approved by the Institutional Review Board of Luiss University (June 5, 2020) and it was partially financed thanks to a grant from the Research Council of the European University Institute and by the 2022 PRIN "Fact-checking Politicians" (MUR 2022ZR5FJ8) and by the 2023 EMIF "FacTS" Grant (287894). The field intervention is registered in the American Economic Association Registry for randomized control trials: Mattozzi, Andrea, Samuel Nocito and Francesco Sobbrio. 2022. "Fact-Checking Politicians." AEA RCT Registry. January 28. https://doi.org/10.1257/rct.6432-1.0

1 Introduction

The use of false or unsubstantiated claims by politicians is hardly new. Otto Von Bismarck contended that "people never lie so much as after a hunt, during a war, or before an election," and a century later, Ronald Reagan ventured to claim that "trees cause more pollution than automobiles do." It has only been in the last decade, however, that expressions such as fake news, alternative facts, or post-truth politics have become recurrent in public discourse, largely due to a rise in inaccurate or blatantly false public statements (Nyhan, 2020). During the same period, a number of independent organizations committed to verifying the factual accuracy of public statements emerged worldwide.[1] What remains a matter of ongoing debate is whether fact-checking is indeed an effective tool to curb fake news and misleading statements that could undermine the accuracy of individuals' beliefs and, in turn, the efficiency of democracy.[2] While there is recent experimental evidence on individual response to fact-checking (Swire *et al.*, 2017; Nyhan *et al.*, 2020; Henry *et al.*, 2022; Guriev *et al.*, 2023), little is known on the impact of real-world fact-checking on politicians.

This paper provides the first policy evaluation of the impact of the daily activities of independent fact-checking organizations on the supply of misinformation by elected politicians. Specifically, we estimate how Italian MPs respond to being exposed to a negative fact-checking by the leading Italian fact-checking organization (*Pagella Politica*). Our empirical design relies on a staggered difference-in-differences (DiD) analysis comparing politicians exposed to negative fact-checking after versus before fact-checking takes place, with respect to politicians that are not or not-yet fact-checked.

A key concern of such a DiD analysis is the endogeneity of the fact-checking itself. While independent fact-checking organizations adhere to rigorous standards in the verification of news

[1]In 2016, there were more than 100 independent fact-checking organizations in more than 50 countries. The 90% of these organizations have been established after 2010 and more than half are not affiliated with media companies (Graves and Cherubini, 2016). Fact-checkers are trained to acquire specific skills allowing them to judge the quality of information quickly and accurately (Wineburg and McGrew, 2017), and tend to outperform crowd-sourced evaluations of news stories (Godel *et al.*, 2021).

[2]DellaVigna and Gentzkow (2010): "The efficiency of market economies and democratic political systems depends on the accuracy of individuals' beliefs." *Annual Review of Economics*, 2(1), p. 644. For a recent survey on the political economy of internet and social media see Zhuravskaya *et al.* (2020).

and public statements they choose to fact-check (IFCN, 2025), their selection is not random.[3] Accordingly, DiD estimates alone are unlikely to recover causal effects: the endogenous selection process by fact-checkers introduces a potential bias when simply considering the timing and the sample of fact-checked statements at face value. For example, assessing the impact of fact-checking on a highly newsworthy statement might lead to an upward bias in the estimates. Conversely, to avoid appearing politically biased, fact-checking organizations may sometimes abstain from repeatedly fact-check politicians from a given party or they may decide to focus on less contentious and less newsworthy claims, which could result in a downward bias in the estimated effects.

To address this issue, our empirical design complements the DiD with a randomized field intervention in collaboration with *Pagella Politica*, which eliminates the endogeneity in the selection of political statements in terms of whether, who, what, and when to fact-check. The protocol we follow in any given week of the intervention period is as follows. We fix a pool of politicians at the beginning of our sample period. Each week, the fact-checking agency commits to fact-check one (and only one) of the politicians in this pool. We randomly select without replacement one politician belonging to the initial pool who made an incorrect public statement the week before. Then, we randomly select a statement among the incorrect ones made in the week before by the randomly selected politician. *Pagella Politica* rigorously fact-checks the selected statement. As per *Pagella Politica*'s usual practice, the verdict along with a link to the background analysis is published on the fact-checking organization's website and social media pages. Importantly, the Tweets promoting the fact-check tag and mention the politician's Twitter account to ensure that each treated politician is aware of the fact-checking. We standardize the timing of the fact-checking publication and related social media posts such that each fact-checking is published on a Monday morning of the week following the one when the fact-checked politician made an incorrect statement. We complement the public fact-checking with an informational campaign

[3]The *Washington Post* states that "We especially try to examine statements that are newsworthy or concern issues of importance. [...] We strive to be dispassionate and non-partisan, drawing attention to inaccurate statements on both left and right." See https://www.washingtonpost.com/politics/2019/01/07/about-fact-checker/. Similar criteria are applied by the Italian fact-checking agency *Pagella Politica*, which selects politicians statements "on the basis of their relevance and resonance in the media and in the political debate, always trying to avoid an undue concentration of our articles on a single politician or political party", translation from https://pagellapolitica.it/progetto.

targeting all politicians, to isolate the direct impact of experiencing a negative fact-checking from a simple awareness or salience effect (Nyhan and Reifler, 2015a; Avis *et al.*, 2018). Notice that our randomization protocol only affects—"behind the curtains"—the decision of whether, who, what, and when to fact-check. Aside from this, our intervention does not alter the *modus operandi* of the fact-checking organization. As such, the experimenter demand effect is mechanically absent in our context.

We analyze the universe of political statements publicly released by a sample of 55 Italian MPs over a period of sixteen weeks. The sample encompasses 82% of all mid-rank Italian MPs with an active Twitter account present in the registry of media-exposed politicians provided by the Italian Communication Authority AGCOM (and 65% of all mid-rank MPs present in the registry). We show that politicians respond to negative fact-checking by significantly reducing the number of incorrect statements made in the weeks following the fact-checking, in the order of more than a quarter of a standard deviation. Notably, the effects of the treatment are not short-lived, as they persist for as long as eight weeks.

In addition to robustness checks and event-study specifications, we validate the causal interpretation of our results using randomization inference p-values, following Young (2019) and Dell and Olken (2020). This approach compares the observed estimates to a distribution of placebo effects obtained by randomly reassigning the treatment, also providing a non-parametric robustness check against finite sample bias. Finally, we show that the reduction in incorrect statements is partially explained by a reduction in the total number of claims, as well as an increase in the proportion of *unverifiable statements*. Politicians, when exposed to negative fact-checking, also respond by resorting to non-factual claims or vague political rhetoric, thus increasing the ambiguity of their statements. At the same time, we find that fact-checking neither increases nor displaces correct statements. Importantly, the observed response of politicians—both in terms of the reduction in incorrect statements and changes in other dimensions of political communication—remains externally valid when examined across alternative samples, including a pilot study (alone or combined with the main sample) and an observational study.

In order to uncover why politicians respond to fact-checking, we explore a number of potential channels: salience, politicians making unintentional mistakes, politicians revising their beliefs regarding the probability of being fact-checked in the future, and career concerns inside and outside the party. We argue that the most plausible narrative that could account for the observed behavior is that politicians might face convex costs from being repeatedly exposed to negative fact-checking. This could be due to voters becoming progressively less tolerant of politicians who repeatedly make incorrect statements. It may also be driven by career concerns beyond direct electoral accountability, such as prospects of appointments to higher offices or high-profile positions in the private sector requiring a certain level of perceived competence. More generally, it may be related to concerns about the self-image of politicians (Bursztyn and Jensen, 2017; Abeler *et al.*, 2019).

The existing literature on fact-checking has primarily examined its impact on citizens through survey experiments that randomly expose individuals to fact-checking (Swire *et al.*, 2017; Barrera *et al.*, 2020; Nyhan *et al.*, 2020; Henry *et al.*, 2022; Guriev *et al.*, 2023). These studies yield mixed evidence regarding the effectiveness of fact-checking in shaping citizens' beliefs and attitudes toward politicians who spread misinformation. Conversely, Nyhan and Reifler (2015a) investigate the impact of fact-checking on politicians, employing a randomized informational campaign targeting state legislators across nine U.S. states. Their study uses mail-based interventions designed to increase legislators' awareness of fact-checking activities, finding that treated politicians exhibit a lower likelihood of having their statements questioned in news outlets. However, they do not detect a statistically significant effect on the probability of having the accuracy of a statement questioned by fact-checkers. Importantly, their multifaceted informational intervention complicates attributing outcomes exclusively to politicians' responses to fact-checking itself.[4]

Our study contributes to the literature along several dimensions. First, departing from most existing studies, we investigate the impact of fact-checking on one of the primary direct sources

[4]The treatment letter informs politicians of the presence of a fact-checking agency in their state; that the authors are working on a research project aimed at studying how politicians in their state respond to fact-checking; that "politicians who lie put their reputations and careers at risk, but only when those lies are exposed" (Nyhan and Reifler 2015b, page 3). Furthermore, the treatment letter implicitly reminds politicians of the possible consequences of being exposed to negative fact-checking in terms of reputational and electoral prospects and of negative advertising more generally. The placebo (Hawthorne) letter instead informs politicians that the authors are working on a project studying the accuracy of legislators' statements in that state.

of misinformation: politicians.[5] Second, we provide evidence on the effects of *real-world, publicly observable* fact-checking within consolidated democracies (Tucker *et al.*, 2018).[6] This is relevant in terms of external validity, as publicly observed real-world fact-checking may lead others—voters, party hierarchies, or institutions—to update their beliefs about the treated politicians in ways that may alter those politicians' incentives (Ashworth and De Mesquita, 2014). Third, our study examines the impact of real-world fact-checking as conducted by *Pagella Politica*, which is the *only* organization in Italy that systematically fact-checks statements from *all* politicians—not just party leaders or high-profile politicians. This allows us to capture potential general equilibrium effects. Fourth, by complementing public fact-checking with an informational campaign targeting all politicians in our sample, we are better able to isolate the impact of experiencing a negative fact-checking from a simple awareness or salience effect (Nyhan and Reifler, 2015a; Avis *et al.*, 2018). Finally, by considering the universe of politicians' statements, we examine the impact of fact-checking on both the number and verifiability of all statements, documenting a nuanced substitution pattern in which politicians also respond to negative fact-checking by increasing the ambiguity of their language.

Before we develop our analysis, an important remark is in order. There are two conceivable counterfactuals when thinking about the potential impact of fact-checking. The first would require comparing a world where politicians face the risk of incurring in a negative fact-checking with one where they do not. Assessing this type of counterfactual is unfeasible, as all politicians face the risk of being fact-checked. The second counterfactual involves comparing the behavior of politicians after being fact-checked with the behavior of those who were not fact-checked but could be fact-checked in the future. This counterfactual is the one that is observed in reality and that we focus on. Accordingly, our estimates are informative of the intensive margin of fact-

[5] Regarding the role of fact-checking in mitigating misinformation, our findings align with Henry *et al.* (2022), who document that fact-checking effectively reduces misinformation sharing on social media. Both studies underscore how fact-checking diminishes incentives to spread misinformation, directly or indirectly. The literature focusing on the demand side also show that audiences often value ideological congruence over accuracy and display mixed willingness to consume fact-checked content, especially when sources are aligned (Chopra *et al.*, 2022, 2024). In this respect, our findings point out that public fact-checking still disciplines the supply of false claims by politicians, albeit with partial substitution into unverifiable rhetoric.

[6] For evidence highlighting the nuanced impacts of transparency initiatives in less institutionalized contexts, see Malesky *et al.* (2012); Anderson (2013); Fergusson *et al.* (2013); Bowles *et al.* (2023).

checking activities, that is, the effect of being exposed to a negative fact-checking contingent on fact-checking activities existing, rather than of the extensive margin, that is, the impact of the existence of a fact-checking organization per se.

2 Design of Field Intervention

2.1 Partnership with Fact-Checking Organization

Our field intervention was implemented in partnership with *Pagella Politica*: the first and most important Italian fact-checking organization. *Pagella Politica* has been active since 2012, and is the only Italian organization entirely dedicated to political fact-checking. Along with its sister company (*Facta*), it is the only independent fact-checking organization in Italy belonging to the International Fact-Checking Network (IFCN) and one of the signatories of the related Code of Principles.[7] Furthermore, *Pagella Politica* is the only fact-checking organization in Italy specialized in verifying the statements of politicians. None of the founders of *Pagella Politica* or staffers are members of political parties, political organizations, and entities related to political parties, and the lack of political involvement is a key prerequisite in order to work or cooperate with them.[8]

The usual business activity of *Pagella Politica* consists of monitoring political statements from traditional and online media, social media, and news agencies. Clearly, it only focuses on *verifiable* statements, that is based on verifiable facts or numbers. While a number of news media and websites sometimes provide fact-checking on selected statements by political leaders or politicians in key government positions, *Pagella Politica* is, effectively, the monopolist supplier of fact-checking on mid-rank politicians.

[7] *Pagella Politica* is independent from any media organization and it is mostly financed by selling content and services to third parties (e.g., Facebook) and by participating in international projects and calls. Independence from media organizations is particularly important for our purposes. Indeed, Louis-Sidois (2025) shows that the six main French fact-checkers connected to media organizations tend to fact-check less entities that align with their ideology and are more likely to agree with them.

[8] Online Appendix B offers further details on *Pagella Politica* for the interested reader.

2.2 Time Frame and Politicians Sample

The field intervention lasted 16 weeks between March and July 2021 and comprised 3 pre-intervention weeks, 10 intervention weeks, and 3 post-intervention weeks.[9] Throughout the entire 16-week period, *Pagella Politica* committed not to publish any fact-checking involving politicians in our sample other than those randomly included in the intervention.

Our final sample is composed of 55 mid-rank politicians. The sample was drawn starting from the registry of media-exposed politicians provided by the Italian Communication Authority (AGCOM), by selecting mid-rank MPs with an active Twitter account. It encompasses 82% of all mid-rank Italian MPs with an active Twitter account present in the registry of media-exposed politicians provided by AGCOM, and 65% of all mid-rank MPs present in the registry.[10]

2.3 Politicians' Statements

We collected the universe of politicians' statements provided by the main Italian news agencies (ANSA; AGI; Adnkronos; Askanews). The focus on news agencies allows us to avoid issues of biased coverage as their core business is precisely to monitor any possible public statements made by politicians on any type of news media and on politicians' social media accounts.

During the intervention weeks, we pre-screened verifiable or fact-checkable (FC, henceforth) statements using the following procedure (the interested reader may find a detailed explanation of the entire data-generating process in Online Appendix C). First, we identify FC statements by using a supervised machine-learning classifier. Then, *Pagella Politica* identifies the incorrect statements within the set of FC statements. Next, we randomly draw one politician from those who made at least one incorrect statement in the previous week, and we randomly select one incorrect statement made by the selected politician. We adopt this two-step procedure to avoid oversampling politicians who make more verifiable but incorrect statements in a week. Politicians

[9] Online Appendix A discusses the advantages of our intervention with respect to alternative designs. Online Appendix G—following the guidelines of Banerjee *et al.* (2020)—presents a report detailing the ("populated") pre-analysis plan (PAP) that we submitted prior to the start of our field intervention, along with any deviations from the PAP.

[10] See Online Appendix B for further details on the sample. Table B.1, in online Appendix B, shows that the sample is balanced across the six main political parties present in the Italian parliament at the time of our field intervention.

are randomly drawn without replacement. That is, once a politician is fact-checked at time t, she is not part of the randomization pool of the following weeks. Once the statement is selected, *Pagella Politica* produces a fact-checking according to its standard rules. Starting from our sample of 55 politicians, our procedure leads to 10 treated politicians (staggered over time) and 45 never-treated politicians.

Table 1 reports descriptive statistics of individual characteristics for treated and control politicians in the pre-treatment period. Observable characteristics are broadly balanced across groups, notably also for the number of verifiable statements and the number of incorrect statements made before the intervention. The only exceptions concern *right-wing* and *populist* politicians, who are over-represented among treated units.[11] This imbalance is not a concern for identification, since it does not require balance in politicians' pre-treatment characteristics as discussed in Section 3.

2.4 Fact-checking Campaign

Pagella Politica publishes the fact-checking on its website. Furthermore, to ensure that the politician is aware of being exposed to fact-checking, *Pagella Politica* simultaneously sends two Tweets from its official account—with a link to the fact-checking—mentioning the politician's Twitter account.[12] Finally, we advertise a video featuring the fact-checking on a number of popular websites and social media.[13] This advertising campaign was aimed at informing politicians in the control group about the presence of the fact-checking. To increase its effectiveness, the campaign was geo-targeted in two zip-codes (00186 and 00187) around the Italian Parliament. For each fact-checked politician the campaign started right after the publication of the fact-checking on the *Pagella Politica* website (Monday afternoon) and lasted five days, i.e. until the Friday evening of the week when the fact-checking was published.

[11] These two categories substantially overlap in the Italian context: Rooduijn *et al.* (2024) classifies all right-wing Italian parties as populist (i.e., the list of populist parties encompasses all right-wing parties plus the Five Stars Movement). This imbalance reflects that a necessary condition to enter the treated group is having made a verifiable and incorrect statement during the intervention weeks; during this period—and, in line with the focus on populist politicians of Swire *et al.* (2017); Barrera *et al.* (2020); Nyhan *et al.* (2020); Henry *et al.* (2022)—populist politicians were more likely to do so than non-populists.

[12] Figures B.1 and B.2, in Online Appendix B, report examples of the fact-checking web page and the fact-checking Tweets.

[13] An English version of the video can be found at the following link: `https://tinyurl.com/283yut75`.

3 Empirical Strategy

The design of our field intervention involves a staggered treatment of politicians: the first treated politician is fact-checked in intervention week one, the second in week two, and so on. Given that our design involves ten sequential treatments, and the two-way fixed effects (TWFE) estimator may be biased under staggered treatment timing and heterogeneous effects, we adopt the stacked difference-in-differences approach (CDLZ, hereafter) proposed by Cengiz *et al.* (2019, 2022). Specifically, we create ten event-specific datasets (h) by using only within-event variation between the treated unit and clean control units. Each of the ten event-specific datasets is composed of one treated politician and the corresponding "clean" controls, that is, never-treated and not-yet-treated politicians. We then stack such event datasets and estimate:

$$Y_{h,i,t} = \beta D_{h,i,t} + \delta_{h,i} + \delta_{h,t} + \varepsilon_{h,i,t} \tag{1}$$

where $Y_{h,i,t}$ is the observed outcome of politician i (e.g., number of incorrect statements) at time to event t in the event-level h. Furthermore, $D_{h,i,t} = \mathbb{1}\{t \geq G_{h,i}\}$, where $G_{h,i}$ is the time when politician i is fact-checked in the event-level h. Finally, $\delta_{h,i}$ and $\delta_{h,t}$ represent politician-event and time-event fixed effects, respectively. We follow Cengiz *et al.* (2022) and Butters *et al.* (2022) and cluster standard errors by politician, which is the level at which the treatment is assigned. This accounts for the usual concern of serially correlated residuals by assuming that all observations from the same politician i may be dependent, even if they appear in different event-levels h.[14] The estimator is robust to heterogeneous treatment effects and it does not suffer from negative weighting bias since it rules out the potentially dangerous group comparisons that use "already-treated" units as controls (De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). Indeed, as shown by Gardner (2022), CDLZ estimates a convex weighted average of the group-time average treatment effects on the treated ($WATT_{g,t}$) under parallel trends and no anticipation.[15]

[14] An alternative approach would be to cluster standard errors at the politician-event level as in Cengiz *et al.* (2019), which we present in Column (2) of Table 2. Yet, when one cluster is fully nested within another, clustering at the higher level yields more conservative standard errors (Cameron *et al.*, 2011). Nevertheless, the two approaches produce nearly identical results (Wing *et al.*, 2024).

[15] Since we have one treated unit for each stacked event, the CDLZ estimator almost equally weights each stacked event. Generally, stacked events are equally weighted in frameworks with a homogeneous time-to-event

Consistent with this identifying assumption, Figure 1 in Section 4 documents parallel pre-treatment trends in outcomes.[16] We also assess the sensitivity of our estimates to potential deviations from the parallel-trends condition using the honest DiD bounds outlined by Rambachan and Roth (2023).[17]

4 Results

Table 2 presents our main results, where the dependent variable in Equation (1) is the number of incorrect statements. At the bottom, we also report summary statistics for treated politicians in the three pre-intervention weeks (i.e., before treatment), which serve as a benchmark to compare the estimated effects. As long as the set of treated politicians is not altered, these summary statistics remain unchanged across different sample restrictions, ensuring consistency in comparisons. Column (1) shows that fact-checking leads to a reduction in incorrect statements made by politicians, amounting to a quarter of a standard deviation. Column (2) presents estimates when clustering standard errors at the politician-event level, rather than the politician level as in the baseline specification. Column (3) provides estimates when using a simple two-way-fixed effects model, i.e., using a panel specification with time and politician fixed effects. Finally, in Column (4), as the number of incorrect statements is a count variable, we estimate a fixed effect Poisson model (Correia *et al.*, 2019) on the stacked-events sample.[18]

Estimates of the Equation (1) rely on the parallel trends assumption between treated and

window across stacked events, only never-treated as controls, and having the same number of treated units for each treatment group. Numerical calculations of weights according to Gardner (2022) are available upon request. We use CDLZ for its flexibility in accommodating our empirical analyses. In Column (8) of Table 3, as a robustness check, we also estimate the $ATT_{g,t}$ using the imputation-based DiD estimator proposed by Borusyak *et al.* (2024).

[16] Conversely, as Equation (1) also accounts for time-invariant differences across politicians, identification does not hinge on balance in pre-treatment characteristics. Nonetheless, to further reassure that the over-representation of populist/right-wing politicians in the randomization sample (Table 1) does not play a role in our results, we conduct two targeted robustness checks. First, we allow for heterogeneous time effects by interacting week fixed effects with observable characteristics, including right-wing and populist indicators (Table 3, Col. 7). Second, we estimate the models restricting the sample to right-wing or to populist politicians (Online Appendix Table F.1).

[17] In particular, we report results after allowing the slope of the pre-trend to change by an amount M across consecutive periods.

[18] The number of observations is lower than that of Column (1) as the fixed effect Poisson model is estimated excluding observations that are either singletons or separated by fixed effects (Correia *et al.*, 2019). The point estimate of Column (4) has to be interpreted as in a log-linear model; given the dichotomic nature of the treatment variable, it represents a semi-elasticity.

control politicians. Accordingly, Panel A of Figure 1 presents the results of an event-study specification estimating leads and lags from/to the fact-checking event. Formally, we estimate the equation:

$$Y_{h,i,t} = \sum_{j=a}^{b} \beta_j D^j_{h,i,t} + \delta_{h,i} + \delta_{h,t} + \varepsilon_{h,i,t} \tag{2}$$

where $D^j_{h,i,t}$ are leads and lags of the treatment variable defined in Equation (1) with respect to the time of the fact-checking in the event-level h.[19] The graph shows the absence of significative pre-trends in the number of incorrect statements by fact-checked politicians. It also points out that the reduction in the number of incorrect statements induced by fact-checking is not short-lived. To further assess the robustness of the parallel trend assumption, in Panel B of Figure 1, we employ the "honest approach" to parallel trends outlined by Rambachan and Roth (2023), which provides bounds on the estimates under alternative assumptions about how the outcome would have evolved for treated and control politicians without fact-checking. When considering the average effects through post fact-checking event-study coefficients, we find that the effect remains negative and statistically significant even when considering wide range of deviations from the assumption of a linear pre-trend.[20]

Table 3 presents a series of exercises to further corroborate the robustness of our baseline results.[21] Column (1) reports estimates when adding an interaction term between a dummy capturing whether a politician belongs to the randomization pool of a given event dataset and time-to-event fixed effects. In this way, we can isolate the impact of fact-checking within the pool of randomizable politicians within each event dataset and within each week to/from the fact-checking event. The estimates in Column (1) show that even focusing only on the effects within the randomization sample, we still detect a significant albeit smaller effect in the order of one-

[19]Given the time structure of our experimental design, we fix $a = -6$ and $b = 8$. We follow McCrary (2008) and bind up end-points, results are similar for alternative windows and are available upon request.

[20]According to Rambachan and Roth (2023), the parameter $M \geq 0$ governs the extent to which the slope of the treatment effect can change between consecutive periods, effectively bounding the discrete analogue of its second derivative. $M = 0$ corresponds to cases where differences between treated and control are exactly linear, while larger values of M allow for larger deviations from linearity.

[21]We report summary statistics for treated politicians in the three pre-intervention weeks, and as long as the set of treated politicians is not altered, these summary statistics remain unchanged across different sample restrictions, ensuring consistency in comparisons.

fifth of a standard deviation. Notice that, differently from our baseline model, the specification in Column (1) compares differences in behavior between fact-checked politicians and politicians who made an incorrect statement in the pre-treatment week, that is, it does not consider politicians who did not make any incorrect statement in the pre-treatment week even though they may make incorrect statements in the following weeks. Column (2) shows that the estimates of Column (1) are robust when dropping from each event dataset all politicians not in the randomization sample.[22] Column (3) reports estimates when restricting the control group to never treated politicians only. In Column (4), we present estimates when restricting to a homogeneous time window across event-datasets ($-3/+3$ weeks from the fact-checking event) as well as including only never treated in the control group.[23] This specification imposes a rather drastic reduction in the estimation window, as expected, it leads to less precise estimates. Column (5), in the spirit of Sun and Abraham (2021), uses only the last treated unit as a control group; in this specification, each stacked event h is composed of the actually treated unit and the last treated as control with time truncated before the last treatment. In Column (6), we exclude from the control group politicians who do not make any incorrect statements over the sample period, thus focusing on a more homogeneous sample. In Column (7), we augment Equation (1) by also including week fixed-effects interacted with the baseline politicians' characteristics. Finally, Column (8) presents estimates when implementing the imputation-based difference-in-differences estimator proposed by Borusyak *et al.* (2024).

Online Appendix E provides evidence on the external validity of the baseline results. The negative impact of fact-checking on the number of incorrect statements is also present when looking at alternative samples involving a pilot study (also in combination with our main sample), as well as an observational study. Altogether, the baseline results are robust across three datasets (main sample, pilot, and observational studies) which span a total of 86 politicians, 20 of whom were exposed to negative fact-checking, 62,918 total statements by politicians, of which 2,257 verifiable

[22] These coefficients are estimated without taking into account the tenth intervention week since in this week there were no politicians in the randomization pool but the actual treated.

[23] This analysis is equivalent to using a weighted stacked DiD estimator (Wing *et al.*, 2024), in which each sub-event dataset is equally weighted, since the stacked dataset is characterized by a constant share of treated units across sub-events.

ones, including 448 incorrect statements, and a time horizon of 56 weeks. Notably, the results of these different samples are also consistent across all dimensions of political communication studied in Section 5.1.

4.1 Randomization Inference Tests

To further assess the internal validity of our estimates, we implement Randomization Inference Tests (RIT), which estimate how extreme the observed effect is relative to a distribution of placebo effects generated by reassigning the treatment. This also provides a non-parametric robustness check against finite sample bias, relying solely on the randomization mechanism rather than asymptotic inference (Young, 2019; Dell and Olken, 2020). Specifically, we implement a design-based RIT by exploiting the fact that in each intervention week, there are politicians in the randomization pool who could have been subject to fact-checking in that week, but they were not. Accordingly, we estimate placebo counterfactuals by randomly and fictitiously attributing the treatment in week t to a politician in the randomization pool of that week excluding the actually treated politician.[24] These counterfactual estimates allow us to compute a randomization inference (RI) p-value, defined as the proportion of placebo coefficients that are more extreme than the observed baseline coefficient. Formally, $p = \frac{\sum_{k=1}^{N} \mathbb{1}\{\beta_k \leq \hat{\beta}_{obs}\}}{N}$ with $N = 1{,}100$, where β_k denotes the coefficients obtained through permutations, and $\hat{\beta}_{obs}$ is the observed baseline coefficient.[25] Figure 2 plots the distribution of estimated placebo coefficients and displays the observed baseline coefficient—used as a benchmark—as a red vertical line. It also reports the corresponding RI p-value. The graph shows that politicians in the randomization pool also reduce their number of incorrect statements, although to a much lower extent with respect to the actually treated ones. This suggests that even if a reversion-to-the-mean effect was at play among politicians making an incorrect statement in a week, it is unlikely to be the main driver of the observed result. Indeed, the estimated benchmark coefficient is more than twice as large as the average placebo coefficient,

[24] Online Appendix C describes in detail the process used to generate the counterfactuals for the RIT.

[25] With respect to Dell and Olken (2020) we adopt a randomization without replacement since we are interested in estimating all possible combinations of placebo treatments across treatment weeks and politicians entering the randomization pool. When the reference distribution is constructed from the complete set of all possible random assignments, the resulting p-values are exact and do not rely on theoretical approximations or assumptions about the shape of the sampling distribution (Gerber and Green, 2012).

and none of the permutation coefficients exceeds it in magnitude. The RI p-value also supports the same statistical inference as in the baseline results reported in Table 2 (Columns (1) and (2)), thereby addressing potential concerns about the presence of finite sample bias in the estimates.

5 Mechanism

We analyze the mechanism behind our main results by looking at: a) whether fact-checking impacts other dimensions of political communication; b) alternative narratives that may rationalize the observed differential response between treated and control politicians.

5.1 Substitution Patterns in Political Communication

Table 4 provides an overview of the impact of fact-checking on political communication in terms of incorrect, correct, and verifiable statements, as well as the overall number of statements. Columns (1) and (2) show that fact-checking leads to a decrease both in the intensive and extensive margin of incorrect statements.[26] Furthermore, Columns (3) and (4) suggest that there is no discernible impact of fact-checking on correct statements, both in terms of statistical significance and magnitude of the point estimates. Column (5) reports estimates on the number of verifiable statements. Mechanically, this latter point estimate is equal to the sum of the estimates in Columns (1) and (3), and since the variance in the number of verifiable statements is mostly driven by the variance in the number of *correct* statements, the estimate in Column (5) is also not statistically significant. As the number of correct statements is not affected by fact-checking, it follows that the observed reduction in the number of incorrect statements translates into a reduction of verifiable statements.[27] Furthermore, Column (7) shows that politicians respond to fact-checking by also reducing their overall number of statements, in the order of one-seventh of a standard deviation.[28] All in all, these results suggest that politicians respond to negative fact-

[26] Note that Column (1) of Table 4 simply replicates for ease of comparison the baseline results of Table 2 on the number of incorrect statements.

[27] Accordingly, we also see a marginally significant reduction in the probability of fact-checked politicians making any verifiable statements in a week (Column (6)).

[28] Table D.2, in online Appendix D, shows that this pattern is also present when considering the proportions of incorrect, correct, and unverifiable statements made in a given week.

checking by adopting a more ambiguous language: a strategy to avoid potential public scrutiny.

5.2 Fact-checked Politicians vs. Control Group

We now discuss different possible narratives that may rationalize the observed differential response between treated and control politicians.

First, it may be the case that fact-checking operates through a simple information channel: making politicians aware of being potentially fact-checked or increasing the salience of fact-checking (Avis *et al.*, 2018). As discussed in Section 2.4, we complement the fact-checking with an informational campaign by advertising the video of each fact-checking on various popular websites and social media, geo-targeting the two zip codes surrounding the Italian Parliament. The video advertising campaign is explicitly aimed at informing politicians in the control group about the presence of the fact-checking activity. That is, the informational campaign was purposely designed to bring the salience of fact-checking to a comparable level across politicians in both the treatment and control groups. Furthermore, we do not detect any pattern suggesting that politicians respond more when the engagement with the informational campaign is stronger.[29] To this end, we can exploit the heterogeneity present across treated politicians in terms of engagement with the video ads.

Figure 3 illustrates how the response of politicians varies when interacting our treatment variable with a dummy indicating whether a video advertising the fact-checking - relative to a given politician - is above or below the median with respect to different measures of engagement. The results do not suggest a clear pattern. While we observe a stronger response in terms of unique impressions, this is not the case when looking at other measures of engagement such as clicks on the video or total video completed. Indeed, for these two measures, we see that politicians at the lower end of the distribution in terms of video completed or clicks seem to be the ones with a more significant response and also a larger one in terms of the relative magnitude of the effects. Furthermore, the salience effect should be conceivably stronger among politicians

[29]This question is relevant both to shed light on the mechanism as well as in terms of policy evaluation, as the video advertising campaign is the only point of departure from how *Pagella Politica* normally advertises its contents.

"familiar" with a treated one, as they are more likely to be aware of the fact-checking.

In order to test whether there is any evidence of potential spillover effects, for each event-dataset we drop the fact-checked politician and attribute the treatment to politicians in the control group who are "familiar" with the treated ones. That is, we run alternative specifications fictionally attributing the treatment to politicians belonging either to the same party of the fact-checked politicians or to the same party-chamber or sitting in the same parliamentary commission, including never-treated and not-yet treated politicians in the control group. Table 5 reports the estimates of such exercise. The results show no evidence of potential spillover effects across party-peers or party-chamber peers, or parliamentary peers sitting in the same commission.

Overall, taking into account the informational campaign and the absence of spillover effects, our findings offer complementary evidence to the existing literature (Nyhan and Reifler, 2015a; Avis *et al.*, 2018), indicating that the impact of fact-checking does not seem to merely being driven by a salience effect.

A second possible explanation for our findings is that politicians may not intentionally lie but simply make mistakes. This would imply that negative fact-checking allows them to actually learn that their statement was incorrect. Accordingly, they might respond by putting a higher effort in the future to make sure their statements are not incorrect. Yet, this "benevolent" narrative seems at odds with our evidence showing that also politicians who had received a negative fact-checking in the last year have a statistically significant response to the fact-checking (and, if anything, such response is larger in terms of magnitude). Indeed, as shown in Figure 4, we observe a slightly stronger statistical significance and a larger effect for politicians who received a negative fact-checking in the year before the start of the field intervention (0.36 of a standard deviation) relative to those who did not (0.22 of a standard deviation). Furthermore, this potential explanation does not seem to square with the evidence, discussed in Section 5.1, showing that fact-checking does not seem to affect the number of correct statements made by politicians.

A third potential narrative is that treated politicians revise upward their belief about the likelihood of being fact-checked in the near future. While we cannot formally test this channel, this mechanism seems implausible, given that *Pagella Politica* seldom repeatedly fact-checks mid-rank politicians. Indeed, only 31 politicians in our sample had ever been exposed to a negative fact-checking before. Moreover, of these 31 politicians, only 11 had ever been exposed to a negative fact-checking more than once, with a median distance between a negative fact-checking and the subsequent one of 98 days. Hence, if anything, politicians should expect that after being exposed to a negative fact-checking, it is unlikely that the fact-checking organization will target them again in the near future.

Finally, it might be the case that politicians have convex costs from being repeatedly exposed to negative fact-checking. This could be due to voters becoming progressively less forbearing with politicians repeatedly making incorrect statements. It may also be driven by career concerns beyond direct electoral accountability (e.g., moving to higher offices or high-profile positions in the private market that require some minimum level of perceived competence). Moreover, there might also be a convex cost in terms of a politician's self-image concerns (Bursztyn and Jensen, 2017; Abeler *et al.*, 2019). This narrative is—indeed—consistent with the evidence present in Figure 4.[30]

6 Conclusion

We view our results as a first step toward informing both academic and public debates on how real-world fact-checking influences politicians' behavior. Our study represents the first randomized field intervention on fact-checking conducted in a business-as-usual setting, altering only the selection process of statements fact-checked by Italy's leading fact-checking organization.

Our design combines the benefits of preserving external validity through a real-world setting, with the identification of causal effects via exogenous variation in which political statements are fact-checked. This combination, essential for evaluating the real-world effectiveness of fact-checking, naturally constrains the sample size of treated politicians. A purely observational study would offer high external validity but suffer from endogeneity, while a large-scale randomized intervention would yield a larger treated sample but disrupt fact-checkers' operations. This

[30] Online Appendix F outlines why it is unlikely that electoral incentives solely explain the observed behavior of politicians.

would not only compromise the external validity of the results, but it would also be unlikely to be accepted by any major fact-checking company. Conversely, our design allows us to stay closely aligned with the actual practices of fact-checking organizations while providing causal inference. At the same time, the size of our main sample is unlikely to undermine the external validity of the estimates, as evidenced by the robustness of results across alternative samples.

The results show that fact-checking discourages politicians from making factually incorrect statements, with effects lasting several weeks. Since fact-checking is conducted on a regular basis, with a frequency typically increasing during electoral campaigns, our results provide the first evidence that fact-checking is indeed an effective tool in reducing the level of factual misinformation in politics. At the same time, we also document some unexpected drawbacks. Fact-checking neither increases nor displaces correct statements. Instead, fact-checked politicians tend to substitute incorrect statements with either no statements or with unverifiable ones. This suggests that they also increase the ambiguity of their language to escape the possibility of public scrutiny. What can we conclude from a welfare perspective? The net impact depends on how voters respond to verifiable and unverifiable statements. As long as unverifiable statements are less persuasive than verifiable ones, the substitution pattern is likely to be beneficial from a welfare perspective. Yet, one of the pillars of representative democracy is a functioning mechanism of accountability, and this mechanism also relies on the possibility of a constant and thorough scrutiny of politicians' verifiable public statements. In this respect, constructing reliable individual-level measures of ambiguity or non-verifiability in political rhetoric seems a particularly interesting avenue for future research.

References

Abeler, J., Nosenzo, D. and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, **87** (4), 1115–1153.

Anderson, J. H. (2013). Sunshine works: Comment on "the adverse effects of sunshine: A field experiment on legislative transparency in an authoritarian assembly". *World Bank Policy Research Working Paper*, (6602).

Ashworth, S. and De Mesquita, E. B. (2014). Is voter competence good for voters?: Information, rationality, and democratic performance. *American Political Science Review*, **108** (3), 565–587.

Avis, E., Ferraz, C. and Finan, F. (2018). Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *Journal of Political Economy*, **126** (5), 1912–1964.

Banerjee, A., Duflo, E., Finkelstein, A., Katz, L. F., Olken, B. A. and Sautmann, A. (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics. *National Bureau of Economic Research*, (WP No. 26993).

Barrera, O., Guriev, S., Henry, E. and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, **182**, 104123.

Borusyak, K., Jaravel, X. and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, **91** (6), 3253–3285.

Bowles, J., Croke, K., Larreguy, H., Liu, S. and Marshall, J. (2023). Sustaining exposure to fact-checks: Misinformation discernment, media consumption, and its political implications. *American Political Science Review*, pp. 1–24.

Bursztyn, L. and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, **9**, 131–153.

Butters, R. A., Sacks, D. W. and Seo, B. (2022). How do national firms respond to local cost shocks? *American Economic Review*, **112** (5), 1737–1772.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, **29** (2), 238–249.

Cengiz, D., Dube, A., Lindner, A. and Zentler-Munro, D. (2022). Seeing beyond the trees: using machine learning to estimate the impact of minimum wages on labor market outcomes. *Journal of Labor Economics*, **40** (S1), S203–S247.

—, —, — and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, **134** (3), 1405–1454.

Chopra, F., Haaland, I. and Roth, C. (2022). Do people demand fact-checked news? Evidence from U.S. democrats. *Journal of Public Economics*, **205**, 104549.

—, — and — (2024). The demand for news: Accuracy concerns versus belief confirmation motives. *The Economic Journal*, **134** (661), 1806–1834.

Correia, S., Guimarães, P. and Zylkin, T. (2019). Verifying the existence of maximum likelihood estimates for generalized linear models. *arXiv preprint arXiv:1903.01633*.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, **110** (9), 2964–96.

Dell, M. and Olken, B. A. (2020). The development effects of the extractive colonial economy: The dutch cultivation system in java. *The Review of Economic Studies*, **87** (1), 164–203.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: empirical evidence. *Annual Review of Economics*, **2** (1), 643–669.

Fergusson, L., Vargas, J. F. and Vela, M. A. (2013). *Sunlight disinfects? Free media in weak democracies*. Documentos de Trabajo 010484, Universidad del Rosario.

Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.

Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

Godel, W., Sanderson, Z., Aslett, K., Nagler, J., Bonneau, R., Persily, N. and Tucker, J. A. (2021). Moderating with the mob: Evaluating the efficacy of real-time crowd-sourced fact-checking. *Journal of Online Trust and Safety*, **1** (1).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, **225** (2), 254–277.

Graves, L. and Cherubini, F. (2016). *The rise of fact-checking sites in Europe*. Reuters Institute for the Study of Journalism.

Guriev, S., Henry, E., Marquis, T. and Zhuravskaya, E. (2023). Curtailing False News, Amplifying Truth. *CEPR Discussion Paper No. 18650*.

Henry, E., Zhuravskaya, E. and Guriev, S. (2022). Checking and sharing alt-facts. *American Economic Journal: Economic Policy*, **14** (3), 55–86.

IFCN (2025). The commitments of the Code of Principles. `https://www.ifcncodeofprinciples.poynter.org/the-commitments`, [Online; accessed 30-January-2025].

Louis-Sidois, C. (2025). Both judge and party? investigating the political unbiasedness of fact-checkers. *Journal of the European Economic Association*, p. jvaf011.

Malesky, E., Schuler, P. and Tran, A. (2012). The adverse effects of sunshine: A field experiment on legislative transparency in an authoritarian assembly. *American Political Science Review*, **106** (4), 762–786.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, **142** (2), 698–714.

Nyhan, B. (2020). Facts and myths about misperceptions. *Journal of Economic Perspectives*, **34** (3), 220–36.

—, Porter, E., Reifler, J. and Wood, T. J. (2020). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political Behavior*, **42** (3), 939–960.

— and Reifler, J. (2015a). The effect of fact-checking on elites: A field experiment on us state legislators. *American Journal of Political Science*, **59** (3), 628–640.

— and — (2015b). Supplementary materials. The effect of fact-checking on elites: A field experiment on US state legislators. `https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fajps.12162&file=ajps12162-sup-0001-supmat.pdf`, [Online; accessed 01-July-2023].

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, p. rdad018.

Rooduijn, M., Pirro, A. L., Halikiopoulou, D., Froio, C., Van Kessel, S., De Lange, S. L., Mudde, C. and Taggart, P. (2024). The populist: A database of populist, far-left, and far-right parties using expert-informed qualitative comparative classification (eiqcc). *British Journal of Political Science*, **54** (3), 969–978.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, **225** (2), 175–199.

Swire, B., Berinsky, A. J., Lewandowsky, S. and Ecker, U. K. (2017). Processing political misinformation: comprehending the trump phenomenon. *Royal Society open science*, **4** (3), 160802.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. and Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Working Paper, Hewlett Foundation.*

Wineburg, S. and McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information.

Wing, C., Freedman, S. M. and Hollingsworth, A. (2024). Stacked difference-in-differences. *National Bureau of Economic Research*, (WP No. 32054).

Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, **134** (2), 557–598.

Zhuravskaya, E., Petrova, M. and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, **12** (1), 415–438.

Figures and Tables

Table 1: Balancing - Politicians' Characteristics: Treated vs. Control

	Control			Treated			Mean Diff.
	N	Mean	SD	N	Mean	SD	
Woman	45	0.29	0.46	10	0.20	0.42	-0.089
Age	45	51.42	9.00	10	54.60	7.68	3.178
College Graduate	45	0.80	0.40	10	0.90	0.32	0.100
Lower-chamber MP	45	0.71	0.46	10	0.50	0.53	-0.211
First Parliamentary Experience	45	0.40	0.50	10	0.40	0.52	0.000
Elected in a Single-member District	45	0.31	0.47	10	0.30	0.48	-0.011
N. of Parliamentary Commissions	45	1.67	1.00	10	2.10	1.37	0.433
Right Wing	45	0.36	0.48	10	0.70	0.48	0.344**
Populist	45	0.56	0.50	10	0.90	0.32	0.344***
Opposition	45	0.11	0.32	10	0.20	0.42	0.089
Number of incorrect statements	45	0.07	0.33	10	0.60	1.26	0.533
Number of verifiable statements	45	0.42	0.84	10	1.40	2.17	0.978
Negative Fact-check in the Last Year	45	0.38	0.49	10	0.20	0.42	-0.178

Notes: This table shows summary statistics of observable characteristics of politicians, divided into control and treatment groups. The last column reports the test of differences in means across groups. Standard errors are clustered at the politician level. *** p<0.01, ** p<0.05, * p<0.1, + p<0.15.

Table 2: Politicians' Response to Negative Fact-Checking

	(1) CDLZ Baseline	(2) CDLZ Cluster Pol-event	(3) TWFE Model	(4) Poisson Model
Fact-Checked	-0.378**	-0.378**	-0.370**	-0.901**
	(0.146)	(0.146)	(0.147)	(0.398)
Observations	8,035	8,035	880	3,715
Statistics on treated before treatment:				
Mean	0.67	0.67	0.67	0.67
SD	1.47	1.47	1.47	1.47
Politician-event FE	YES	YES	NO	YES
Time-event FE	YES	YES	NO	YES
Cluster SE at Politician-event	NO	YES	NO	NO
Politician FE	NO	NO	YES	NO
Time FE	NO	NO	YES	NO
Cluster SE at Politician	YES	NO	YES	YES

Notes: This table shows difference-in-differences estimates using as a dependent variable the number of incorrect statements made by a politician in a week. In columns (1), (2), and (4), we estimate Equation (1) on a stacked sample at the politician-time-event level. In columns (1), (3), and (4), we cluster standard errors at the politician level, while in column (2) we adopt clustering at the politician level-event. In column (4), we estimate Equation (1) using a high-dimensional fixed effects Poisson model. These regressions control for both politician-event and time-event fixed effects, where the event dimension identifies each sub-treatment in the stacked sample. In column (3), we estimate a standard two-way fixed effects (TWFE) model on a panel dataset at the politician-time level, controlling for politician and time fixed effects with standard errors clustered at the politician level. We also report outcome summary statistics for treated politicians in the first three pre-intervention weeks (i.e., before treatment). Results refer to the field intervention conducted over the period from March 22 to July 11, 2021 (16 weeks). *** $p<0.01$, ** $p<0.05$, * $p<0.1$, + $p<0.15$.

Table 3: Politicians' Response to Negative FC: Robustness

	(1) Interaction Randomizable pol. time-event FE	(2) Only randomizable politicians	(3) Only never treated	(4) Hom. Time $(-3/+3)$ (& only NT)	(5) Last treated as control	(6) Exclude pol. with no incorrect	(7) Week-FE interacted with pol. characteristics	(8) BJS Imputation DiD
Fact-checked	-0.268**	-0.268*	-0.369**	-0.287*	-0.620***	-0.377**	-0.412***	-0.430***
	(0.117)	(0.140)	(0.146)	(0.158)	(0.157)	(0.149)	(0.138)	(0.016)
Observations	7.226	452	7,360	3,220	216	3,715	8,023	880
Statistics on treated before treatment:								
Mean	0.67	0.67	0.67	0.17	0.67	0.67	0.67	0.67
SD	1.47	1.47	1.47	0.41	1.47	1.47	1.47	1.47
Politician-event FE	YES	YES	YES	YES	YES	YES	YES	YES
Time-event FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table shows difference-in-differences estimates using as a dependent variable the number of incorrect statements (grade 3 and below: half-false, mostly-false, utterly-false) made by a politician in a week. We estimate Equation (1) on a stacked sample at the politician-time-event level, controlling for politician-event and time-event fixed effects, where the event dimension identifies each sub-treatment in the stacked sample. Each column of this table refers to a specific robustness check. In Column (1), we include an interaction term between a dummy capturing whether a politician belongs to the randomization pool of a given event dataset and time-to-event fixed effects. In Column (2), we drop from each event dataset all politicians not in the randomization sample. In Column (3), we show results when using only never-treated politicians in the control group. In Column (4), we use a homogeneous time window across event-datasets $(-3/+3$ weeks from the fact-checking event) as well as including only never treated in the control group. In Column (5), we use only the last treated unit as a control group. In Column (6), we exclude from the control group politicians who do not make any incorrect statements over the sample period. In Column (7), we augment our baseline specification by also including week fixed-effects interacted with the politicians' characteristics. In Column (8), we estimate the treatment effect using the imputation-based DiD estimator proposed by Borusyak *et al.* (2024). We also report outcome summary statistics for treated politicians in the first three pre-intervention weeks (i.e., before treatment). Results refer to the field intervention conducted over the period from March 22 to July 11, 2021 (16 weeks). Standard errors in parentheses are clustered at the politician level. *** p<0.01, ** p<0.05, * p<0.1, + p<0.15.

Table 4: Politicians' Response to Negative Fact-Checking:
Incorrect, Correct, Verifiable and Overall Statements

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Incorrect Statements		Correct Statements		Verifiable Statements		Overall Statements	
	N. of	Any	N. of	Any	N. of	Any	N. of	Any
Fact-checked	-0.378**	-0.228***	-0.038	-0.039	-0.417	-0.153+	-6.526*	0.004
	(0.146)	(0.064)	(0.287)	(0.090)	(0.386)	(0.093)	(3.363)	(0.043)
Observations	8,035	8,035	8,035	8,035	8,035	8,035	8,035	8,035
Statistics on treated before treatment:								
Mean	0.67	0.33	1.07	0.50	1.73	0.63	52.17	0.90
SD	1.45	0.47	1.42	0.50	2.44	0.48	49.54	0.30
Politician-event FE	YES	YES	YES	YES	YES	YES	YES	YES
Time-event FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table shows difference-in-differences estimates of Equation (1) on a stacked sample at the politician-time-event level, controlling for politician-event and time-event fixed effects, where the event dimension identifies each sub-treatment in the stacked sample. Each column of this table refers to a specific dependent variable. In columns (1), (3), (5), and (7) we use as an outcome the number of incorrect statements, the number of correct statements, the number of verifiable ones, and the overall number of statements made by a politician in a week, respectively. In columns (2), (4), (6), and (8) we use as an outcome a dummy variable for any incorrect, correct, verifiable, and overall statements made by a politician in a week, respectively. We also report outcome summary statistics for treated politicians in the first three pre-intervention weeks (i.e., before treatment). Results refer to the field intervention conducted over the period from March 22 to July 11, 2021 (16 weeks). Standard errors in parentheses are clustered at the politician level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$, + $p<0.15$.

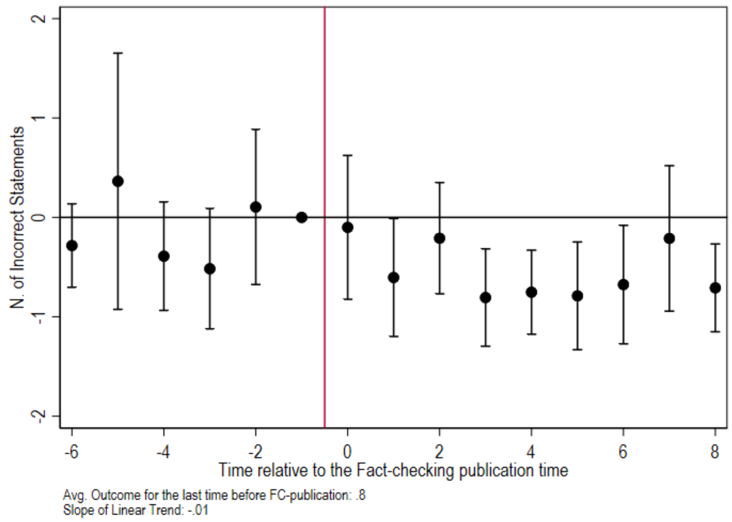
Table 5: Politicians' Response to Negative FC: Spillovers

	(1)	(2)	(3)
FC party peers	0.003		
	(0.025)		
FC party-chamber peers		0.005	
		(0.025)	
FC parliamentary commission peers			0.032
			(0.022)
Observations	4,416	5,600	5,104
Statistics on treated before treatment:			
Mean	0.06	0.05	0.06
SD	0.27	0.22	0.27
Politician-event FE	YES	YES	YES
Time-event FE	YES	YES	YES

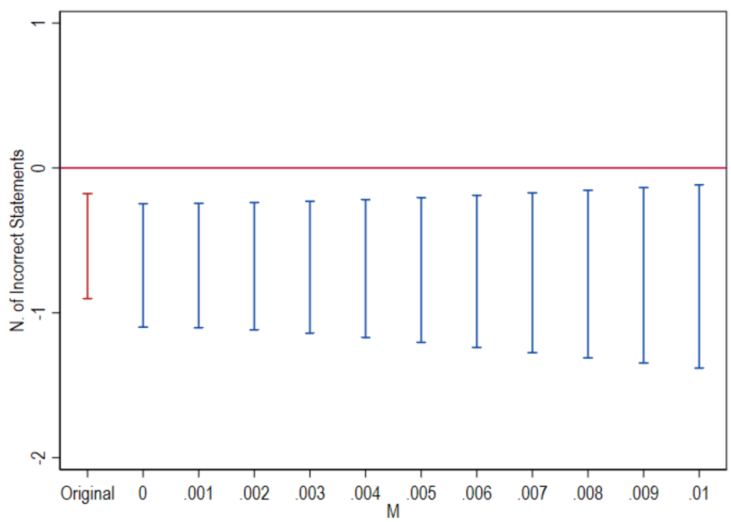
Notes: This table shows difference-in-differences estimates using as a dependent variable the number of incorrect statements made by a politician in a week. We estimate Equation (1) on a stacked sample at the politician-time-event level, controlling for politician-event and time-event fixed effects, where the event dimension identifies each sub-treatment in the stacked sample. Each column of this table refers to a specific test for potential spillover effects in the control group, and for this purpose, we drop fact-checked politicians and attribute a fake treatment to politicians in the control group as follows. In Column (1), we attribute the treatment to party peers of the treated politicians in the control group. In Column (2), the treatment is attributed to politicians belonging to the sample party-chamber of the treated one. In Column (3), the treatment is attributed to parliamentary peers sat in the same commission of the politician that has been actually treated during the field intervention. We also report outcome summary statistics for treated politicians in the first three pre-intervention weeks (i.e., before treatment). Results refer to the field intervention conducted over the period from March 22 to July 11, 2021 (16 weeks). Standard errors in parentheses are clustered at the politician level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$, + $p<0.15$.

Figure 1: Dynamic Effect of Fact-checking on the Number of Incorrect Statements

Panel A: Event Study

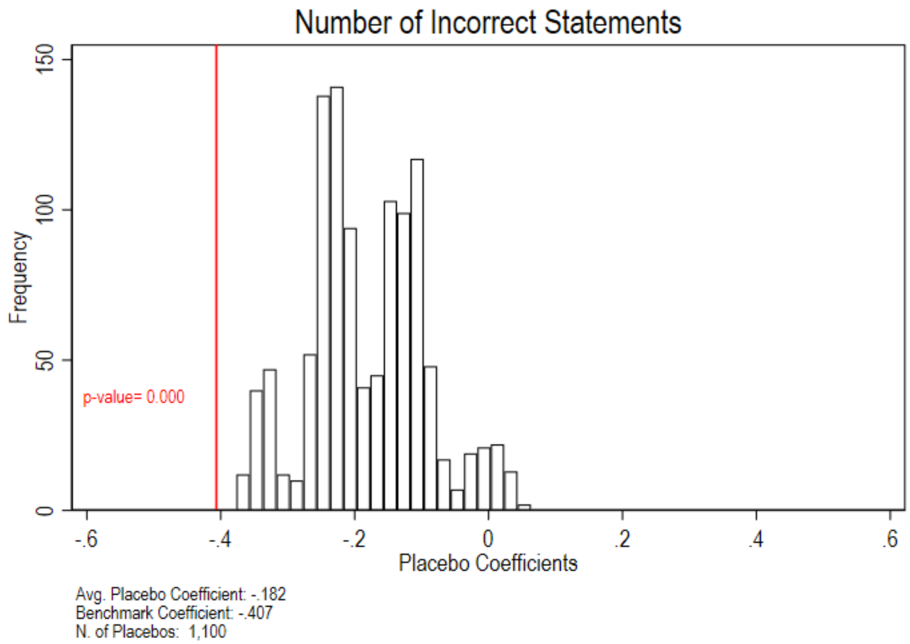


Panel B: Honest bounds of Rambachan and Roth (2023)



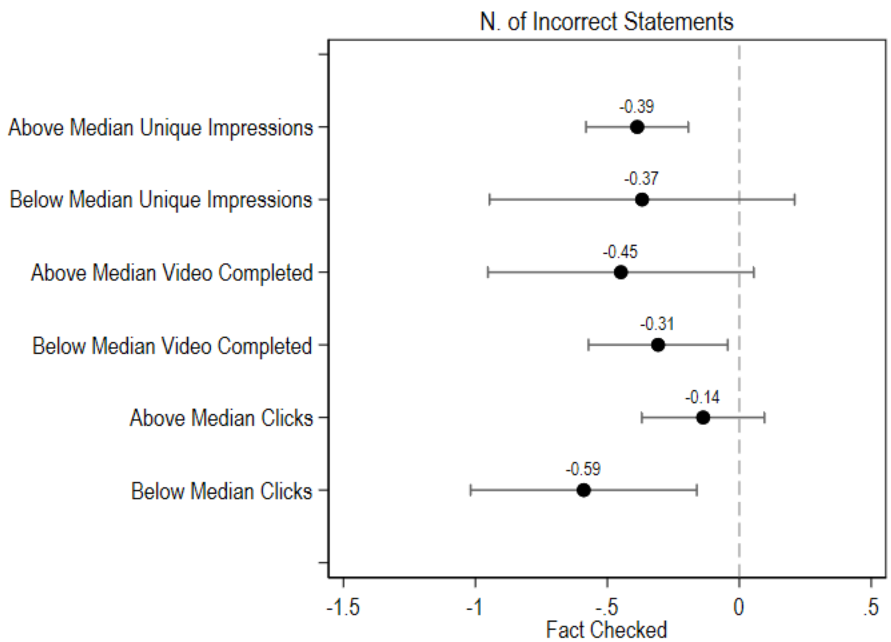
Notes: In Panel A, the figure shows estimates of the dynamic Equation (2) on a stacked dataset of panel data at the politician-week level over the period from March 22 to July 11 (16 weeks per event level). The regression controls for politician-by-event fixed effects and time-by-event fixed effects. Event-study estimates are normalized relative to the week before the fact-checking publication. The straight line indicates the time of the negative fact-checking treatment. Below the graph on the left, we report the slope of the pre-treatment event-study coefficients as well as the average of the outcome variable for the last week in the pre-treatment period. 95% confidence intervals are obtained after clustering the standard errors at the politician level. In Panel B, the figure reports the confidence sets described in Rambachan and Roth (2023) for the average of all post-fact-checking coefficients on the outcomes described in Panel A when we allow the slope of the pre-trend coefficients to change by no more than M—reported on the x-axis—across consecutive weeks. Value of $M = 0$ corresponds to the case where differences between treated and control are exactly linear.

Figure 2: Randomization Inference Tests



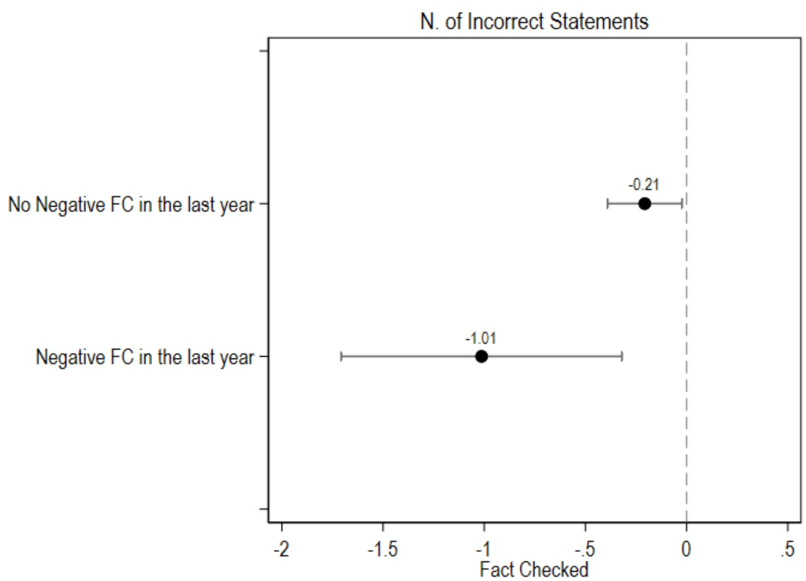
Notes: This figure plots the distribution of 1,100 difference-in-differences placebo estimates using as a dependent variable the number of incorrect statements made by a politician in a week. We estimate Equation (1) on a stacked sample at the politician-time-event level, controlling for politician-event and time-event fixed effects, where the event dimension identifies each sub-treatment in the stacked sample, with standard errors clustered at the politician level. We estimate placebo counterfactuals by randomly and fictitiously attributing the treatment in week t to a politician in the randomization pool of that week excluding the actually treated politician. The RI p-value is calculated as the proportion of placebo coefficients that are more extreme than the observed baseline coefficient that is shown as a red vertical line. Below the graph on the left, we report the number of placebo estimates, the average placebo coefficient, and the benchmark coefficient. These coefficients are estimated without taking into account the tenth intervention week since in this week there were no politicians in the randomization pool but the actual treated.

Figure 3: Politicians' Response by Engagement with Informational Campaign



Notes: This figure presents separate estimates of Equation (1), in which we sequentially add interaction terms of the treatment with a dummy indicating whether a politician is above the median in terms of unique video impressions, total video completed, and total clicks on the video, respectively. We plot the treatment coefficient to report the effect of the excluded heterogeneity category, and the linear combination of the treatment and the interaction term, showing the effect for the heterogeneity category. The dataset encompasses panel data at the politician-week level over the period from March 22 to July 11 (16 weeks per event level). The regression controls for politician-by-event fixed effects and time-by-event fixed effects. We display the point estimates above each coefficient spike and provide detailed estimates and corresponding summary statistics in Appendix Table F.2. 95% confidence intervals are obtained after clustering the standard errors at the politician level.

Figure 4: Politicians' Response by Exposure to Negative Fact-checking in the Last Year



Notes: This figure presents separate estimates of Equation (1), in which we add an interaction terms of the treatment with a dummy variable identifying whether a politician was exposed to negative fact-checking in the past year. We plot the treatment coefficient to report the effect on treated politician not exposed to a negative fact-checking in the last year, and the linear combination of the treatment and the interaction term, showing the effect for those who were actually exposed to a negative fact-checking in the last year. The dataset encompasses panel data at the politician-week level over the period from March 22 to July 11 (16 weeks per event level). The regression controls for politician-by-event fixed effects and time-by-event fixed effects. We display the point estimates above each coefficient spike. Summary statistics on treated before treatment for politicians with no negative FC in the last year: Mean 0.50, SD 0.93; with negative FC in the last year: Mean 1.33, SD 2.80. 95% confidence intervals are obtained after clustering the standard errors at the politician level.