Federal Reserve Bank of Cleveland Working Paper Series

Communicating Data Uncertainty: Multi-Wave Experimental Evidence for UK GDP

Ana Beatriz Galvão and James Mitchell

Working Paper No. 21-28

December 2021

Suggested citation: Galvão, Ana Beatriz, and James Mitchell. 2021. "Communicating Data Uncertainty: Multi-Wave Experimental Evidence for UK GDP." Working Paper No. 21-28. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202128.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Communicating Data Uncertainty: Multi-Wave Experimental Evidence for UK GDP*

Ana Beatriz Galvão[†] and James Mitchell[‡]

This version: December 2021

Abstract

Economic statistics are commonly published without any explicit indication of their uncertainty. To assess if and how the UK public interprets and understands data uncertainty, we conduct two waves of a randomized controlled online experiment. A control group is presented with the headline point estimate of GDP, as emphasized by the statistical office. Treatment groups are then presented with alternative qualitative and quantitative communications of GDP data uncertainty. We find that most of the public understands that uncertainty is inherent in official GDP numbers. But communicating uncertainty information improves understanding. It encourages the public not to take estimates at face-value, but does not decrease trust in the data. Quantitative tools to communicate data uncertainty - notably intervals, density strips, and bell curves - are especially beneficial. They reduce dispersion of the public's subjective probabilistic expectations of data uncertainty, improving alignment with objective estimates.

Keywords: Experiments; Data Uncertainty; Uncertainty Communication; Data Revisions

JEL Codes: C82, E01, D80

*An earlier version of this paper entitled "Communicating Data Uncertainty: Experimental Evidence for UK GDP," available at https://escoe-website.s3.amazonaws.com/wp-content/uploads/2020/07/14163259/ESCoE-DP-2019-20.pdf (co-authored with Johnny Runge), also included the analysis of a survey of "expert" users of economic statistics. We thank Sumit Dey-Chowdhury, Alexandra Freeman, Rob Kent-Smith, Ed Knotek, Sanjiv Mahajan, Heather Rolfe, David Spiegelhalter, Sally Srinivasan, Anne Marthe van der Bles, and Garry Young for helpful comments. Thanks also to conference and seminar participants at the Bank of Canada, Bundesbank, ESCoE, Federal Reserve Bank of Cleveland, Halle Institute of Economic Research, OECD, the Royal Statistical Society, and Strathclyde. This research has been funded by the ONS as part of the research program of the Economic Statistics Centre of Excellence (ESCoE). See https://www.escoe.ac.uk/projects/modelling-and-communicating-data-uncertainty/. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

[†]Warwick Business School, University of Warwick; ESCoE and CEPR.

[‡]Federal Reserve Bank of Cleveland and ESCoE.

1 Introduction

Economic statistics, in particular important measures of economic activity such as real GDP growth, are subject to revisions. GDP revisions aim to improve data accuracy, by incorporating information not available at the time of the earlier data release and can include methodological improvements too. More broadly, data revisions are one manifestation of "data uncertainty," with Manski (2015) distinguishing between "transitory," "permanent," and "conceptual" data uncertainties. Data uncertainty implies that agents need to consider how future data revisions affect their assessments of current economic conditions. Uncertainty about current estimates of economic activity and inflation has been used to explain how cautious, smooth changes in monetary policy can be optimal (Aoki (2003)). Data uncertainty can lead to disagreement among private agents about the current state of the economy, even after the first estimate of GDP growth is released; this can result in strategic uncertainties that can cause business cycles due to waves of optimism and pessimism as in Angeletos et al. (2018). Data uncertainty has also been perceived as comparable in size to the forecast uncertainty communicated by central banks. As evidence, note how the Bank of England's "fan charts" for GDP growth are almost as wide one quarter in the past as they are one quarter into the future.[1]

National statistical offices, however, do not typically communicate data uncertainty *explicitly*.[2] They present headline GDP as point estimates, arguably conveying a misleading degree of reliability in these data. This type of communication is common across national statistical offices - as emphasized by Manski (2015, 2019), and van der Bles et al. (2019). Given evidence that the unreliability of initial data releases affects policy decisions (e.g., see Orphanides (2001); Croushore (2011)), as intimated, some policymakers, such as the Bank of England and the Riksbank, provide their own (quantitative) estimates of data uncertainty for historical values of real GDP growth. This evidences a direct link between data uncertainty and monetary policy decisions.

This paper evaluates if and how different methods of communicating GDP data uncertainty affect the public's perceptions of GDP values, their understanding of data uncertainty, and their trust in the statistical office. Clements and Galvão (2017) and Galvão and Mitchell (2020) consider professional forecasters' and policymakers' assessments of data uncertainties, specifically due to data revisions. But it is not known how members of the public perceive data uncertainty. Given that statistical offices do not communicate measures of uncertainty in their GDP press releases, the public may take initial GDP estimates

[1] For example, see page 2 of https://www.bankofengland.co.uk/-/media/boe/files/inflation-report/2017/fan-charts-aug-2017.pdf.

[2] Statistical offices and central banks do increasingly communicate data uncertainty *implicitly* via publication and analysis of real-time databases and revision triangles. Statistical offices, like the Office for National Statistics (ONS) in the UK, also acknowledge data uncertainty in supporting documentation, usually available on their websites, by reminding users that early estimates of GDP have a lower data content than later estimates; e.g. see https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/introducinganewpublicationmodelforgdp/2018-04-27.

at face-value. Or they may infer their own error magnitudes around the numbers presented to them. We do not know. Nor do we know how statistical and data communications should be drafted for maximum impact.

We design and implement two waves of a randomized controlled trial to fill these information gaps. Importantly, the trial is also designed to evaluate how different ways of communicating and visualizing data uncertainty affect user comprehension and interpretation of data uncertainty. This involves measuring the effects of a set of randomized GDP data uncertainty communication *treatments* on a set of *outcomes*. These outcomes include the public's subjective probabilistic expectations of data uncertainty, their understanding of the causes of data revisions, and their trust in the data producer. We also assess whether heterogeneities across members of the public affect understanding and, in turn, whether there are differing implications for how data uncertainty should be communicated to different types of users of economic statistics.

The first wave of the randomized controlled trial, conducted in 2018 at a time of positive and relatively stable GDP growth, randomly sampled more than 3,000 (nationally representative) adults in the UK. The second wave, conducted during the coronavirus pandemic when UK GDP saw its worst-ever contraction in the second quarter of 2020, randomly sampled more than 4,000 adults. In both waves, the GDP data are communicated to individuals in the trial control group in a format that mimics recent Office for National Statistics (ONS) press releases.

We find that most of the public understands that there is uncertainty inherent in GDP numbers, even when presented with headline data releases that do not emphasize data uncertainty. But communicating additional uncertainty information, via one of the communication *tools*, improves the public's understanding. It encourages them not to take GDP estimates at face-value, but does not decrease trust in the data. Our evidence suggests that it is especially helpful to communicate uncertainty information quantitatively using intervals, density strips, and bell curves. Quantitative communication tools help anchor the public's otherwise dispersed subjective probabilistic perceptions of data uncertainty to objective revisions-based estimates of data uncertainty. The treatment effects for these quantitative communication tools are especially effective for individuals who are better informed about the economy and have more trust in the statistical office.

This paper therefore picks up Manski's (2015; 2019) call for empirical studies on how communication of uncertainties associated with economic statistics affects users. Similar calls have been made by Spiegelhalter et al. (2011) and van der Bles et al. (2019) in wider inter-disciplinary contexts.[3] Our use of randomized controlled trials follows a recent literature in macroeconomics that evaluates the impact of monetary policy communication on the public's expec-

[3] Using examples across different fields, Spiegelhalter et al. (2011) show that probabilities (even when known) are notoriously hard to communicate whether via words, numbers, or graphs. Empirical evidence is needed to establish what is understood and by whom. As Visschers et al. (2009) stress, in an inter-disciplinary review, the effects of different communication formats depend on the context, hence the need for empirical evidence for economic statistics.

tations of inflation and the economic outlook (Haldane and McMahon (2018); Coibion et al. (2020a)) and on their trust and understanding of policy messages (Bholat et al. (2019); Coibion et al. (2019)).

The plan of the remainder of this paper is as follows. Section 2 details the main measured responses or outcomes of the two surveys. It motivates our survey questions, including with reference to the small but growing literature on uncertainty communication outside economic statistics, especially meteorology. In addition, Section 2 explains how we measure GDP data uncertainty due to data revisions and sets out our candidate ways of communicating this uncertainty - our *communication tools*. These (with one constituting the control) form the treatments that are then randomized in the two public trials. Section 3 sets out how we measure and characterize the treatment effects of the different communication tools. Section 4 then analyzes the results from the two waves. It provides summary statistics from both surveys, before considering how the survey results let us examine the treatment effects on the outcomes of interest. Section 5 concludes. Online appendices contain supplementary material. Appendix A lists the survey questionnaires and provides summary statistics. Appendix B provides supplementary empirical results, including on the robustness of our main results.

2 Experimental Design, Data, and Empirical Background

In this section, we describe and motivate the design of the surveys.

2.1 Randomized Controlled Surveys

The surveys were conducted online as randomized controlled experiments. Implemented by Dynata, they take a representative sample of the UK population (across age, gender, and region using a quota sample).[4] To keep our surveys manageable, and without much larger sample sizes, in wave 1 we focus on five candidate ways of communicating and visualizing data uncertainty, two of which are qualitative and three quantitative. In wave 2, we expand this with one additional qualitative and one quantitative communication tool. Both of these reflect recent innovations in how the ONS has sought to communicate data uncertainty. These new communication tools were introduced by ONS during the coronavirus pandemic, in part drawing on the findings of our wave 1 survey (as written up in an earlier 2019 version of this paper).

[4]Dynata (formerly Research Now, when the survey was run) is a global online sampling and digital data collection company. Invitations are randomized and a survey router is used to support randomization. The samples are taken from the actively managed online panels maintained by Dynata and draw on a mixture of sources (invitation only, online partnerships, and online sites). Dynata follows the ESOMAR guidelines; which can be found at https://esomar.org/code-and-guidelines/icc-esomar-code.

The effects of these communication tools on the public's understanding of data uncertainty are contrasted with the effects of communicating, in effect, the current ONS headline press release to a *control* group. There is no (explicit) mention of uncertainty in this press release. Our sample size of about 3,000 (4,000 in wave 2) respondents means that around 500 respondents are in each of our six (eight in wave 2) treatment groups. Respondents are randomly allocated into one of these six or eight groups: the *control* group (presented with no uncertainty information) and five or seven *treatment* groups (presented with uncertainty information). This randomization lets us identify the causal effects of different ways of communicating uncertainty information.[5]

2.2 Characteristics of the Surveys

The surveys were structured so that the respondents would not anticipate that the survey is about data uncertainty *per se*, at least until partially through the survey. This was to minimize the chances of framing responses. Respondents were not allowed to go back to previous questions in the survey; that is, operationally, the survey always moves forward, with the respondent retaining sight of his/her randomly allocated communication tool (as shown in Table 1).

The surveys are not intended to capture conceptual uncertainties associated with how GDP is or should be measured. To control for the fact that the public may not know what GDP measures, and that this may affect their responses, prior to treatment they were directly asked what they think GDP is (question 10): "`To the best of your knowledge, which option most accurately describes what GDP is?`" Respondents could then reply that GDP measures the increase in prices, how many people are in employment, the size of the economy, the difference between exports and imports, they have no clue, or they have heard about GDP but are not sure what it is. After this question, if respondents either did not answer correctly (by agreeing that GDP measures the size of the economy) or did not answer the question, the survey provided these respondents with an explanation of what GDP does measure. They are reminded that "Gross domestic product (GDP) growth is the main indicator of economic performance" - a phrase taken directly from the ONS's own GDP press release.

To maximize realism, the surveys in both waves ask questions about the ONS's latest, at the time of the survey, GDP estimates and headline press release. At the time of running wave 1, in November 2018, this concerned the GDP point estimate of 1.5 percent for 2018Q3 published by the ONS on November 9, 2018. At the time of running wave 2, in August 2020, this concerned the GDP point estimate of -21.7 percent for 2020Q2 published by the ONS on August 12, 2020. These are year-on-year growth rates. This is based on the view that the public, arguably, is more familiar with year-on-year growth estimates presented over calendar years than quarterly growth rates. Our intention in the surveys is not to test the public's ability to understand and interpret different change measures. So we chose to frame our questions around, we believe, the

[5] Our focus is written communication; we do not consider oral news reports, such as radio.

most widely understood measure of growth.

2.3 Quantifying Data Uncertainty

In the absence of official information, from the ONS, quantifying GDP data uncertainty in the UK,[6] for the purposes of designing the surveys and testing the public's understanding of uncertainty, we assume a distributional form for this uncertainty. Specifically, we use estimates from Galvão and Mitchell (2020), based on a recent revisions analysis of the ONS's GDP estimates, to quantify "transitory" data uncertainty. Other sources of data uncertainty, for example due to limitations of the survey methodology, are not represented, and methodological work measuring non-sampling errors continues (Manski (2016)).[7] To facilitate cross-wave comparison, we also assume - based on the data - a common distributional form for data uncertainty across the two waves.[8]

We characterize GDP data uncertainty via a Gaussian density, centered on the ONS first-release point estimate, with standard deviation equal to the historical standard deviation of revisions to this first estimate over the subsequent four years. After four years, GDP growth estimates in the UK have gone through two annual (*Blue Book*) benchmarking and balancing processes (with supply and use tables). Revisions beyond this point tend not to reflect the arrival of additional survey information but methodological changes. The standard deviation of these revisions in the 20-year window between 1993Q2 and 2013Q1 is 0.8 percent and the mean absolute revision is 0.7 percent.[9] We assume zero mean revisions; that is, we assume the first release is an unbiased estimate of the revised estimate. This assumption, as shown in Galvão and Mitchell (2020), holds better for more recent ONS data. The Bank of England also assumes that historical GDP data uncertainty is characterized by a Gaussian density in its

[6] To quote the ONS: "The estimate of GDP ... is currently constructed from a wide variety of data sources, some of which are not based on random samples or do not have published sampling and non-sampling errors available. As such it is very difficult to measure both error aspects and their impact on GDP. While development work continues in this area, like all other G7 national statistical institutes, we don't publish a measure of the sampling error or non-sampling error associated with GDP." See https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi.

[7] Although the ONS does report and analyze data revisions, it notes explicitly at https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi that "there is no simple way of measuring the accuracy of GDP" and goes on to emphasize that while revisions tell us something about "reliability" "there are other aspects to accuracy, which revisions analysis cannot attempt to measure" (e.g. if a lower response rate than normal is received, the estimates are more uncertain even if they are not subsequently revised).

[8] There is evidence that, in fact, data revisions' uncertainty varies over time and is often larger at business cycle turning points; see Galvão and Mitchell (2020). It is also anticipated that the COVID-19 pandemic will lead to more revisions than the historical data suggest. This suggests scope for data communicators, such as statistical offices, to use their judgment (as well as past data) when quantifying data uncertainty. Central banks, such as the Bank of England, deploy a similar strategy of using judgment and *data* (including models) when its Monetary Policy Committee quantifies and then communicates forecast uncertainties via fan charts.

[9] We continue to consider year-on-year growth rates.

Inflation, now *Monetary Policy, Reports*. The bank's estimates of the standard deviation to first-release estimates of UK GDP growth have tended to increase since first published in 2007: they have fluctuated between 0.6 percent and 1.1 percent. Accordingly, to be broadly consistent both with the real-time evidence in Galvão and Mitchell (2020) and the practice at the Bank of England, we use a standard deviation estimate of 0.8 percent when quantifying GDP data uncertainty. We again emphasize the likely importance of data uncertainty in influencing households' real-world expectation formation and decisions, by noting how this estimate of 0.8 percent is about 70 percent the size of the Bank of England's typical expectations of one-quarter-ahead GDP growth; for example, the standard deviation of its one-quarter-ahead fan chart made in 2018Q3 is 1.1 percent.

2.4 Data Uncertainty Communication Tools: Treatments

In principle, for a given quantification of data uncertainty, there are a range of ways in which the uncertainty information can be communicated and/or visualized. Van der Bles et al. (2019) delineate nine candidate ways of communicating uncertainty: (i) a full explicit probability distribution (e.g., a fan chart); (ii) a summary of a distribution; (iii) a rounded number, range, or an order-of-magnitude assessment; (iv) a predefined categorization of uncertainty; (v) a qualifying verbal statement; (vi) a list of possibilities or scenarios; (vii) informally mentioning the existence of uncertainty; (viii) no mention of uncertainty; and (ix) explicit denial that uncertainty exists. This list follows a scale from the most comprehensive communication device, (i), to the narrowest one, (vii), including no communication of uncertainty and indeed denial of its existence (viii and ix).[10]

In turn, for each of these nine communication options, there are different ways of communicating and visualizing the uncertainty. Experimental evidence outside economic statistics has investigated how different visualizations of uncertainty and indeed the uncertainty of visualization matter; see Nadav-Greenberg et al. (2008), Joslyn and Savelli (2010), Correll and Gleicher (2014), Padilla et al. (2015), and Tak et al. (2015). Brodie et al. (2012) provide a review.

Even when not presented with a full probability density function to represent the uncertainty (like (i) on the nine-point scale above), users may still try to infer the underlying density function from the incomplete uncertainty information that they are provided. Tak et al. (2015) and Dieckmann et al. (2015, 2017) find, in their experiments, that when presented with range estimates (like (iii) on the scale above) users still seek to impose their underlying (subjective) density function. Accordingly, in our experiments we entertain a range of communication tools increasing in the degree of uncertainty information.

Each group in our survey is presented with a statement based on the latest GDP growth point estimate (of 1.5 percent in wave 1 and -21.7 percent in wave

[10] As Spiegelhalter et al. (2011) discuss, there are in fact a broader set of candidate ways of representing the uncertainty about continuous quantities like GDP growth, including interactive web-based and infographic formats that we do not explore in this paper.

2). Specifically, after 10 introductory questions (see Appendix A) that identify individual characteristics and the test and reminder of what GDP measures, the survey informs the respondents that:

> `The Office for National Statistics (ONS) publishes estimates of GDP growth. You will be asked a number of questions about this, so please take time to read the ONS statement below.`

Then each of the randomized groups, six in wave 1 and eight in wave 2, is presented with a different GDP communication tool. These tools are shown in Table 1.

As seen from Table 1, the control group is presented with something that closely resembles the current ONS headline press release. They are therefore not presented, directly, with any uncertainty information beyond the textual reference to uncertainty, given that the ONS does refer to its GDP numbers as "estimates." Groups 2, 3, and 7 (in wave 2) are then presented with a qualitative, qualifying verbal statement. Specifically, Group 2 respondents are warned explicitly that the number is approximate. This communication tool is deliberately only a minor tweak on the baseline stimulus above, in that it now also includes *about*. We therefore follow in the spirit of the Intergovernmental Panel on Climate Change (IPCC) (see Budescu et al. (2009)) in providing a textual confidence indicator. For Group 3, we add a warning that the number is approximate but we also provide more textual information on the fact that the values are subject to revisions, so that the point estimate communicated by the ONS is likely to change. Group 7, in wave 2, is presented with the textual confidence statement actually issued by the ONS in the summer of 2020 when publishing GDP estimates during the pandemic. This involves respondents being reminded, in words, that GDP estimates are especially uncertain due to challenges in collecting data under pandemic-induced lockdowns.

In contrast to these qualitative treatment tools, Table 1 shows how Groups 4, 5, 6, and 8 (in wave 2) are presented with alternative and, arguably, increasingly sophisticated quantitative impressions of GDP data uncertainty. These quantitative communications of uncertainty reflect the knowledge we as survey designers have (but the survey respondent does not) on what the *true* data density is assumed to be - given our quantification of data uncertainty, as explained in Section 2.3 above.

The amount of uncertainty information communicated increases from Group 4 through Group 8. For Group 4, in addition to the qualitative information presented to Group 3, we present a 60 percent confidence interval. We also include some details on how to interpret the probabilistic information communicated.[11] Group 5 is then presented with a density strip that provides additional information on how the probability mass is allocated across three 30 percent probability bands. Group 6 is provided with a distributional form for this uncertainty; this involves presenting Group 6 with a bell curve. It is shaded like a fan chart,

[11] There was a typo in one instance of the online wave 1 survey that meant Group 4 was told there was a 3 in 10 chance that GDP growth fell outside the blue line, not a 4 in 10 chance.

following recent practice at the ONS.[12] In turn, this builds on the Bank of England's pioneering approach to the communication of both historical and future uncertainty via its fan charts.[13] Group 8 is then presented with confidence intervals around the historical time-series of first estimates of GDP. This visualization of now both current and historical data uncertainty is taken directly from the ONS itself. Drawing on an earlier version of this paper, in April 2020 the ONS published an online article proposing how to convey data uncertainty. This included the proposal to publish 68 percent confidence intervals of the type shown in Table 1.[14] Inclusion of the communication tool shown to Group 8 in our second wave survey therefore provides one way to test the efficacy of this ONS proposal. We also emphasize that the confidence intervals shown to Group 8 involve the ONS quantifying GDP data revisions similarly to how we quantify data uncertainty for the other quantitative communication tools shown in Table 1, as explained in Section 2.3 above.

The ONS statement and the allotted data uncertainty communication tool are kept in front of respondents throughout the survey. Therefore, as the respondents move through the survey questions, they can always see their randomly allocated GDP communication treatment tool. We do not wish to test a respondent's memory.

2.5 Outcomes

This section delineates the main measured responses or outcomes evaluated in the surveys. As the surveys start by asking the public to characterize their perceptions of uncertainty qualitatively (or verbally), before asking for their quantitative perceptions, we discuss these first. We emphasize that it is the questions that elicit quantitative responses that enable meaningful interpersonal comparisons. The detailed survey questions, along with some descriptive statistics, are listed in full in Appendix A. Question numbers are referred to as q#.

[12] For example, see https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/migrationstatisticsquarterlyreport/july2018revisedfrommaycoveringtheperiodtodecember2017.

[13] In choosing how to communicate uncertainty to survey participants we made some choices in the interests of parsimony. For example, while the color of an uncertainty graph may well matter, we just use a common color across treatments, to avoid this affecting cross-group behavior.

[14] See https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/communicatinggrossdomesticproduct/2020-04-16. This article notes that "Depending on user feedback, we could implement the proposed confidence intervals around the latest estimates of calendar quarter GDP, which would be a marked departure from how we have previously communicated data uncertainty. This would reflect our first efforts to produce estimates of the inherent levels of uncertainty around GDP, which we would look to implement in the future if users would find this a helpful addition."

2.5.1 Qualitative Perceptions of Uncertainty in GDP Numbers

To gauge perceptions of single-valued GDP numbers, having observed their randomized communication tool, respondents are asked (q11): "`How accurate do you think the first estimate of GDP growth of 1.5% (-21.7% in wave 2) is likely to be?`" Respondents reply on a 4-point scale (`4=very accurate`, `3=fairly inaccurate`, `2=not very accurate`, and `1=very inaccurate`). Respondents are also asked for their views on the degree of informativeness of the communication tool presented (q19): "`Thinking back to the ONS statement about GDP growth, how much information did it give that the 1.5% (21.7%) estimate may be uncertain?`" Responses are on a 4-point scale (`not at all` = 1, through `a lot` = 4).

2.5.2 Quantitative Perceptions of Uncertainty in GDP Numbers

To measure quantitatively the perceived uncertainty of the estimate being communicated, the surveys ask (q14): "`What do you think is the chance that GDP grew (fell) by exactly 1.5% (21.7%)?`" Respondents answer on a 7-point scale (`7 = virtually certain - about a 99 in 100 chance (99%)`, `6 = Very likely - about a 9 in 10 chance (90%)`, through `1 = exceptionally unlikely - about a 1 in 100 chance (1%)`). The surveys deliberately use both words and numerical probabilities to describe the possibilities. This is because, as Spiegelhalter et al. (2011) emphasize, it can be hard to use words to convey precise probabilistic (uncertainty) information. One person's *very certain* may be different from another's.[15] Textual or verbal uncertainty statements have been found to be interpreted differently by different people; for example, experiments reported by Budescu et al. (2009) reveal large differences in the way people understand the verbal uncertainty phrases used by the IPCC. They recommend that both verbal terms and numerical values be used to communicate uncertainty and our surveys follow this practice.

The surveys go on to evaluate respondents' ability to interpret and quantify the uncertainty information provided by asking (q15): "`What do you think is the chance that GDP grew (fell) by between 1.2% (21.4%) and 1.8% (22%)?`" Possible replies are from `virtually certain - about a 99 in 100 chance (99%)`, through `very likely - about a 9 in 10 chance (90%)...` to `exceptionally unlikely - about a 1 in 100 chance (1%)`.

We posit a general *desiderata* that the public's understanding and use of any uncertainty information should be consistent with how the data communicator would like them to use it. In other words, we should hope that the *better* uncertainty information is communicated, the more the public's understanding of data uncertainty should align with the (assumed) objective interval/density estimate. As a consequence, for q15 we define an outcome variable equal to unity (zero otherwise) if a respondent's answer is correctly aligned with the uncertainty information actually communicated; that is, if the respondent an-

[15] And if words are used, which ones: natural frequencies (e.g., 1-in-10) or probabilities (e.g., 0.1)?

swered "`quite unlikely - about a 3 in 10 chance (30%)`" - as based on our quantification of data uncertainty, there is a 30 percent chance that GDP falls between the specific intervals given in q15.

Questions 12 and 13 ask for quantitative assessments of interval ranges around the GDP estimate. While lacking statistical interpretability, as now these bounds are not defined probabilistically, this sort of question is used in the weather forecasting literature - as a simple indicator of respondents' perceptions of uncertainty (e.g., see Joslyn and Savelli (2010)). The public is asked to place a number at the end of the following statement: "`I would not be surprised if actual GDP growth was as high (or low) as:_`" (given the negative GDP estimate in wave 2, the question is reworded as described in Appendix A). For each respondent, we compute the range between his/her high and low numbers and use this as an alternative measure of perceived uncertainty, albeit one, unlike q15 and q16, to which we now turn, that cannot be interpreted as a specific confidence interval.

2.5.3 Subjective Probabilistic Assessments of Data Uncertainty

We added to the wave 2 (2020) survey a question asking the public to express their expectations of data uncertainty as a subjective probability distribution (reported as a histogram). As emphasized by Manski (2004), an attraction of eliciting quantitative probabilistic responses is that probability provides a well-defined absolute numerical scale and thus better facilitates interpersonal comparisons. A disadvantage is that the public's understanding of quantitative uncertainty communication tools may be related to their ability to understand probabilities, as suggested by the weather forecasting literature.[16]

Our choice of probabilistic/histogram question is inspired by those included in the Survey of Professional Forecasters conducted by the Philadelphia Fed for the US and the European Central Bank for Europe. Specifically, q16 in wave 2 asked: "`Please provide (best-guess) estimates of the percentage probabilities you would attach to various outcomes for GDP growth. The probabilities should sum to 100% as indicated.`" Centered on the 21.7 percent outcome, with the central bin containing this outcome highlighted in bold in the question seen by respondents to aid interpretation, respondents are asked to report probabilities attached to interval bins of width 0.5 percentage points. The online form forced their probability estimates to sum to 100 percent.

We estimate the mean and standard deviation of each individual's reported histogram without making specific parametric assumptions about any under-

[16] For example, the survey evidence in Handmer and Proudley (2007) indicates that most lay users of probabilistic weather forecasts do understand probabilities, but that it matters whether the uncertainty is communicated verbally or numerically. Joslyn and Savelli (2010) find, using an online survey, that the public understands that there is uncertainty inherent in point forecasts. And they argue that the provision of explicit uncertainty estimates may be necessary to overcome some of the anticipated forecast biases that may affect the usefulness of weather forecasts given their uncertainties. Complementing this, Joslyn and LeClerc (2012) find that providing uncertainty forecasts associated with weather forecasts increases trust in the forecast and gives people a helpful idea of the range of possible outcomes.

lying continuous density that the respondent may subjectively have. As the first and last intervals are open-ended, we follow, e.g., Abel et al. (2016), and assume that the first and last intervals have a length double that of the central intervals. Results are not especially sensitive to this assumption. And following Zarnowitz and Lambros (1987), we assume that the probability mass is uniformly distributed within each interval rather than concentrated at the midpoint of each interval, although results are again robust to this.

The mean, μ_i, and standard deviation, σ_i, of individual i's histogram are then estimated as:

$$\mu_i = \sum_j \left(\frac{(u_j - l_j)}{2} \right) p_{i,j} \tag{1}$$

$$\sigma_i = \sqrt{\left[\sum_j \left(\frac{(u_j^3 - l_j^3)}{3\,(u_j - l_j)} \right) p_{i,j} - \left[\sum_j \left(\frac{(u_j^2 - l_j^2)}{2\,(u_j - l_j)} \right) p_{i,j} \right]^2 - \frac{w^2}{12} \right]} \tag{2}$$

where u_j and l_j are the upper and lower limits of the jth interval, w is the width of the central intervals, and $p_{i,j}$ is the probability that forecaster i assigns to the jth interval. The last term in the formula for σ_i is the commonly applied Sheppard correction for the variance.

To analyze the effects of the communication tool treatments on the public's probabilistic perceptions of data uncertainty as elicited via this question, we use the Cramer-von-Mises (CM) distance to measure the distance between each respondent's subjective histogram and the objective histogram as quantified via the communication tools seen in Table 1. Specifically, the CM distance is defined as:

$$CM_i = \sum_j \left(p_{i,j} - p_j^* \right)^2 \tag{3}$$

where $p_{i,j}$ is the reported probability respondent i attached to the j-th interval and p_j^* is the objective probability attached to this j-th interval, given the assumed Gaussian density with mean -21.7 percent and standard deviation 0.8 percent.

2.5.4 Understanding of Data Revisions

There is sometimes said to be a risk that communicating uncertainty information will erode trust in the data or indeed the data producer and/or communicator themselves. In turn, that trust may be affected by how the uncertainty information is communicated.[17] As a consequence, we also evaluate the impact of

[17]We do not pursue this here, but Raftery (2016) considers how statistical calibration may affect the confidence or trust in the (density) estimate/forecast, with confidence and trust increasing as calibration improves. One could imagine this working the other way round too. If the data communicator fears users will lose trust in it if the *final* estimate ends up outside the communicated uncertainty bands, even though this can still be consistent with correct calibration (e.g., 10 percent of *final* estimates should fall outside the 90 percent interval), they may apply judgment when quantifying data uncertainty to offset this.

uncertainty communication tools both on trust in the statistical office and on the public's beliefs about the sources of data revisions.

Research outside economics has found that simple indicators of uncertainty can be preferable; for example, see Budescu et al. (2009). Communicating uncertainty information may increase trust. For example, Joslyn and LeClerc (2013) find that including numerical uncertainty estimates with weather forecasts increases trust. But trust in the data producer might be related to how well uncertainty, and its sources, is understood.[18] It may well be that attitudes as well as trust affect how people interpret and react to uncertainty information. This has been found to be important when communicating climate change nowcasts and forecasts Visschers (2018).

Our surveys therefore seek to capture aspects of trust in GDP numbers and if and how this relates to attitudes to and understanding of revisions to these numbers. Question 9, presented before the GDP estimate is communicated, asks: "`Personally, how much trust do you have in economic statistics produced by the Office for National Statistics (ONS)? For example, on unemployment, inflation or economic growth?`" Replies are on a 4-point scale from `Trust them greatly` = 4 through `Distrust them greatly` = 1. Respondents are also allowed to reply `Not sure/don't know`.

After respondents receive the communication tool treatment, the surveys again test respondents' awareness of data revisions, by asking (q18): "`Are you surprised that estimates of GDP growth are regularly revised?`" Replies are on a 4-point scale from `not at all surprised` = 4 through `very surprised` = 1. Respondents were also allowed to reply: `N/A. I had never thought about it before doing this survey`. We treat this response separately below. The surveys also explicitly ask (q17) for views on the causes of data revisions: "`ONS regularly publishes revisions to their GDP estimates. Why do you think they do this?`" Respondents are invited to choose from among seven possible reasons for revisions, including mistakes at the ONS, vested interests, and/or the availability of more information.

3 Measuring the Treatment Effects of the Communication Tools

This section describes how we measure and test the treatment effects of the five/seven alternative communication tools of Section 2.4 on the set of outcomes detailed in Section 2.5.

Consider the outcome variable of interest y_i observed for individual i. The effect of communication treatment j on individual i is defined as β_{ij}:

$$\beta_{ij} = E(y_i|D_i^j = 1) - E(y_i|D_i^j = 0), (j = 1, ..., J) \tag{4}$$

[18] For example, people may not understand the process around data collection for economic data and therefore misinterpret information communicated to them about economic data uncertainty as evidence that the ONS has made mistakes or been incompetent.

where the dummy variable $D_i^j = 1$ (0 otherwise) if individual i was randomly allocated to Group j (where $j = 1$ is the control group). $J = 6$ in wave 1 (2018) and $J = 8$ in wave 2 (2020).

Both of these potential outcomes cannot be observed for individual i. But randomization of treatment, D_i^j, implies that we can measure average treatment effects via the difference in mean outcomes between the five or seven groups presented with uncertainty information and the control group told only that the GDP value is a point estimate. These average treatment effects, β_j, can be characterized via the generic linear model:

$$y_i = \alpha + \sum_{j=2}^{J} \beta_j D_i^j + \varepsilon_i \tag{5}$$

where $\varepsilon_i = \sum_{j=2}^{J} (\beta_{ij} - \beta_j) D_i^j + \upsilon_i$ and J is the number of communication tools ($j = 1$ is treated with the control group communication tool). The composite error, ε_i, includes the difference between the individual treatment, β_{ij}, and the average treatment β_j effects.

The null hypothesis that the average effect of treatment j ($j = 2, ..., J$) on outcome y is zero involves testing $\beta_j = 0$ in (5). Test statistics are obtained by least squares using robust standard errors. This hypothesis testing strategy assumes *iid* sampling for both y_i and D_i^j.

We also consider randomization tests. In these tests, the only stochastic element is due to the randomized allocation of treatment, as y_i is taken as fixed. Athey and Imbens (2017) argue for such tests, as developed by Fisher (1925), when using randomized experimental data; also see Young (2019). The randomization null hypothesis is that *all* of the treatment effects are zero:

$$\beta_{ij} = 0, j = 2, ..., J, \forall_i \tag{6}$$

and involves looking at all possible random allocations in the data, tabulating the distribution of the differences in the two means and then computing the probability of generating an outcome greater than the actual difference. This (sharp) null hypothesis is stronger than testing $\beta_j = 0$: when it holds it implies the weaker hypothesis of no *average* treatment effect, $\beta_j = 0$.

The communication tool treatments may affect different types of individuals heterogeneously. So we consider whether treatment effects differ along reported characteristics of the public, as elicited in our surveys. Understanding such heterogeneity is useful for the statistical office if interested in maximizing the effects of communications on beliefs by targeting specific subgroups that are more responsive. Specifically, we add to the model a $k \times 1$ vector of exogenous variables, W_i, capturing individual characteristics of the respondents as elicited via the first 10 questions to the survey. The W_i have associated coefficient vector, γ, allowing the treatment effects to vary with these:

$$y_i = \alpha + \gamma W_i + \sum_{j=2}^{J} \left(\beta_j + \beta_j^W W_i\right) D_i^j + \varepsilon_i, \tag{7}$$

where $\varepsilon_i = \sum_{j=2}^{J}(\beta_{ij} - \beta_j - \beta_j^W W_i) D_i^j + \upsilon_i$. The W_i are not affected by the treatments. Their consideration, by in effect dividing the N-sample into stratified sub-samples, assuming $\beta_{ij} = \left(\beta_j + \beta_j^W W_i\right)$, provides one measure of heterogeneity in the communication treatments. In Section 4.5 below, we report these *conditional* average treatment effects, focusing on respondents who have heard of the ONS (q8), trust the ONS (q9), and understand what GDP is (q10). This is complemented by use of the non-parametric tests of Crump et al. (2008) to examine heterogeneities across all subgroup characteristics, W_i.

We lead our analysis in Section 4 by presenting average and then heterogeneous (conditional) treatment effects estimated via least squares estimation of (5) and (7). Such regression-based estimators are popular, including in the growing literature in macroeconomics using randomized controlled trials (see Haldane and McMahon (2018); Bholat et al. (2019); Coibion et al. (2019); Binder (2020)). We note that for those y_i where the responses are discrete, results are robust to the use of probit or ordered probit estimation. When analyzing the histogram question (q16), due to evidence of outliers, we estimate quantile regressions and thereby report average treatment effects by quantile. The results of the randomization tests are summarized in Section 4.4, along with robustness checks.

4 Survey Results

Appendix A lists the survey questions and summarizes responses across the two waves. Some summary statistics to mention upfront are: about half of respondents claimed some knowledge of economics (q6); a similar proportion correctly stated what GDP measures (q10), had heard of the ONS before the survey (q8), and said they tended to trust the ONS (q9).[19]

Individual characteristics and opinions (i.e., answers to q1 through q9) are generally very similar across the two waves, as we should expect given the representative nature of the samples. Two apparently little differences in their opinions are worth mentioning, however. First, respondents in August 2020 appear more aware of the existence of the ONS (q8): 58 percent had heard of the ONS, compared to 49 percent in November 2018. This heightened awareness may be due to the prominent role that ONS statistics played during the 2020 pandemic. Second, wave 2 respondents had a better understanding of GDP as a concept, with 55 percent answering the test question correct compared to 46 percent in wave 1.

[19]This is consistent with independent survey evidence. The 2019 *Public Confidence in Official Statistics* report, produced by the National Centre for Social Research (NatCen) on behalf of the UK Statistics Authority, similarly finds that 85 percent of people who gave a view trusted the statistics produced by the ONS; see https://uksa.statisticsauthority.gov.uk/news/pcos-2019/.

4.1 Qualitative and Quantitative Perceptions of Uncertainty in GDP Numbers

To test how perceptions of data uncertainty are affected by the different communication tools, Table 2 reports average treatment effects, by communication tool, for the five survey outcomes measuring, in different ways, perceptions of data uncertainty. For each of these five outcomes, the first column of Table 2 presents the average response in the control group. The remaining columns report the average treatment effect, relative to the control group (G1), for each of the five or seven treatments. We report estimated robust standard errors below. And, for ease of reading but without wishing to emphasize a particular significance level, the average treatment effect is placed in bold when suggesting statistical significance at the 10 percent level.

4.1.1 Control Group Perceptions

Before evaluating the effects of the communication tools, we summarize the responses of the control group: those shown only the regular headline GDP data release. Looking at the G1 column in Table 2, we see that the control group, on average, felt that they were only given "`some`" indication that the GDP data are uncertain (q19). Despite this, they do perceive the ONS's GDP point estimate to be subject to inaccuracies: the average response (to q11) is to expect the GDP data to be "`not very accurate`." They also anticipate data uncertainty: the average response (to q14) is to attribute a "`fifty-fifty`" chance to GDP growing (or falling) by exactly the number shown in the headline press release. The mean width of the range interval (from q12 and q13) for the control group was 2.7 percentage points in 2018 and 12.6 percentage points in 2020.[20] This further supports the view that the public does understand that uncertainty is inherent in the ONS's GDP estimates, even when not treated with an uncertainty communication tool. It also shows that uncertainty perceptions were substantially higher in 2020 than in 2018. Finally, we see that only about 10 percent of the control group correctly attributed a 30 percent chance to GDP growing between the stated interval (q15): this is 4 percentage points lower than we would expect if respondents replied to this question randomly.

4.1.2 Evaluating Treatment Effects

Table 2 shows that the communication tools do affect the public's qualitative and quantitative perceptions of the accuracy of GDP estimates. Overall, looking across the five outcomes, the interval estimates shown to Groups 4 and 8 and the bell curve, shown to Group 6, stand out as having the largest causal effects on the public's perceptions of data uncertainty. These effects are often statistically signficant.

[20]We note that the median width of the range interval for the control group was 1.00 percentage points in 2018 and 10.00 percentage points in 2020. This fits with evidence that respondents tend to reply with rounded numbers; see Manski and Molinari (2010)

Looking at the sign of the treatment effects for q11 in Table 2, we see that all of the communication tools encourage the public to believe that the GDP point estimate is less accurate than if they were not presented with any uncertainty information. Similarly, we also see from q14 that all of the communication tools lead the public to decrease the probability they attach to GDP growing at exactly the rate communicated in the headline press release. These effects were stronger in 2020, with its extreme GDP data realization of -21.7 percent, than in 2018. The communication tool treatment effects are strongest, and statistically significant, for the interval estimate communicated to Group 4 and for the bell curve communicated to Group 6. In contrast, the textual uncertainty qualifier given to Group 2 tends to have little effect.

The responses to q19 confirm that the communication tools are perceived to be more informative when either a qualifying verbal assessment of data uncertainty (as shown to Group 3) or a quantitative impression of uncertainty (shown to Groups 4 to 8) is provided. The positive sign of these estimates suggests that these treatments cause more respondents to agree that they were being shown more uncertainty information.

Next, we test whether the communication tools increase the probability that the public correctly infers that there is a 30 percent chance of GDP growing between the interval stated in q15. The summary statistics in Appendix A show that only 13 percent (12 percent in 2020) of the public overall clicked on this answer. They also confirm the impression that the majority of the public does not take the GDP estimate at face-value: fewer than 20 percent (14 percent in 2020) of the public thinks it is "`very likely`" or "`virtually certain`" that GDP, in fact, grew by the exact GDP estimate communicated (see Appendix A, q15). The average treatment effect estimates in Table 2 suggest that, as before, the quantitative communication strategies improve the likelihood of a correct answer. That is, the predictive interval (Group 4) and the bell curve (Group 6) communication tools, respectively, lead to individuals being 3 to 4, and 6 to 7, percentage points more likely to answer q15 correctly than the control group.

A related outcome assessed in Table 2 considers the answers from questions 12 and 13. Recall that these questions asked respondents to provide high and low numbers that they would not be surprised to observe for *actual* GDP growth. For each respondent, we compute the range between his/her high and low numbers. As shown in Appendix A, about 35 percent of respondents in wave 1 and about 32 percent in wave 2 chose not to provide answers to these questions, perhaps suggesting an inability or reluctance to quantify data uncertainty.[21] Focusing here on those respondents who replied, Table 2 reports average treatment effects for this interval question.[22] In the 2018 survey, we see that only the bell curve has a significant effect: its communication, on av-

[21] A small(er) number of individuals (77 in wave 1 and 194 in wave 2) failed to report a lower bound value lower than the upper bound; these individuals are added to the group of respondents who chose not to reply and are effectively treated as missing.

[22] Note that, due to randomization of the treatment, these estimates remain valid even if individuals who replied are not a random sample from the population as a whole. In Section 4.4, for robustness, we estimate treatment effects explicitly conditioning on response.

erage, increased the width of the reported interval. As the interval ranges for the control group in wave 1 appear rather narrow compared with the estimates of data uncertainty in Section 2.3, the bell curve helps align individuals' perceptions of data uncertainty with revisions-based estimates. But quantitative communication tools have more impact on the interval range in the second wave of the survey, conducted during the pandemic. In 2020, quantitative communication tools dramatically decrease the width of the interval. Individuals who were not treated with a quantitative measure of uncertainty perceived more data uncertainty than the objective revisions-based estimates.

4.1.3 Summary

The results in Table 2 show that the UK public does perceive GDP point estimates as uncertain. They also demonstrate that providing the public with, in particular, quantitative expressions of data uncertainty further encourages the public to view GDP data as uncertain. The quantitative communication tools importantly lead to more of the public correctly inferring the degree of data uncertainty. During the heightened uncertainty of the pandemic, these quantitative communication tools lead to the public not overestimating data uncertainties. By contrast, the qualitative communication tools have less causal effect on assessments of data uncertainty.

When we consider that a large proportion of the public is neither sure what GDP measures nor what the ONS does, it is perhaps encouraging that we are able to find statistically significant improvements in terms of how the public understands data uncertainty when quantitative impressions of data uncertainty are communicated to them.

4.2 Probabilistic Perceptions of Data Uncertainty

Question 16 in wave 2 elicited probabilistic perceptions of data uncertainty from each respondent. Section 2.5.3 above describes how we compute the moments from each individual's histogram.

Figure 1 plots, for each respondent except those who attach 100 percent to a single bin, his/her mean and standard deviation estimates as estimated from the reported histograms. Figure 1 shows considerable dispersion both in the reported means and standard deviations, although there is a tendency for the mean estimates to be anchored around the ONS's point estimate of -21.7 percent. But the uncertainty estimates, as measured by the standard deviation, are very disperse. Recall that the correct (objective) revisions-based estimate of data uncertainty, as reported via the quantitative communication tools, is of a standard deviation estimate of 0.8 percent. Weather forecasting communication studies have also found that where uncertainty information is not shown, people tend to make their own assumptions (see Morss et al. (2010); Joslyn and Savelli (2010)), often over-estimating uncertainty.

As described in Section 2.5.3, we use the Cramer-von-Mises (CM) distance to quantify the distance between the objective and each individual's subjective assessment of data uncertainty. Table 3 reports the effects of treatment, by communication tool, on the CM distance. Given that, as shown in Figure 1, there is considerable heterogeneity in respondents' quantitative perceptions of data uncertainty, we report quantile treatment effects to offer robustness to outliers. Specifically, Table 3 reports average treatment effects for the 0.25, 0.5, and 0.75 quantiles. The results in Table 3 show that the quantitative communication tools, with the exception of the time-series interval shown to Group 8, continue to have statistically significant effects. The negative sign of the quantile estimates shows that these communication tools close the distance between the public's and the assumed objective probabilistic estimates of the GDP data density. They encourage the public to infer the degree of GDP data uncertainty correctly.

4.3 Sources of Data Revisions

Toward the end of both surveys, respondents were asked if they were aware of data revisions and, then, why they think the ONS revises its GDP estimates. Recall that all our communication tools, with the exception of those given to the control group (Group 1) and Group 2, contain the phrase "`but this estimate is likely to be revised as updated information becomes available`." As indicated in the summary statistics in Appendix A, across the two waves about 70 percent of the public claim awareness of data revisions (i.e., they answered: `not that/not at all surprised` to question 18). But, in both waves, about 15 percent of the public admits to never having previously thought about data revisions.

In Table 4 we evaluate whether the communication tools affect awareness of revisions. The estimates in the first main row of Table 4 (for *revisions awareness*) suggest that the communication tools tend to raise awareness of data revisions (the effects are mostly positive). But these effects appear small and are rarely statistically signficant. In turn, the communication tools do not obviously decrease the proportion of the public that shows *no awareness* of data revisions (see the second main row of Table 4). As we expand on in Section 4.5, perceptions of data revisions depend on individual characteristics.

Table 4 then shows estimates of the effects of treatment on the public's explanations for data revisions (q17). The outcome variable is defined as a binary variable equal to unity (zero otherwise) if the respondent felt that revisions were explained by: "`vested interests`," defined as either the ONS or the government having vested interests in data production and collection; `mistakes at the ONS`; or when they identify revisions as due to `more information becoming available`. As Appendix A shows, 26 percent of the public in 2020 and 29 percent in 2018 thought that vested interests are at work; 9 percent (11 percent in 2020) stated that ONS mistakes are to blame; and 53 percent (59 percent in 2020) understood (in general, we should add, correctly) that revisions are explained by updated information.

Table 4 shows that the different communication treatments do not have strong causal effects on whether the public believes data revisions are due to either vested interests or mistakes at the ONS. While 19 of the 25 treatment effects (across the two waves and the different communication tools) are negative in sign - suggesting that treatment does discourage the public from viewing data revisions as due to these malign factors - the effects are small in absolute terms (less than 5 percent relative to the control group) and not statistically significant. Similarly, while the communication tools do encourage the public to view data revisions as due to more information arriving, with 10 of 12 treatment effects positively signed, again these effects are weak both in absolute terms and as evidenced by statistically insignificant effects. Only the density strip in 2018 has a positive and statistically significant effect.

Overall, we conclude that communicating uncertainty about early releases of GDP by providing quantitative information alongside the point estimate (as in the density strip and bell curve) improves the public's quantitative perceptions of data uncertainty. But these treatments do not affect public trust in the statistical office. They do not lead to individuals thinking that data revisions are because of vested interests or mistakes at the ONS or the government.

4.4 Robustness Checks

Use of a randomization test, with the stronger null hypothesis, (6), confirms the finding from Tables 2 and 3 that it is the quantitative communication tools that most often have statistically significant effects on the public's qualitative and quantitative assessments of data uncertainty; see Table B1. When a specific communication tool is found to have a statistically significant *average* effect, in Tables 2 or 4, it tends to also have in Table B1 a lower p-value for the null hypothesis that *all* individuals' treatment effects are zero. Table 4's conclusion that communicating uncertainty information does not erode trust in the ONS is also robust to the use of the randomization test (see Table B2). To mitigate the risk of spurious treatment effects, due to multiple hypothesis testing across the different outcome variables seen in Tables 2 and 4, we also report in Tables B3 and B4 p-values controlling for joint testing. Results are again consistent across the tables.

As discussed above, about a third of respondents chose not to reply to the range interval questions (q12 and q13), perhaps suggesting an inability or reluctance of some individuals to quantify data uncertainty. Heckman (1976) selection models, where the treatment effects are conditioned on selection, i.e., on the individual replying to q12-q13, were therefore estimated. Selection is explained by the individual characteristics, as elicited through the introductory survey questions. As we would expect with our experimental data, the treatment effects from the Heckman selection model presented in Table B5 are similar to those shown in Table 2. Interestingly, Table B5 also indicates that individuals who have heard of and trust the ONS and correctly understand GDP are more likely to reply to questions 12 and 13. This motivates the heterogeneity analysis that follows.

4.5 Do the Communication Tool Treatments Affect Individuals Differently? Heterogeneity in Average Treatment Effects

We now evaluate whether the treatment effects are heterogeneous, i.e., whether they differ by reported characteristics of the respondent.

We initially focus on nine sub-samples of our data, as identified by the introductory questions in the surveys. Before treatment, these questions elicit information on characteristics and opinions of the respondents, specifically their gender, age, education, employment status, background in economics, how frequently they follow news about the economy, whether they have heard of and/or trust the ONS, and on whether they understand what GDP measures.[23] Preliminary analysis, using the non-parametric tests for heterogeneous treatment effects developed by Crump et al. (2008), suggests that of these nine characteristics, having heard of the ONS, trusting the ONS, and correctly identifying what GDP measures often stand out as important (see Tables B6 and B7).[24] This is consistent with the Heckman selection results of Table B5, where again these three characteristics were found to best correlate with the outcome variables.[25]

This motivates further analysis of how treatment effects differ for these, what we call, "informed and trusting" individuals. We note that these informed and trusting individuals: i) tend to be older (the proportion of individuals age 34 or less in the informed and trusting group is 16 percent in 2018 (15 percent in 2020) but 38 percent in 2018 (41 percent in 2020) for the uninformed and untrusting; ii) are more likely to have studied economics at the graduate level (33 percent versus 11 percent in 2018 and 28 percent versus 9 percent in 2020); and iii) more frequently consult the news (60 percent versus 14 percent in 2018 and 48 percent versus 10 percent in 2020).

Specifically, Table 5 presents for the qualitative and quantitative data uncertainty outcomes, average treatment effects for these informed and trusting members of the public. This sub-group comprises just over a quarter of the total sample in wave 1 and just over a third in wave 2. Estimates are contrasted with those for uninformed and untrusting individuals (i.e., individuals who have not heard of the ONS, do not trust the ONS, and who incorrectly identified what

[23]Information on where the respondents live was also gathered. But as this had no relationship with the outcome variables, it is dropped from our analysis.

[24]These characteristics, especially for the data revision outcomes, are also often selected by the Bayesian information criterion (BIC) when the BIC is used to select that subset of characteristics to be included in the model for the chosen outcome variable. If we use the less parsimonious Akaike information criterion, we again see these three characteristics most commonly being selected.

[25]Tables B6 and B7 do show, however, at best weak evidence that these correlations translate into statistically significant heterogeneities in the treatment effects themselves. The conditional (on observable characteristics) treatment effect tests reported in Tables B6 and B7 align with the average treatment effect tests: the p-values from the two sets of tests are similar. This, in turn, is consistent with the tests of constant conditional average treatment effects. These tend not to indicate statistical evidence for heterogeneities except for the two quantitative uncertainty questions (q15 and q16), where we see in Table B6 a greater tendency to reject the null of a constant treatment effect across observable characteristics.

GDP measures). This sub-group is smaller, about a fifth of the total sample across the two waves; it is also smaller in wave 2 than in wave 1, suggestive of the public using and trusting data more during the pandemic. This is consistent with the aforementioned heightened public awareness of the ONS in wave 2.

Table 5 reveals that the quantitative communication tools, in particular, tend to have stronger effects on the informed and trusting. These treatments encourage these individuals, relative to the uninformed, to view the reported GDP point estimate as uncertain (q14), to acknowledge that the communication tool is informative (q19), and to classify the probability of GDP falling within the stated bounds correctly (q15). The effects on the CM distance (q16) between the subjective and objective probabilistic assessments of data uncertainty are especially revealing. While the quantitative communication tools, with the exception of the time series interval, do encourage the informed and trusting to report more accurate probabilistic assessments of data uncertainty, they have little or no effect on the uninformed and untrusting. Indeed, the qualitative communication tools cause the uninformed and untrusting to make even worse probabilistic assessments of data uncertainty.[26] This suggests that ONS communications of data uncertainty will be more effective the greater the proportion of the public that is "informed and trusting." In turn, this points to gains from investments in improving the public's understanding of economic data, with scope for experimental research to again inform on the most effective means of achieving this.

But Table 5 shows that the communication tools do have larger and stronger effects on the uninformed and untrusting when measuring the treatment effects on the reported width of the interval, as elicited via q12 and q13. In 2020, when the reported width of the interval was, on average, much wider than in 2018, the quantitative communication tools cause the uninformed and untrusting to decrease their perceptions of data uncertainty far more drastically than seen for the informed and trusting.

Table 6 considers the outcomes based on knowledge of data revisions. Here there is less difference between the effects of the communication tools on the informed/trusting and the uninformed/untrusting. This is consistent with Table 4, which also showed the communication tools to not, across all members of the public, have strong effects on awareness and understanding of data revisions. However, Table 6 does reveal some heterogeneities underlying the weak treatment effects seen in Table 4. In particular, the communication tools especially discourage the uninformed and untrusting from believing that data revisions are due to vested interests or mistakes at the ONS. Indeed, in 2018 the bell curve causes more of the uninformed and untrusting to believe data revisions are due to the arrival of additional information. But this effect is not seen in 2020.

[26] Table B8 presents additional details on the breakdown of these treatment effects on the CM distance. It shows that the conditional treatment effects are larger (and statistically stronger) for those members of the public who know what GDP is (but may not have heard of or trust the ONS), who have heard of the ONS (but may not trust it or know what GDP is), and for those who trust the ONS (but may not have heard of the ONS or know what GDP is).

5 Conclusions

Official estimates of GDP, as published by national statistical offices, are revised over time. Data uncertainty obscures decisions that depend on current estimates of economic growth. Despite growing awareness of the importance of data uncertainty and, acknowledging this, increased availability and analysis of real-time data vintages on statistical office and central bank websites, statistical offices continue to communicate headline GDP as a point estimate. This paper contributes new insights into the implications of this communication strategy. It considers how data communications could be designed to improve the public's understanding of data uncertainty and to increase trust in data.

Using two waves of a randomized controlled trial, with a combined sample of more than 7,000 adults representative of the UK population, this paper finds that most of the UK public does not actually take initial GDP point estimates at face-value. They attribute a degree of inaccuracy and uncertainty to single-valued GDP numbers, as commonly communicated in headline data releases. Treatment groups are then presented with alternative communications and visualizations of GDP data uncertainty, with individuals randomly assigned to a given treatment group. The key finding, across the two waves of the experiment run at times of economic growth and during the pandemic recession, is that if and how uncertainty information is communicated to the public matters.

Communicating uncertainty information alongside the GDP point estimate improves the public's understanding of data uncertainty, but does not reduce its trust in the statistical office. It encourages more of the public to view the point estimate as just that: a point within a range of possible outcomes. The most effective communication tools are those that quantify and visualize data uncertainty, via either confidence intervals, density strips or bell curves. These results are consistent with emerging inter-disciplinary evidence that providing quantitative uncertainty information leads to a better understanding of the range of possible outcomes, but need not erode trust in the data (see Joslyn and LeClerc (2013)).

Absent communication of data uncertainty, the public's probabilistic perceptions of GDP data uncertainty are dispersed and inaccurate. When the public is treated with quantitative communication tools, we find that the public's perceptions become better aligned with objective estimates of data uncertainty, as measured by data revisions. Treatment effects are stronger for individuals who are better informed about the economy and have more trust in the statistical office.

Our experimental findings suggest that by directly communicating data uncertainty, statistical offices can better anchor the public's, at times wild, expectations of data uncertainty to their own estimates. This should facilitate improved decision making, at least to the degree that the public's expectations of data uncertainty better anticipate future data revisions. These results are consistent with recent experimental evidence finding that how a central bank communicates with the public also affects expectations of macroeconomic variables (see Haldane and McMahon (2018); Coibion et al. (2019)).

This paper focuses on UK GDP data uncertainty. Future research should carry out similar experiments for other countries and consider estimates for other economic variables. As van der Bles et al. (2019) review, some statistical offices do compute sampling error estimates for some economic variables, such as unemployment, which could be exploited when testing the public's understanding of uncertainty information if and when communicated to the public in different forms.

References

Abel, Joshua, Robert Rich, Joseph Song, and Joseph Tracy (2016). "The measurement and behavior of uncertainty: Evidence from the ECB Survey of Professional Forecasters." *Journal of Applied Econometrics*, 31(3), pp. 533–550. doi:10.1002/jae.2430.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2018). "Quantifying confidence." *Econometrica*, 86, pp. 1689–1726. doi:10.3982/ECTA13079.

Aoki, Kosuke (2003). "On the optimal monetary policy response to noisy indicators." *Journal of Monetary Economics*, 50, pp. 501–523. doi:10.1016/S0304-3932(03)00023-0.

Athey, Susan and Guido W. Imbens (2017). "The econometrics of randomized experiments." In *Handbook of Economic Field Experiments*, volume 1, pp. 73–140. Elsevier. doi:10.1016/BS.HEFE.2016.10.003.

Bholat, David, Nida Broughton, Janna Ter Meer, and Eryk Walczak (2019). "Enhancing central bank communications using simple and relatable information." *Journal of Monetary Economics*, 108, pp. 1–15. doi:10.1016/j.jmoneco.2019.08.007.

Binder, Carola (2020). "Coronavirus fears and macroeconomic expectations." *The Review of Economics and Statistics*, 102, pp. 721–730. doi:10.1162/rest_a_00931.

Brodie, Ken, Rodolfo A. Osorio, and Adriano Lopes (2012). "A review of uncertainty in data visualization." In John Dill, Rae Earnshaw, David Kasik, John Vince, and Pak Chung Wong, editors, *Expanding the Frontiers of Visual Analytics and Visualization*. Springer. doi:10.1007/978-1-4471-2804-5_6.

Budescu, David V., Stephen Broomell, and Han-Hui Por (2009). "Improving communication of uncertainty in the reports of the intergovernmental panel on climate change." *Psychological Science*, 20, pp. 299–308. doi:10.1111/j.1467-9280.2009.02284.x.

Clements, Michael P. and Ana B. Galvão (2017). "Predicting early data revisions to US GDP and the effects of releases on equity markets." *Journal of Business and Economic Statistics*, 35(5), pp. 389–406. doi:10.1080/07350015.2015.1076726.

Coibion, Olivier, Yuriy Gorodnichenko, Edward S. Knotek II, and Raphael Schoenle (2020a). "Average inflation targeting and household expectations." Working paper 20-26, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202026.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele (2020b). "Inflation expectations and firm decisions: New causal evidence." *Quarterly Journal of Economics*, 135, pp. 165–219. doi:10.1093/qje/qjz029.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber (2019). "Monetary policy communications and their effects on household inflation expectations." Working paper 25482, National Bureau of Economic Research. doi:10.3386/w25482.

Correll, Michael and Michael Gleicher (2014). "Error bars considered harmful: Exploring alternate encodings for mean and error." *IEEE Transactions on Visualization and Computer Graphics*, 20, pp. 2142–2151. doi:10.1109/TVCG.2014.2346298.

Croushore, Dean (2011). "Frontiers of real-time data analysis." *Journal of Economic Literature*, 49, pp. 72–100. doi:10.1257/jel.49.1.72.

Crump, Richard, V. J. Hotz, Guido Imbens, and Oscar A. Mitnik (2008). "Nonparametric tests for treatment effect heterogeneity." *The Review of Economics and Statistics*, 90(3), pp. 389–405. doi:10.1080/07350015.2020.1737080.

Dieckmann, Nathan, Robin Gregory, Ellen Peters, and Robert Hartman (2017). "Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning." *Risk Analysis*, 37, pp. 471–486. doi:10.1111/risa.12639.

Dieckmann, Nathan F., Ellen Peters, and Robin Gregory (2015). "At home on the range? Lay interpretations of numerical uncertainty ranges." *Risk Analysis*, 35, pp. 1281–1295. doi:10.1111/risa.12358.

Driver, Rebecca, Nick Chater, Benny Cheung, Mark Latham, Rich Lewis, and Henry Stott (2010). "Helping consumers understand investment risk: Experimental research into the benefits of standardising risk disclosure." Research Paper 25, Association of British Insurers (ABI). URL https://www.dectech.co.uk/behavioural_science/public_research/dectech_research_abi.pdf.

Fisher, Ronald A. (1925). *Statistical methods for research workers.* Oliver and Boyd, London, 1 edition.

Galvão, Ana B. and James Mitchell (2020). "Real-time perceptions of historical GDP data uncertainty." EMF Research Papers 35, Economic Modelling and Forecasting Group. URL https://ideas.repec.org/p/wrk/wrkemf/35.html.

Haldane, Andrew and Michael McMahon (2018). "Central bank communications and the general public." *American Economic Review, Papers and Proceedings*, 108, pp. 578–83. doi:10.1257/pandp.20181082.

Handmer, John and Beth Proudley (2007). "Communicating uncertainty via probabilities: the case of weather forecasts." *Environmental Hazards*, 7, pp. 79–87. doi:10.1016/j.envhaz.2007.05.002.

Heckman, James (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." In *Annals of Economic and Social Measurement*, volume 5, pp. 475–492. National Bureau of Economic Research. URL http://www.nber.org/chapters/c10491.

Joslyn, Susan and Sonia Savelli (2010). "Communicating forecast uncertainty: public perception of weather forecast uncertainty." *Meteorological Applications*, 17, pp. 180–195. doi:10.1002/met.190.

Joslyn, Susan L. and Jared E. LeClerc (2012). "Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error." *Journal of Experimental Psychology: Applied*, 18, pp. 126–140. doi:10.1037/a0025185.

Joslyn, Susan L. and Jared E. LeClerc (2013). "Decisions with uncertainty: the glass half full." *Current Directions in Psychological Science*, 22(4), pp. 308–15. doi:10.1177/0963721413481473.

Manski, Charles F. (2004). "Measuring expectations." *Econometrica*, 72(5), pp. 1329–1376. doi:10.1111/j.1468-0262.2004.00537.x.

Manski, Charles F. (2015). "Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern." *Journal of Economic Literature*, 53(3), pp. 631–653. doi:10.1257/jel.53.3.631.

Manski, Charles F. (2016). "Credible interval estimates for official statistics with survey nonresponse." *Journal of Econometrics*, 191, pp. 293–301. doi:10.1016/j.jeconom.2015.12.002.

Manski, Charles F. (2019). "Communicating uncertainty in policy analysis." *Proceedings of the National Academy of Sciences*, 116(16), pp. 7634–7641. doi:10.1073/pnas.1722389115.

Manski, Charles F. and Francesca Molinari (2010). "Rounding probabilistic expectations in surveys." *Journal of Business and Economic Statistics*, 28(2), pp. 219–231. doi:10.1198/jbes.2009.08098.

Morss, Rebecca E., Jeffrey K. Lazo, and Julie L. Demuth (2010). "Examining the use of weather forecasts in decision scenarios: results from a US survey with implications for uncertainty communication." *Meteorological Applications*, 17, pp. 149–162. doi:10.1002/met.196.

Nadav-Greenberg, Limor, Susan L. Joslyn, and Meng U. Taing (2008). "The effect of uncertainty visualizations on decision making in weather forecasting." *Journal of Cognitive Engineering and Decision Making*, 20, pp. 24–47. doi:10.1518/155534308X284354.

Orphanides, Athanasios (2001). "Monetary policy rules based on real-time data." *American Economic Review*, 94, pp. 964–985. doi:10.1257/aer.91.4.964.

Padilla, Lace M. K., Grace Hansen, Ian Tanner Ruginski, Heidi Kramer, William B. Thompson, and Sarah H. Creem-Regehr (2015). "The influence of different graphical displays on non-expert decision making under uncertainty." *Journal of Experimental Psychology: Applied*, 21(1), pp. 37–46. doi:10.1037/xap0000037.

Raftery, Adrian E. (2016). "Use and communication of probabilistic forecasts." *Statistical Analysis and Data Mining*, 9, pp. 397–410. doi:10.1002/sam.11302.

Spiegelhalter, David, Mike Pearson, and Ian Short (2011). "Visualizing uncertainty about the future." *Science*, 333, pp. 1393–1400. doi:10.1126/science.1191181.

Tak, Susanne, Alexander Toet, and Jan van Erp (2015). "Public understanding of visual representations of uncertainty in temperature forecasts." *Journal of Cognitive Engineering and Decision Making*, 9, pp. 241–62. doi:10.1177/1555343415591275.

van der Bles, Anne Marthe, Sander van der Linden, Alexandra L. J. Freeman, James Mitchell, Ana B. Galvão, Lisa Zaval, and David J. Spiegelhalter (2019). "Communicating uncertainty about facts, numbers and science." *Royal Society Open Science*, 6, pp. 1–42. doi:10.1098/rsos.181870.

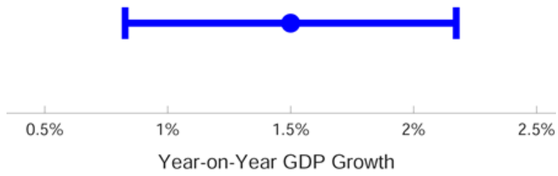
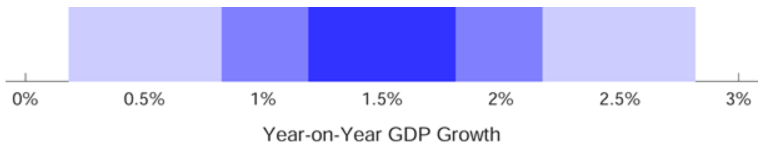
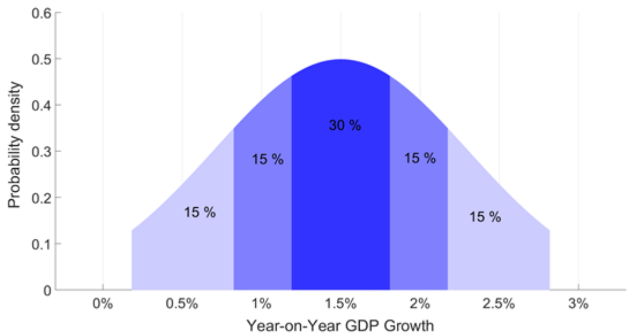
Visschers, Vivianne H. M. (2018). "Public perception of uncertainties within climate change science." *Risk Analysis*, 38, pp. 43–55. doi:10.1111/risa.12818.

Visschers, Vivianne H. M., Ree M. Meertens, Wim W. F. Passchier, and Nanne N. K. De Vries (2009). "Probability information in risk communication: A review of the research literature." *Risk Analysis*, 29, pp. 267–287. doi:10.1111/j.1539-6924.2008.

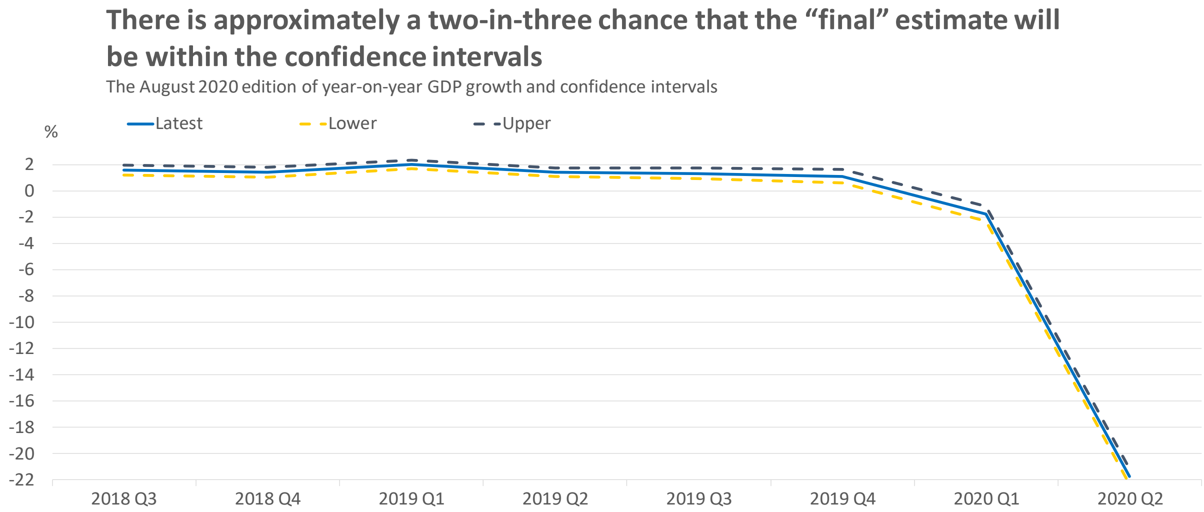
Young, Alwyn (2019). "Channeling Fisher: randomization tests and the statistical insignificance of seemingly significant experimental results." *Quarterly Journal of Economics*, 134, pp. 557–598. doi:10.1093/qje/qjy029.

Zarnowitz, Victor and Louis A. Lambros (1987). "Consensus and uncertainty in economic prediction." *Journal of Political Economy*, 95(3), pp. 591–621. doi:10.1086/261473.

Table 1: Data uncertainty communication tools

Group	Tool
G1	"GDP is estimated to have increased by 1.5% during the last year."
G2	"GDP is estimated to have increased by about 1.5% during the last year."
G3	"GDP is estimated to have increased by about 1.5% during the last year. But this estimate is likely to be revised as updated information becomes available."
G4	G3 phrase above **+** " - When this happens, it is still quite likely that GDP growth will be somewhere on the blue line between 0.8% and 2.2% (a 6 in 10 chance, or 60%). And it is less likely that GDP growth will be outside the blue line (a 4 in 10 chance, or 40%)." 
G5	G3 phrase above **+** " - When this happens, ONS estimates that GDP growth is most likely to be in the dark blue area (3 out of 10 times) and within each pair of lighter blue areas on a further 3 out of 10 occasions. ONS are very confident that GDP growth is somewhere in the total blue area, and will fall outside very rarely (1 out of 10 times)  The shading around the central estimate of 1.5% represents the uncertainty of the GDP estimates based on historical revisions, with 30%, 60% and 90% confidence intervals shown. The highlighted central estimate is the most likely value, while the values towards the upper and lower limit are possible but less likely. Other sources of uncertainty, for example due to limitations of the survey methodology, are not represented."
G6	G3 phrase above **+** " - When this happens, ONS estimates that GDP growth is most likely to be somewhere around 1.5% (where the graph is highest) but there is also a chance that GDP growth will be different. GDP growth is most likely to be in the dark blue area (3 out of 10 times), and within each pair of lighter blue areas on a further 3 out of 10 occasions. ONS are very confident that GDP growth is somewhere in the total blue area, and will fall outside very rarely (1 out of 10 times)." 
G7*	'GDP is estimated to have fallen by about 21.7% during the last year. GDP estimates are subject to more uncertainty than usual as a result of the challenges the ONS face in collecting the data under government imposed public health restrictions.'

G8*	‘GDP is estimated to have fallen by about 21.7% during the last year. But this estimate is likely to be revised as updated information becomes available. There is approximately a two-in-three chance that the “final” GDP estimate will be within the confidence intervals shown.’



Notes: In wave 2 (run in 2020), Groups 1 to 6 are shown equivalent communication tools but about the ONS point estimate of -21.7%. * The G7 and G8 communication tools feature in wave 2 only.

Table 2: Effects of the communication tools on qualitative and quantitative assessments of data uncertainty: Average treatment effects by Group (G#)

Outcomes	wave	G1	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
GDP Accuracy	2020	3.105	-0.046	**-0.107**	**-0.109**	**-0.061**	**-0.095**	**-0.154**	**-0.086**
q11: 1=very inaccurate to 4=very accurate			(0.033)	(0.034)	(0.033)	(0.032)	(0.031)	(0.035)	(0.034)
	2018	2.915	**-0.067**	-0.045	**-0.069**	-0.011	-0.015		
			(0.037)	(0.035)	(0.035)	(0.034)	(0.035)		
Certainty on GDP value	2020	4.580	-0.020	**-0.313**	**-0.408**	**-0.304**	**-0.332**	**-0.191**	**-0.256**
q14: 1=exceptionally unlikely (1% chance) to 7=virtually certain (99% chance)			(0.084)	(0.086)	(0.086)	(0.083)	(0.081)	(0.087)	(0.087)
	2018	4.333	-0.125	-0.001	**-0.185**	-0.080	**-0.193**		
			(0.083)	(0.079)	(0.081)	(0.080)	(0.080)		
Informative Comms Tool	2020	2.284	-0.007	**0.311**	**0.443**	**0.415**	**0.382**	**0.226**	**0.399**
q19: 1=not at all to 4=a lot			(0.046)	(0.046)	(0.046)	(0.045)	(0.046)	(0.045)	(0.046)
	2018	2.276	-0.042	**0.299**	**0.376**	**0.410**	**0.435**		
			(0.048)	(0.046)	(0.047)	(0.047)	(0.046)		
Range Interval*	2020	12.553	-1.067	-0.303	**-4.693**	**-4.495**	-1.282	0.998	**-2.198**
q12-q13			(0.941)	(1.318)	(0.992)	(0.971)	(1.680)	(2.289)	(0.978)
	2018	2.710	-0.789	0.008	0.124	-0.412	**2.348**		
			(0.452)	(0.654)	(0.600)	(0.555)	(0.800)		
Prob (GDP bet. bounds) = 30%	2020	0.095	0.016	**0.043**	**0.066**	0.026	**0.068**	0.015	**0.053**
			(0.019)	(0.019)	(0.020)	(0.019)	(0.021)	(0.018)	(0.020)
q15: Binary variable=1 for a correct answer	2018	0.102	-0.006	0.002	**0.038**	0.026	**0.035**		
			(0.019)	(0.019)	(0.021)	(0.020)	(0.020)		

Notes: q# refers to the survey question number (see Appendix A). Robust standard errors in parentheses. Treatment effects for G2-G8 in bold when statistically significant at 10%. N=4,201 in the 2020 wave and N=3,045 in the 2018 wave. * For q12 and q13, N=2,582 in the 2020 wave and N=1,736 in the 2018 wave, as not all individuals replied to these questions (individuals who reported lower bound higher than the upper bound are also dropped from the analysis). Group 1 (G1) is the average outcome for the control group shown the current headline ONS GDP point estimate press release.

Table 3: The effects of communication tools on probabilistic perceptions of data uncertainty: Quantile treatment effects for CM distance by Group (G#)

Outcome: CM distance between the individual histograms and the histogram that underlies the communication tools							
quantile	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
25%	-0.003 (-0.56)	0.005 (0.89)	**-0.009** (-2.09)	**-0.021** (-4.09)	**-0.011** (-2.59)	-0.005 (-1.03)	-0.006 (-1.34)
median	0.000 (0.02)	0.016 (1.28)	**-0.024** (-2.62)	**-0.029** (-3.27)	**-0.022** (-2.33)	0.003 (0.29)	0.008 (0.71)
75%	0.049 (1.00)	0.050 (1.17)	**-0.056** (-1.45)	**-0.074** (-2.19)	-0.061 (-1.55)	0.007 (0.19)	-0.020 (-0.56)

Notes: 2020 wave. Robust t-stats in parentheses. Values in bold indicate that treatment effect is statistically significant at the 10% level using quantile regressions at the indicated quantile. N=4,201.

Table 4: The effects of communication tools on knowledge about data revisions: Average treatment effects by Group (G#)

Outcomes	Wave N	G1	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Revisions Awareness q18: 1=very surprised to 4=not at all surprised	2020 N=3,594	3.164	**0.103** (0.051)	0.011 (0.055)	0.064 (0.054)	0.034 (0.054)	0.012 (0.055)	0.038 (0.054)	-0.003 (0.058)
	2018 N=2,583	3.084	0.013 (0.056)	0.020 (0.056)	0.063 (0.057)	0.063 (0.055)	-0.016 (0.056)		
No awareness q18: Binary variable=1 when no awareness	2020 N=4,201	0.139	-0.020 (0.021)	0.003 (0.021)	0.009 (0.022)	0.017 (0.022)	-0.013 (0.021)	0.034 (0.022)	0.011 (0.022)
	2018 N=3,045	0.158	-0.006 (0.023)	-0.028 (0.022)	0.008 (0.023)	-0.012 (0.023)	0.001 (0.023)		
Revisions due to more info q17: Binary variable=1, 0 otherwise	2020 N=4,201	0.568	0.027 (0.030)	0.028 (0.030)	0.025 (0.030)	0.037 (0.030)	0.028 (0.030)	-0.004 (0.030)	0.046 (0.030)
	2018 N=3,045	0.502	-0.010 (0.031)	0.040 (0.031)	0.029 (0.031)	**0.059** (0.031)	0.051 (0.031)		
Revisions due to vested interests q17: Binary variable=1, 0 otherwise	2020 N=4,201	0.267	-0.024 (0.031)	-0.014 (0.032)	-0.034 (0.031)	0.010 (0.032)	0.019 (0.033)	0.016 (0.032)	0.001 (0.033)
	2018 N=3,045	0.310	-0.018 (0.033)	-0.012 (0.034)	-0.017 (0.034)	-0.037 (0.033)	-0.026 (0.033)		
Revisions due to ONS mistakes q17: Binary variable=1, 0 otherwise	2020 N=4,201	0.118	-0.023 (0.019)	-0.009 (0.019)	-0.017 (0.019)	-0.019 (0.019)	0.004 (0.020)	-0.002 (0.019)	-0.002 (0.019)
	2018 N=3,045	0.101	-0.014 (0.018)	-0.023 (0.017)	-0.003 (0.018)	-0.012 (0.018)	-0.008 (0.018)		

Notes: q# refers to the survey question number (see Appendix A). Robust standard errors in parentheses. Estimates in bold are statistically significant at 10%. Group 1 (G1) is the average outcome for the control group shown the current headline ONS GDP point estimate press release.

Table 5: Effects of the communication tools on qualitative and quantitative assessments of data uncertainty: Average treatment effects by Group (G#), conditional on respondents having heard of and trusting the ONS and correctly knowing what GDP measures

Outcomes	wave	Condition	G1	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
GDP Accuracy q11: 1=very inaccurate to 4=very accurate	2020	Know	3.169	-0.035	-0.082	**-0.118**	0.016	-0.085	**-0.181**	-0.082
		N=1455		(0.053)	(0.055)	(0.055)	(0.055)	(0.052)	(0.057)	(0.057)
		Don't	2.945	-0.103	**-0.178**	-0.035	-0.096	-0.122	**-0.217**	-0.109
		N=691		(0.083)	(0.080)	(0.091)	(0.075)	(0.082)	(0.099)	(0.086)
	2018	Know	2.992	0.031	-0.063	**-0.121**	0.008	-0.043		
		N=771		(0.053)	(0.054)	(0.059)	(0.055)	(0.062)		
		Don't	2.692	-0.094	-0.068	-0.022	0.076	0.039		
		N=670		(0.082)	(0.083)	(0.084)	(0.075)	(0.074)		
Certainty GDP value q14: 1=exceptionally unlikely (1% chance) to 7=virtually certain (99% chance)	2020	Know	4.746	-0.054	**-0.530**	**-0.735**	**-0.413**	**-0.509**	**-0.420**	**-0.740**
		N=1455		(0.146)	(0.160)	(0.160)	(0.150)	(0.153)	(0.159)	(0.163)
		Don't	4.193	-0.010	**-0.445**	-0.073	-0.251	-0.155	-0.149	-0.015
		N=691		(0.171)	(0.174)	(0.196)	(0.180)	(0.166)	(0.196)	(0.202)
	2018	Know	4.515	-0.149	-0.226	**-0.475**	**-0.308**	**-0.414**		
		N=771		(0.148)	(0.149)	(0.155)	(0.141)	(0.159)		
		Don't	3.880	-0.198	0.019	-0.107	0.144	-0.040		
		N=670		(0.178)	(0.158)	(0.172)	(0.154)	(0.159)		
Informative Comms Tool q19: 1=not at all to 4=a lot	2020	Know	2.365	-0.110	**0.354**	**0.567**	**0.520**	**0.539**	**0.320**	**0.501**
		N=1455		(0.071)	(0.077)	(0.076)	(0.072)	(0.072)	(0.074)	(0.075)
		Don't	2.174	0.131	**0.301**	**0.378**	**0.384**	**0.269**	0.174	0.168
		N=691		(0.114)	(0.111)	(0.123)	(0.112)	(0.115)	(0.118)	(0.128)
	2018	Know	2.308	-0.086	**0.434**	**0.620**	**0.628**	**0.599**		
		N=771		(0.089)	(0.088)	(0.085)	(0.085)	(0.086)		
		Don't	2.359	-0.181	0.036	0.136	0.029	0.170		
		N=670		(0.116)	(0.106)	(0.116)	(0.102)	(0.105)		
Range Interval* q12-q13	2020	Know	11.070	0.229	-0.605	**-3.872**	**-5.091**	0.244	-0.772	-1.921
		N=1142		(0.976)	(0.970)	(1.136)	(0.889)	(3.427)	(0.992)	(1.031)
		Don't	18.561	**-6.965**	3.086	**-8.502**	**-6.961**	**-8.389**	26.789	**-7.521**
		N=218		(3.345)	(8.690)	(3.781)	(3.709)	(3.000)	(30.953)	(3.415)
	2018	Know	1.963	-0.659	-0.882	0.084	-0.028	0.545		
		N=607		(0.784)	(0.748)	(1.067)	(0.980)	(0.899)		
		Don't	2.869	-0.066	**-1.730**	2.798	0.122	0.418		
		N=203		(1.318)	(1.026)	(2.611)	(1.402)	(1.365)		
Prob (GDP bounds) = 30% q15: Binary variable=1 for a correct answer	2020	Know	0.085	0.040	**0.108**	**0.113**	**0.079**	**0.119**	**0.078**	**0.072**
		N=1455		(0.031)	(0.036)	(0.036)	(0.034)	(0.037)	(0.034)	(0.034)
		Don't	0.064	0.033	0.062	**0.085**	0.006	**0.100**	0.066	0.018
		N=691		(0.041)	(0.041)	(0.050)	(0.036)	(0.048)	(0.042)	(0.040)
	2018	Know	0.069	0.030	-0.007	0.044	**0.074**	**0.117**		
		N=771		(0.034)	(0.031)	(0.036)	(0.037)	(0.042)		
		Don't	0.111	-0.008	0.017	0.013	0.029	-0.002		
		N=670		(0.041)	(0.043)	(0.044)	(0.043)	(0.041)		
CM distance q16	2020	Know	0.116	-0.012	-0.006	**-0.041**	**-0.046**	**-0.035**	0.008	-0.001
		N=1455		(0.020)	(0.023)	(0.018)	(0.018)	(0.020)	(0.021)	(0.021)
		Don't	0.090	**0.034**	**0.059**	0.007	-0.003	0.025	**0.045**	0.067
		N=691		(0.019)	(0.028)	(0.020)	(0.015)	(0.027)	(0.026)	(0.044)

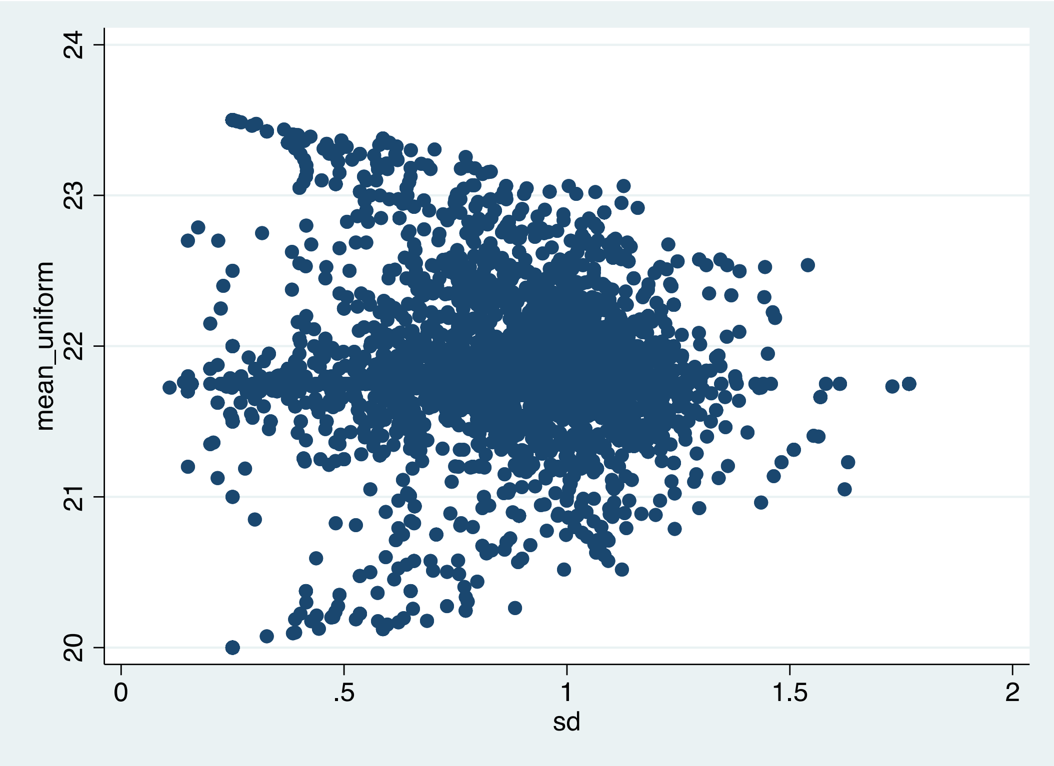
Notes: q# refers to the survey question number (see Appendix A). "Know" refers to individuals who have heard of the ONS, trust the ONS, and correctly identified what GDP measures. "Don't know" refers to individuals who have not heard of the ONS, do not trust the ONS, and who incorrectly identified what GDP measures. Robust standard errors in parentheses. Treatment effects in bold when statistically significant at 10%. Group 1 (G1) is the average outcome for the control group shown the current headline ONS GDP point estimate press release. The CM treatment effects are for the 50% quantile.

Table 6: Effects of the communication tools on knowledge of data revisions: Average treatment effects by group (G#), conditional on respondents having heard of and trusting the ONS and correctly knowing what GDP measures

Outcomes	wave	Condition	G1	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Revisions Awareness q18: 1=very surprised to 4=not at all surprised	2020	Know N=1406	3.409	0.027	0.030	0.062	0.070	0.046	-0.010	0.067
				(0.070)	(0.077)	(0.072)	(0.076)	(0.075)	(0.081)	(0.078)
		Don't N=376	2.937	0.170	0.147	0.020	0.118	-0.022	0.194	0.230
				(0.128)	(0.140)	(0.161)	(0.153)	(0.147)	(0.130)	(0.164)
	2018	Know N=752	3.405	-0.116	0.030	0.099	0.033	-0.060		
				(0.087)	(0.091)	(0.081)	(0.082)	(0.089)		
		Don't N=383	3.065	-0.081	0.104	-0.024	-0.022	-0.079		
				(0.146)	(0.132)	(0.147)	(0.132)	(0.135)		
No awareness q18: Binary variable=1 when no awareness	2020	Know N=1455	0.016	0.013	0.025	0.029	0.041	0.011	0.007	0.019
				(0.015)	(0.018)	(0.018)	(0.020)	(0.015)	(0.014)	(0.017)
		Don't N=691	0.422	0.005	-0.005	0.078	0.026	0.031	-0.004	**0.167**
				(0.073)	(0.068)	(0.071)	(0.077)	(0.072)	(0.073)	(0.075)
	2018	Know N=771	0.031	-0.008	0.016	-0.023	-0.009	-0.014		
				(0.020)	(0.024)	(0.017)	(0.020)	(0.019)		
		Don't N=670	0.470	-0.021	-0.121	0.025	-0.057	-0.067		
				(0.067)	(0.065)	(0.069)	(0.065)	(0.065)		
Revisions due to more info q17: Binary variable=1, 0 otherwise	2020	Know N=1455	0.825	**-0.080**	-0.024	-0.011	0.016	-0.000	0.005	0.012
				(0.041)	(0.041)	(0.040)	(0.039)	(0.039)	(0.040)	(0.040)
		Don't N=691	0.239	0.018	0.043	0.022	0.015	0.041	0.027	0.063
				(0.064)	(0.061)	(0.062)	(0.067)	(0.064)	(0.065)	(0.068)
	2018	Know N=771	0.792	-0.037	-0.003	0.022	0.036	0.030		
				(0.052)	(0.051)	(0.050)	(0.048)	(0.050)		
		Don't N=670	0.214	0.029	0.080	-0.018	0.092	**0.122**		
				(0.056)	(0.058)	(0.056)	(0.057)	(0.058)		
Revisions due to vested interests q17: Binary variable=1, 0 otherwise	2020	Know N=1455	0.201	0.020	-0.002	-0.016	0.002	0.001	-0.009	-0.056
				(0.048)	(0.050)	(0.049)	(0.049)	(0.049)	(0.048)	(0.045)
		Don't N=691	0.266	-0.095	**-0.111**	-0.070	-0.042	-0.068	**-0.114**	**-0.129**
				(0.068)	(0.065)	(0.067)	(0.070)	(0.069)	(0.066)	(0.069)
	2018	Know N=771	0.185	0.037	0.003	0.001	0.008	0.010		
				(0.057)	(0.051)	(0.054)	(0.054)	(0.058)		
		Don't N=670	0.256	-0.060	-0.045	-0.009	-0.083	-0.038		
				(0.064)	(0.063)	(0.071)	(0.059)	(0.063)		
Revisions due to ONS mistakes q17: Binary variable=1, 0 otherwise	2020	Know N=1455	0.106	-0.010	-0.018	0.046	0.002	0.003	0.007	-0.007
				(0.030)	(0.031)	(0.035)	(0.032)	(0.032)	(0.033)	(0.032)
		Don't N=691	0.128	**-0.104**	-0.031	**-0.074**	**-0.099**	-0.024	-0.002	-0.060
				(0.037)	(0.044)	(0.040)	(0.038)	(0.046)	(0.050)	(0.044)
	2018	Know N=771	0.100	-0.039	-0.022	-0.035	**-0.057**	-0.015		
				(0.034)	(0.036)	(0.034)	(0.032)	(0.037)		
		Don't N=670	0.026	0.040	-0.007	**0.077**	**0.049**	0.008		
				(0.028)	(0.020)	(0.034)	(0.028)	(0.022)		

Notes: See notes to Table 5.

Figure 1: Mean and standard deviation of respondents' reported histogram estimates of GDP data uncertainty



Notes: Mean and standard deviation (sd) calculated non-parametrically from the responses to question 16 (see Appendix A). Individuals with 100% probability in 1 bin removed.

Online Appendix A: Questions and summary statistics for wave 1 and 2 surveys

Wave 1: surveyed November 2018: N=3150. Wave 2: surveyed August 2020: N=4201.
&: indicates questions where the respondent could choose more than one answer.

	Wave 1		Wave 2	
	Count	**%**	**Count**	**%**
Q1. What is your gender?				
Male	1490	48.93%	2045	48.68%
Female	1548	50.84%	2137	50.87%
Other (please specify)	3	0.10%	4	0.10%
Prefer not to state	4	0.13%	15	0.36%
Q2. What is your age?				
18-24	357	11.72%	546	13.00%
25-34	556	18.26%	663	15.78%
35-44	513	16.85%	719	17.11%
45-54	521	17.11%	748	17.81%
55-64	479	15.73%	618	14.71%
65 and above	619	20.33%	907	21.59%
Q3. Where do you live?				
East of England	273	8.97%	351	8.36%
East Midlands	224	7.36%	308	7.33%
London	369	12.12%	563	13.40%
North East	125	4.11%	191	4.55%
North West	346	11.36%	455	10.83%
Northern Ireland	69	2.27%	128	3.05%
Scotland	246	8.08%	351	8.36%
South East	450	14.78%	577	13.73%
South West	264	8.67%	350	8.33%
Wales	150	4.93%	221	5.26%
West Midlands	265	8.70%	378	9.00%
Yorkshire & Humberside	264	8.67%	328	7.81%
Q4. What is your highest educational qualification?				
PhD or equivalent doctoral level qualification	81	2.66%	133	3.17%
Masters or equivalent higher degree level qualification (MA, MSc, PGCE etc.)	294	9.66%	478	11.38%
Bachelors or equivalent degree level qualification (BA, BSc etc.)	680	22.33%	1113	26.49%
Post-secondary below-degree level qualification	264	8.67%	357	8.50%
A Level / NVQ Level 3	708	23.25%	889	21.16%
GCSE / O Level / NVQ Level 1 / NVQ Level 2	769	25.25%	892	21.23%
CSE	74	2.43%	97	2.31%
Any other qualification	58	1.90%	82	1.95%
None of the above	117	3.84%	160	3.81%

	Wave 1		Wave 2	
	Count	%	Count	%
Q5. What's your current employment status?				
Employed full-time	1176	38.62%	1604	38.18%
Employed part-time	448	14.71%	522	12.43%
Unemployed and currently looking for work	136	4.47%	211	5.02%
Unemployed and not currently looking for work	235	7.72%	225	5.36%
Retired	671	22.04%	937	22.30%
Self-employed	113	3.71%	166	3.95%
Unable to work	131	4.30%	169	4.02%
Student	135	4.40%	223	5.31%
Furloughed (from full-time job)	n/a	n/a	76	1.81%
Furloughed (from part-time job)	n/a	n/a	68	1.62%
Q6. In which, if any, have you ever studied economics?[&]				
At school	819	26.90%	1047	24.92%
In higher education (e.g. university, college)	719	23.61%	955	22.73%
Through self-directed study (books)	186	6.11%	279	6.64%
Self-motivated study (course)	186	6.11%	230	5.47%
Other	26	0.85%	24	0.57%
Don't know / can't recall	97	3.19%	164	3.90%
Not applicable – I have never studied economics	1346	44.20%	1,949	46.39%
Q7. How frequently do you read/watch/listen to news stories related to economics or the economy?				
Never	227	7.45%	348	8.28%
Rarely	557	18.29%	797	18.97%
Monthly	292	9.59%	514	12.24%
Weekly	748	24.56%	1024	24.38%
Almost every day	732	24.04%	942	22.42%
Every day	372	12.22%	392	9.33%
Not sure	117	3.84%	184	4.38%
Q8. The Office for National Statistics (ONS) is the UK's largest independent producer of official statistics and the recognised national statistical institute of the UK. Before answering this survey, had you ever heard of the ONS?				
Yes, I had heard of them, and knew what they did	1480	48.60%	2427	57.77%
Yes, I had heard of them, but didn't know what they did	797	26.17%	983	23.40%
No, I had never heard of them	598	19.64%	599	14.26%
Not sure / don't know	170	5.58%	192	4.57%
Q9. Personally, how much trust do you have in economic statistics produced by the Office for National Statistics (ONS)? For example, on unemployment, inflation or economic growth?				
Trust them greatly	349	11.46%	591	14.07%
Tend to trust them	1566	51.43%	2346	55.84%
Tend not to trust them	414	13.60%	429	10.21%
Distrust them greatly	65	2.13%	74	1.76%
Not sure / don't know	651	21.38%	761	18.11%

	Wave 1		Wave 2	
	Count	%	Count	%
Q10. To the best of your knowledge, which option most accurately describes what GDP is?				
GDP measures the increase in prices	247	8.11%	288	6.86%
GDP measures how many people are in employment	200	6.57%	208	4.95%
GDP measures the size of the economy	1405	46.14%	2308	54.94%
GDP measures the difference between exports and imports	352	11.56%	421	10.02%
I don't have a clue what GDP is	462	15.17%	499	11.88%
I have heard about GDP but not sure what it is	379	12.45%	477	11.35%
Random allocation to a group – each group shown their allocated communication tool. See Table 1				
GROUP1	507	16.65%	524	12.47%
GROUP2	508	16.68%	527	12.54%
GROUP3	508	16.68%	526	12.52%
GROUP4	506	16.62%	525	12.50%
GROUP5	507	16.65%	525	12.50%
GROUP6	509	16.72%	524	12.47%
GROUP7	n/a	n/a	525	12.50%
GROUP8	n/a	n/a	525	12.50%
Q11. How accurate do you think the first estimate of GDP growth of 1.5% is likely to be? *(wave 1)* How accurate do you think the estimate that GDP fell by 21.7% is likely to be? *(wave 2)*				
Very accurate	261	8.57%	634	15.09%
Fairly accurate	2205	72.41%	3074	73.17%
Not very accurate	533	17.50%	447	10.64%
Very inaccurate	46	1.51%	46	1.09%
Q12. I would not be surprised if actual GDP growth was as high as: *(wave 1)* I would not be surprised if actual GDP fell by as much as: (*wave 2*)				
Don't know	1025	33.66%	1027	24.45%
Q13. I would not be surprised if actual GDP growth was as low as: *(wave 1)* I would not be surprised if actual GDP fell by as little as: (*wave 2*)				
Don't know	1085	35.63%	1310	31.18%
Q14. What do you think is the chance that GDP grew (fell) by exactly 1.5% (21.7%)?				
Virtually certain – about a 99 in 100 chance (99%)	80	2.63%	141	3.36%
Very likely – about a 9 in 10 chance (90%)	399	13.10%	702	16.71%
Quite likely – about a 6 in 10 chance (60%)	808	26.54%	1339	31.87%
Fifty-fifty – about a 1 in 2 chance (50%)	1018	33.43%	993	23.64%
Quite unlikely – about a 3 in 10 chance (30%)	474	15.57%	571	13.59%
Very unlikely – about a 1 in 10 chance (10%)	144	4.73%	249	5.93%
Exceptionally unlikely – about a 1 in 100 chance (1%)	122	4.01%	206	4.90%

	Wave 1		Wave 2	
	Count	%	Count	%
Q15: What do you think is the chance that GDP grew by between 1.2% and 1.8%? *(wave 1)* Q15: What do you think is the chance that GDP fell by between 21.4% and 22.0%? *(wave 2)*				
Virtually certain – about a 99 in 100 chance (99%)	152	4.99%	168	4.00%
Very likely – about a 9 in 10 chance (90%)	549	18.03%	773	18.40%
Quite likely – about a 6 in 10 chance (60%)	836	27.45%	1357	32.30%
Fifty-fifty – about a 1 in 2 chance (50%)	941	30.90%	1063	25.30%
Quite unlikely – about a 3 in 10 chance (30%)	360	11.82%	553	13.16%
Very unlikely – about a 1 in 10 chance (10%)	128	4.20%	182	4.33%
Exceptionally unlikely – about a 1 in 100 chance (1%)	79	2.59%	105	2.50%
Q16: Please provide (best-guess) estimates of the percentage probabilities you would attach to various outcomes for GDP growth during the last year. The probabilities should sum to 100% as indicated.[&]				
			Average answer for each bin:	
Fall by 23% or more			16.73%	
Fall by 22.5% to 23%			11.33%	
Fall by 22% to 22.5%			12.28%	
Fall by 21.5% to 22%			27.56%	
Fall by 21% to 21.5%			12.02%	
Fall by 20.5% to 21%			9.16%	
Fall by 20.5% or less			10.93%	
Q17. ONS regularly publishes revisions to their GDP estimates. Why do you think they do this? [&]				
Mistakes at the ONS	275	9.03%	460	10.95%
More information becomes available	1617	53.10%	2488	59.22%
The ONS has vested interests in results / manipulates production or collection	280	9.20%	382	9.09%
The Government has vested interests in the results / interferes in production or collection	606	19.90%	727	17.31%
Limitations to the way GDP is measured	607	19.93%	989	23.54%
Other [please write any other reasons]	25	0.82%	36	0.86%
Don't know / not sure	533	17.50%	710	16.90%
Q18. Are you surprised that estimates of GDP growth are regularly revised?				
Very surprised	107	3.51%	149	3.55%
Fairly surprised	413	13.56%	487	11.59%
Not that surprised	1157	38.00%	1465	34.87%
Not at all surprised	906	29.75%	1493	35.54%
N/A. I had never thought about it before doing this survey	462	15.17%	607	14.45%
Q19. Thinking back to the ONS statement about GDP growth, how much information did it give that the 1.5% estimate may be uncertain?				
None at all	259	8.51%	315	7.50%
Very little	1193	39.18%	1605	38.21%
Some	1336	43.88%	1914	45.56%
A lot	257	8.44%	367	8.74%

Online Appendix B: Supplementary Empirical Results

Table B1. Effects of the communication tools on qualitative and quantitative assessments of data uncertainty: *p*-values by group (G#) for randomized tests for zero treatment effects

Outcomes	wave	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
GDP Accuracy	2020	0.16	0.00	0.00	0.06	0.00	0.00	0.01
q11: 1=very inaccurate to 4=very accurate	2018	0.07	0.21	0.05	0.73	0.65	-	-
Certainty GDP value	2020	0.81	0.00	0.00	0.00	0.00	0.03	0.00
q14: 1=exceptionally unlikely (1% chance) to 7=virtually certain (99% chance)	2018	0.13	0.99	0.02	0.31	0.02	-	-
Informative Comms Tool	2020	0.87	0.00	0.00	0.00	0.00	0.00	0.00
q19: 1=not at all to 4=a lot	2018	0.39	0.00	0.00	0.00	0.00		
Range Interval*	2020	0.25	0.83	0.00	0.00	0.52	0.82	0.02
q12-q13	2018	0.08	0.99	0.83	0.45	0.00	-	-
Prob (GDP bounds) = 30%	2020	0.36	0.03	0.00	0.19	0.00	0.47	0.01
q15: Binary variable=1 for a correct answer	2018	0.73	0.97	0.07	0.20	0.09	-	-
CM distance q16	2020	0.39	0.04	0.21	0.03	0.23	0.89	0.75

Notes: q# refers to the survey question number (see Appendix A). Randomized p-value from Young (2019) randomized-t test with 5000 replications. CM distance is winsorized at the 10% level. *For q12 and q13, N=2,582 in the 2020 wave and N=1,736 in the 2018 wave as not all individuals replied to these questions. Individuals who reported lower bound higher than the upper bound are also dropped from analysis.

Table B2: Effects of the communication tools on knowledge about data revisions: *p*-values by group (G#) for randomized tests for zero treatment effects

Outcomes	wave	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Revisions Awareness	2020	0.05	0.84	0.23	0.53	0.82	0.48	0.96
q18: 1=very surprised to 4=not at all surprised	2018	0.82	0.74	0.27	0.25	0.77		
Revisions due to more info	2020	0.35	0.36	0.40	0.22	0.38	0.89	0.13
q17: Binary variable=1, 0 otherwise	2018	0.74	0.20	0.35	0.06	0.11		
Revisions due to vested interests	2020	0.43	0.66	0.28	0.75	0.60	0.63	0.98
q17: Binary variable=1, 0 otherwise	2018	0.58	0.71	0.61	0.27	0.42		
Revisions due to ONS mistakes	2020	0.20	0.60	0.36	0.31	0.92	0.89	0.90
q17: Binary variable=1, 0 otherwise	2018	0.43	0.17	0.88	0.51	0.60		

Note: q# refers to the survey question number (see Appendix A). Randomized p-value from Young (2019) randomized-t test with 5000 replications.

Table B3. Effects of the communication tools on qualitative and quantitative assessments of data uncertainty: Romano-Wolf joint tests for zero average treatment effects across the five uncertainty outcomes: *p*-values by Group (G#)

Outcomes		G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Accuracy of GDP	2020	0.59	0.00	0.00	0.13	0.01	0.00	0.03
q11: 1=very inaccurate to 4=very accurate	2018	0.28	0.56	0.14	0.73	0.65		
Certainty on GDP value	2020	0.96	0.00	0.00	0.00	0.00	0.22	0.02
q14: 1=exceptionally unlikely (1% chance) to 7=virtually certain (99% chance)	2018	0.36	1.00	0.08	0.66	0.05		
Informative Comms Tool	2020	0.96	0.00	0.00	0.00	0.00	0.00	0.00
q19: 1=not at all to 4=a lot	2018	0.63	0.00	0.00	0.00	0.00		
Range interval*	2020	0.68	0.83	0.00	0.00	0.49	0.73	0.03
q12-q13	2018	0.28	1.00	0.84	0.70	0.01		
Chance between	2020	0.75	0.06	0.00	0.17	0.01	0.73	0.03
L and U is 30% 2020 q15: Binary variable=1 for a correct answer	2018	0.74	0.99	0.14	0.58	0.16		

Notes: q# refers to the survey question number (see Appendix A). Romano-Wolf step-down adjusted p-values with 5000 replications. * For q12 and q13, N=2,582 in the 2020 wave and N=1,736 in the 2018 wave as not all individuals replied to these questions. Individuals who reported lower bound higher than the upper bound are also dropped from analysis. Romano-Wolf test implemented as in Stata; see Clarke et al. (2020).

Table B4. Effects of the communication tools on knowledge of data revisions: Romano-Wolf joint tests for zero average treatment effects across the revisions outcomes: *p*-values by Group (G#)

Outcomes	Wave	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Revisions Awareness	2020	0.16	0.94	0.65	0.76	0.97	0.92	1.00
q18: 1=very surprised to 4=not at all surprised	2018	0.93	0.92	0.73	0.58	0.88		
Revisions due to more info	2020	0.60	0.82	0.65	0.61	0.81	0.98	0.39
q17: Binary variable=1, 0 otherwise	2018	0.93	0.54	0.74	0.20	0.34		
Revisions due to vested interests	2020	0.60	0.94	0.65	0.76	0.92	0.94	1.00
q17: Binary variable=1, 0 otherwise	2018	0.93	0.92	0.85	0.58	0.81		
Revisions due to ONS mistakes	2020	0.51	0.94	0.65	0.67	0.97	0.98	1.00
q17: Binary variable=1, 0 otherwise	2018	0.90	0.54	0.85	0.58	0.88		

Note: q# refers to the survey question number (see Appendix A). Romano-Wolf step-down adjusted p-values with 5000 replications. Romano-Wolf test implemented as in Stata; see Clarke et al. (2020).

Table B5: Effects of the communication tools on the range interval outcome: Heckman corrected treatment effects and selection equations by Group (G#)

Treatment effect	wave	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Range interval q12-q13	2020	-0.845 (-0.58)	-0.407 (-0.27)	-4.396 (-2.89)	-4.422 (-2.89)	-1.158 (-0.78)	1.220 (0.80)	-2.092 (-1.41)
	2018	-0.771 (-1.20)	-0.0671 (-0.10)	0.0799 (0.12)	-0.454 (-0.71)	2.311 (3.51)		

Selection	wave	man	young	grad	grad econ	Full time	Freq-news	Know ONS	Trust ONS	Know GDP
Range interval q12-q13	2020	0.133 (3.04)	-0.045 (-0.97)	0.072 (1.51)	-0.004 (-0.09)	0.105 (2.41)	0.206 (4.31)	0.326 (6.92)	0.393 (8.07)	0.387 (8.78)
	2018	0.210 (4.07)	-0.115 (-2.10)	0.165 (2.75)	-0.026 (-0.44)	-0.087 (-1.72)	0.127 (2.34)	0.277 (5.01)	0.483 (9.01)	0.394 (7.68)

Notes: q# refers to the survey question number (see Appendix A). t-stats in parentheses. Wave 1 (selected = 1,736; nonselected=1,309). Wave 2 (selected=2,582; nonselected=1,619). The variables in the selection equation are nine dummy variables equal to unity, zero otherwise, capturing, in turn, when the respondent is male, young (age 34 or less), is a graduate (bachelor's degree), is a graduate in economics, works full-time, follows the economic news at least almost every day, has heard of the ONS, trusts the ONS, and knows what GDP measures (as identified by answering question 10 correctly).

Table B6: Effects of the communication tools on qualitative and quantitative assessments of data uncertainty: P-values of non-parametric tests for zero conditional, constant and zero average treatment effects by Group (G#)

		Selected variables	G2: textual 'about'			G3: likely revised			G4: interval			G5: density strip			G6: bell curve			G7: Covid effects			G8: time interval		
	wave		CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE
GDP Accuracy (q11)	2018	trust ONS	0.18	0.89	0.06	0.42	0.86	0.20	0.01	0.06	0.05	0.40	0.21	0.74	0.61	0.33	0.66	-	-	-	-	-	-
	2020	trust ONS full-time young	0.08	0.21	0.17	0.04	0.67	0.00	0.00	0.02	0.00	0.03	0.07	0.06	0.01	0.74	0.00	0.00	0.41	0.00	0.03	0.32	0.01
Certainty on GDP value (q14)	2018	trust ONS male grad econ	0.56	0.87	0.13	0.34	0.21	0.99	0.01	0.11	0.02	0.30	0.33	0.31	0.07	0.26	0.02	-	-	-	-	-	-
	2020	trust ONS full-time young	0.44	0.44	0.81	0.01	0.79	0.00	0.00	0.10	0.00	0.00	0.10	0.00	0.00	0.04	0.00	0.03	0.19	0.03	0.00	0.06	0.00
Comm. Informative (q19)	2018	grad econ	0.10	0.05	0.39	0.00	0.11	0.00	0.00	0.26	0.00	0.00	0.19	0.00	0.00	0.35	0.00	-	-	-	-	-	-
	2020	freq-news	0.25	0.10	0.87	0.00	0.05	0.00	0.00	0.22	0.00	0.00	0.20	0.00	0.00	0.53	0.00	0.00	0.75	0.00	0.00	0.52	0.00
Range interval (q13-q12)	2018	young full-time	0.10	0.05	0.08	0.11	0.05	0.99	0.11	0.11	0.84	0.57	0.36	0.46	0.00	0.38	0.00	-	-	-	-	-	-
	2020	freq-news	0.42	0.26	0.51	0.08	0.03	0.53	0.00	0.11	0.00	0.00	0.27	0.00	0.01	0.19	0.00	0.60	0.34	0.98	0.05	0.64	0.01
Prob(GDP bet. Bounds)= 30% (q15)	2018	trust ONS	0.87	0.76	0.75	0.92	0.70	0.93	0.11	0.43	0.07	0.38	0.67	0.20	0.19	0.55	0.09	-	-	-	-	-	-
	2020	freq-news grad	0.10	0.20	0.38	0.00	0.00	0.03	0.00	0.26	0.00	0.00	0.00	0.16	0.00	0.03	0.00	0.01	0.01	0.42	0.00	0.02	0.01
CM dist (q16)	2020	freq-news	0.28	0.18	0.40	0.01	0.03	0.04	0.14	0.14	0.21	0.01	0.05	0.03	0.29	0.34	0.23	0.45	0.21	0.89	0.03	0.01	0.75

Notes: *p*-values from the chi-squared test of Crump et al. (2008). "CATE" is their non-parametric test for zero conditional average treatment effect, i.e., the test of no treatment effect for all values of the covariates. The covariates capture individual characteristics, specifically whether the respondent is male, young (age 34 or less), is a graduate (bachelor's degree), is a graduate in economics, works full-time, follows the economic news at least almost every day, has heard of the ONS, trusts the ONS, and knows what GDP measures (as identified by answering question 10 correctly). "Cons" is the test of constant conditional average treatment effect, i.e., the test that the average effect conditional on the covariates is identical for all subpopulations implying no heterogeneity in the treatment effects. "ATE" is Crump et al.'s non-parametric test of no average treatment effect. Variables (characteristics) are selected similarly to the top-down selection strategy of Crump et al. (2008), using the BIC to select the preferred number of variables using only the data for the control group (G1). The CM distance estimates are winsorized at the 10% level.

Table B7: Effects of the communication tools on knowledge of data revisions: P-values of non-parametric tests for zero conditional, constant and zero average treatment effects by Group (G#)

		Selected variables	G2: textual 'about'			G3: likely revised			G4: interval			G5: density strip			G6: bell curve			G7: Covid effects			G8: time interval		
			CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE	CATE	Cons	ATE
Revisions Awareness (q18: 1 to 4)	2018	know GDP young full-time male trust ONS	0.04	0.03	0.81	0.03	0.02	0.73	0.04	0.03	0.27	0.10	0.08	0.25	0.23	0.16	0.77	-	-	-	-	-	-
	2020	know GDP young, grad full-time grad econ grad full-time grad econ	0.13	0.25	0.04	0.60	0.47	0.84	0.06	0.05	0.24	0.70	0.61	0.52	0.04	0.02	0.82	0.92	0.89	0.48	0.86	0.76	0.96
Revisions due to more information (q17)	2018	know GDP freq-news trust ONS, know ONS grad econ young	0.23	0.19	0.73	0.34	0.40	0.20	0.44	0.44	0.36	0.08	0.15	0.06	0.16	0.38	0.10	-	-	-	-	-	-
	2020	know GDP young trust ONS know ONS	0.45	0.31	0.37	0.43	0.55	0.35	0.99	0.99	0.40	0.67	0.72	0.22	0.72	0.60	0.35	0.98	0.96	0.87	0.82	0.93	0.13
Revisions due to vested interests (q17)	2018	young	0.41	0.22	0.59	0.90	0.89	0.71	0.46	0.25	0.62	0.42	0.48	0.26	0.12	0.06	0.43	-	-	-	-	-	-
	2020	grad econ	0.66	0.56	0.44	0.88	0.92	0.66	0.51	0.72	0.27	0.93	0.86	0.74	0.75	0.71	0.56	0.86	0.74	0.61	0.99	0.98	0.97
Revisions due to ONS mistakes (q17)	2018	know GDP	0.16	0.07	0.45	0.30	0.32	0.18	0.53	0.26	0.84	0.43	0.23	0.52	0.81	0.59	0.66	-	-	-	-	-	-
	2020	full-time	0.12	0.09	0.22	0.14	0.05	0.61	0.49	0.39	0.37	0.25	0.17	0.32	0.62	0.35	0.85	0.64	0.35	0.91	0.71	0.41	0.91

Notes: See notes to Table B6.

Table B8: The effects of the communication tools on probabilistic perceptions of data uncertainty: Treatment effects for CM distance at the median, conditional on observed characteristics, by Group (G#)

	G2: textual 'about'	G3: likely revised	G4: interval	G5: density strip	G6: bell curve	G7: Covid effects	G8: time interval
Know GDP N=2308	-0.009 (-0.58)	0.001 (0.04)	**-0.042** (-3.01)	**-0.050** (-3.68)	**-0.034** (-2.43)	-0.007 (-0.44)	-0.006 (-0.37)
Don't know GDP N=1893	0.005 (0.38)	0.025 (1.23)	-0.008 (-0.66)	-0.014 (-0.97)	-0.014 (-1.20)	0.010 (0.71)	0.017 (1.29)
Know ONS N=2427	-0.014 (-0.83)	-0.008 (-0.43)	**-0.041** (-2.94)	**-0.044** (-3.08)	**-0.037** (-2.50)	0.002 (0.15)	-0.001 (-0.05)
Don't know ONS N=1774	0.020 (1.49)	**0.041** (2.72)	0.000 (0.01)	-0.014 (-1.38)	-0.003 (-0.25)	0.007 (0.56)	0.017 (1.25)
Trust ONS N=2937	-0.016 (-1.20)	-0.017 (-1.09)	**-0.036** (-3.28)	**-0.044** (-3.84)	**-0.037** (-3.22)	-0.015 (-1.14)	-0.017 (-1.29)
Don't trust ONS N=1264	**0.041** (2.33)	**0.055** (3.13)	0.011 (0.61)	-0.008 (-0.55)	0.028 (1.53)	**0.039** (1.95)	**0.062** (3.09)
Young=1 N=1209	-0.026 (-1.25)	-0.022 (-0.97)	**-0.047** (-2.29)	**-0.048** (-2.01)	**-0.049** (-2.44)	-0.032 (-1.50)	-0.007 (-0.31)
Young=0 N=2992	0.013 (1.01)	**0.031** (2.24)	-0.015 (-1.53)	**-0.024** (-2.57)	-0.010 (-0.93)	0.021 (1.70)	0.012 (1.07)

Notes: The CM treatment effects for G2-G8 are for the 50% quantile relative to the control group (G1) shown the current headline ONS GDP point estimate press release. Robust t-statistics in parentheses. Treatment effects in bold when statistically significant at 10%. Young is characterized as those respondents under the age of 34.

Appendix References:

Clarke, Damian, Joseph P. Romano, and Michael Wolf (2020). "The Romano–Wolf multiple-hypothesis correction in Stata." *The Stata Journal*, 20(4), pp. 812-843. doi:10.1177/1536867X20976314.

Young, Alwyn (2019). "Channeling Fisher: randomization tests and the statistical insignificance of seemingly significant experimental results." *Quarterly Journal of Economics*, 134, pp. 557–598. doi:10.1093/qje/qjy029.