

Online Appendix

A Newspaper data

Table 2 summarises the data sources and time period for each newspaper used in our dataset, and figure 17 shows the reach of each newspaper. For newspapers from ProQuest we used all articles in the time period. For both Factiva and LexisNexis, we downloaded articles that contained any of the following keywords: a, e, i, o, u, the, be, to, of, and, in, that, have, it, for, not, he, as, you, do, at, this, but, his, by, from, they, we, say, her, she, an, will, or, my, one, all, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. None of the words after “the” returned any additional results, making us fairly confident that these searches return all articles.

Table 2: Newspaper sources and coverage.

Newspaper	Country	Type	Source	Start Year	End Year	Article Count	Articles Per Year
The Sun	UK	Tabloid	LexisNexis	1996	2022	4,999,583	185,170
Metro	UK	Tabloid	LexisNexis	2008	2022	882,084	58,806
Daily Mail	UK	Tabloid	ProQuest	1998	2022	3,184,404	127,376
Mirror	UK	Tabloid	ProQuest	1996	2022	3,591,628	133,023
Guardian	UK	Broadsheet	ProQuest	1995	2022	2,699,992	96,428
Telegraph	UK	Broadsheet	ProQuest	1995	2022	2,450,994	87,536
Le Figaro	France	Broadsheet	LexisNexis	1996	2022	1,362,361	50,458
Libération	France	Broadsheet	LexisNexis	2018	2022	64,031	12,806
Le Monde	France	Broadsheet	ProQuest	1995	2022	1,276,459	45,588
Les Echos	France	Broadsheet	ProQuest	2010	2022	410,187	31,553
New York Post	US	Tabloid	Factiva	1997	2021	785,106	31,404
Washington Post	US	Broadsheet	ProQuest	1995	2022	2,705,073	96,610
Wall Street Journal	US	Broadsheet	ProQuest	1995	2022	1,884,655	67,309
New York Times	US	Broadsheet	ProQuest	1995	2022	5,070,395	181,086
USA Today	US	Broadsheet	ProQuest	1995	2022	575,593	20,557
Bild	Germany	Tabloid	Factiva	2011	2022	182,822	15,235
Total						32,125,367	1,240,944

For some newspapers sourced from ProQuest, the earliest years of coverage contain abstracts or truncated text rather than full articles. For *USA Today* (1995–January 1997), all articles are replaced by short abstracts. For the *Washington Post* (1995–1996) and *New York Times* (1995–early 1996), approximately 40% of articles are abstracts, with the remainder being full text. For the *Wall Street Journal* (1995–September 1997), approximately 30% of articles are abstracts. For the *Telegraph* (1995–October 2000), articles are not replaced by abstracts but are truncated, with a maximum article length of around 2,000 characters compared to 6,000–10,000 characters in later years. In each case, there is a sharp transition to full-text coverage, consistent with a change in ProQuest’s licensing rather than editorial decisions. These data limitations affect only the descriptive analysis, as our causal analysis uses *The Sun* (sourced from LexisNexis) from 2013 onwards.

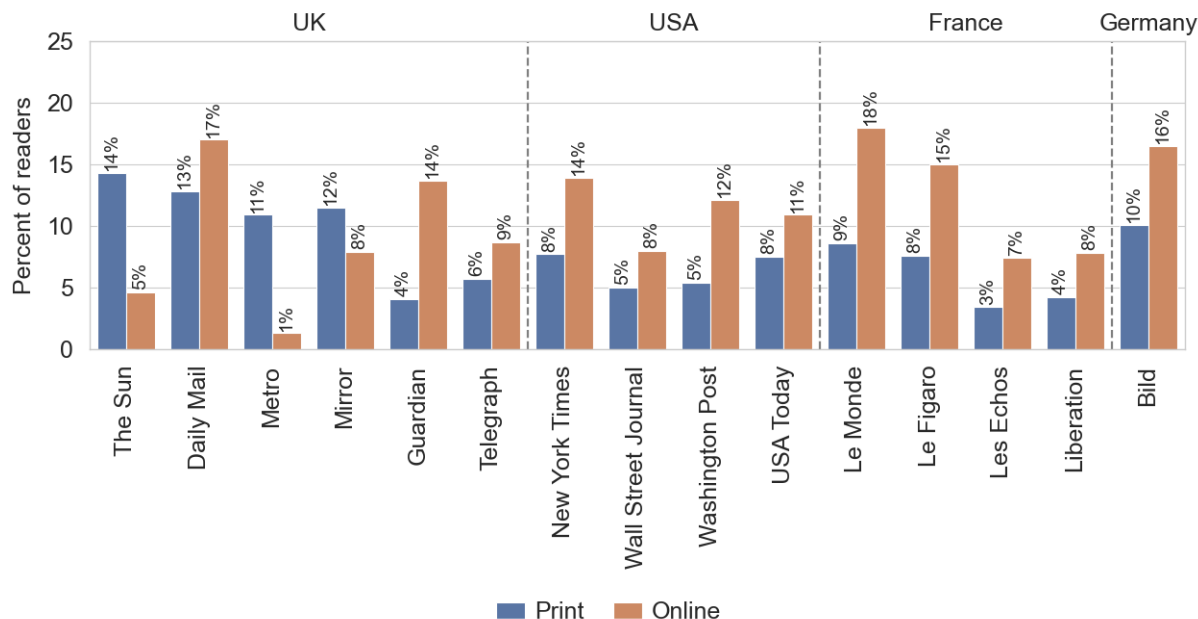


Figure 17: Newspaper reach for newspapers in our dataset.

Notes: Reach is the percentage of survey respondents in a country who report having used a given title to access news in the past week. Respondents may name more than one title, so the bars do not sum to 100%. No data is available for the *New York Post*. Sourced from the Reuters Institute, *Digital News Report 2016* (question B3), conducted by YouGov. See <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2016>.

For *The Sun*, our main source of articles was LexisNexis. However, there are 40 missing days in 1999, three missing days in 2008 and one missing day in 2019, as well as two days in 2019 which have very few articles. For these days, we manually downloaded articles from Factiva and added these to our corpus.

While LexisNexis contains full-text data (and an API to access it), data on which edition (Scotland, Northern Ireland, National Edition, etc.) published each article was not accurate. From correspondence with LexisNexis: “Our coverage of *The Sun* (England) includes these editions: London (Main Edition) Scotland (Scottish Edition) Eire (Ireland edition) Ulster (Northern Ireland Edition). The publisher may choose to only provide us with one version of an article that might appear in multiple editions. We make the content available on our site as it is provided to us by the publisher.”

Given that in the causal analysis we focus only on England, filtering to only articles listed as “London (Main Edition)” in LexisNexis would cause us to miss some articles that were listed under another edition, but were in fact printed in both that edition and in the London

edition, while including articles from all editions would lead us to include some articles that English readers would not have seen. To solve this, we took all articles that were categorised as mainly about benefits, but were listed in LexisNexis with an edition other than London, manually found these articles in Factiva, and used Factiva’s edition information to determine which editions each article appeared in. We use these articles only for the causal analysis. In the descriptive analysis, we use all articles from LexisNexis, regardless of edition.

B Natural language processing of newspaper articles

Each article in our corpus is processed through a natural language processing pipeline, which is summarised in figure 1. Details on each of the steps in this pipeline are provided here, and table 3 summarises the performance of each classifier on hand-labelled test data. The full text of every prompt is given in table S2 of the online supplement.

B.1 ‘Mentions benefits’ classifier

First we pass articles through a custom-trained language model to classify whether a newspaper article contains any mention of welfare benefits or the welfare system.

Splitting articles We are interested in detecting any mention of welfare benefits or the benefits system, but most language models have a ‘context length’, that is, an upper limit on the number of tokens (word pieces) after which a text is truncated. This poses a problem if we feed full articles into the model, as several articles do not mention welfare benefits until the part that is truncated. Therefore for this stage, we chunk articles, so that the text used remains within the context window. Where possible, we split articles at paragraph breaks, with an overlap of one paragraph between chunks. We include the headline with every chunk, featurising the text as “headline [sep token] chunk”. At inference time, we count a full article as mentioning benefits if any of its component chunks were predicted as mentioning benefits.

Training data As articles about welfare benefits make up less than 2% of all newspaper articles during our time period, it would have been infeasible to create training data for our topic classifier by labelling randomly sampled data. Instead, we select data to label using class-balanced discriminative active learning (CB-DAL v1) from Silcock and Löwe (2026).

For the UK, we started with a random sample of 1,000 articles per newspaper, which creates a pool from which we selected articles to label. To create the sample to initiate the

Table 3: Performance of the natural language processing classifiers.

Classifier	Model	Split size			Hyperparameters			Accuracy	F1
		Train	Dev	Test	Epochs	Learning rate	Batch		
<i>Mentions benefits</i>									
UK	RoBERTa-large	2,629	568	564	10	5×10^{-6}	16	98.9	96.2
US	RoBERTa-large	593	115	296	10	1×10^{-5}	16	98.3	95.9
France	CamemBERT(a)-v2	703	143	154	10	5×10^{-5}	16	97.1	93.2
Germany	EuroBERT-210m	589	56	55	10	1×10^{-5}	32	98.2	90.9
<i>Mainly about benefits</i>									
UK	RoBERTa-large	700	150	150	10	2×10^{-6}	16	92.9	90.5
US	RoBERTa-large	350	75	75	10	1×10^{-5}	16	96.0	82.3
France	CamemBERT(a)-v2	350	75	75	10	5×10^{-5}	16	94.7	83.3
Germany	EuroBERT-210m	350	75	75	10	1×10^{-5}	16	94.7	71.4
<i>Frame</i>									
Individual	GPT-4o-mini, o4-mini	–	100	100	–	–	–	98.0	94.1
Policy	GPT-4o-mini	–	100	100	–	–	–	91.0	93.6
Statistics	GPT-4o-mini	–	100	100	–	–	–	96.0	83.3
<i>Slant</i>									
Negative vs. positive or neutral	o4-mini	–	100	100	–	–	–	94.0	94.3
<i>Person characteristics</i>									
Main individual	GPT-4o-mini	–	100	100	–	–	–	96.0	–
Gender	GPT-4o-mini	–	100	100	–	–	–	96.0	–
Employment status	o4-mini	–	100	100	–	–	–	87.0	–
Marital status	GPT-4o-mini	–	100	100	–	–	–	90.0	–
Has children	o4-mini	–	100	100	–	–	–	96.0	–
Number of children	GPT-4o-mini	–	100	100	–	–	–	84.3	–
Native to the country	o4-mini	–	100	100	–	–	–	97.0	–
Benefit fraud	GPT-4o-mini	–	100	100	–	–	–	98.0	–
Years claiming	o4-mini	–	100	100	–	–	–	95.0	–
Won the lottery	GPT-4o-mini	–	100	100	–	–	–	100.0	–
Race (UK)	o4-mini	–	100	100	–	–	–	93.0	78.6
<i>Sun sample filters</i>									
Reader communication	GPT-4o-mini	–	200	200	–	–	–	100.0	100.0
Immigrants, jihadists or terrorists	GPT-4o-mini	–	200	200	–	–	–	96.5	78.9
Not related to the UK	GPT-4o-mini	–	200	200	–	–	–	99.5	98.2
TV program	GPT-4o-mini	–	200	200	–	–	–	98.0	77.8
<i>Tropes</i>									
Trope (majority of three runs)	GPT-5.4	–	–	100	–	–	–	89.0	–

Notes: Accuracy and F1 are in percent, on hand-labelled test sets not used for training or prompt development. For the finetuned classifiers, labelled articles are split 70%, 15% and 15% into training, development and test sets; ‘mentions benefits’ sizes count chunks, with all chunks of an article in the same split. For the GPT classifiers, the development set is used to tune the prompt; the race classifier is evaluated on names rather than articles, and the number of children only on test articles where the individual has children. The finetuned models are RoBERTa-large (Liu et al., 2019), CamemBERT(a)-v2 (Antoun et al., 2024) and EuroBERT-210m (Boizard et al., 2025). – indicates not applicable or not reported.

Active Learning process, we used keyword searches based on terms suggested by Baumberg (2016) and Morrison (2019). We use all keyword hits of ‘social security’, ‘dole’, ‘benefit fraud’, ‘scrounger’, ‘shirker’, ‘sponger’, ‘skiver’, ‘workshy’, ‘work-shy’, ‘something for nothing’, ‘underclass’, ‘benefit tourism’ or ‘benefit tourist’, as there are few hits, but they are highly likely to be on topic. We then randomly sample across keyword hits of ‘benefits’ and ‘welfare’, which return far more results but are less likely to be on topic, to get up to 10 articles per newspaper.¹ This gives an initial sample of 298 articles, which were split into 761 chunks. These were manually labelled.

We then used this first sample to initialise the Active Learning method. We use the remaining articles within the random sample of 1,000 articles per newspaper as the pool. We run this for 30 iterations, with a RoBERTa-base classifier (Liu et al., 2019), a batch size of 32, a learning rate of 5×10^{-5} , and 10 training epochs at each iteration. As run time is quadratic in context length, for the first 22 iterations we used a context window of 100 tokens, and we used a context window of 512 tokens for the final eight iterations. Each iteration pulled out a further 100 chunks, which were manually labelled. This allows us to expand our training data beyond what is returned by keyword searches, as many on-topic articles do not use these keywords. Silcock and Löwe (2026) find that when the initial sample covers only part of a rare target class, this method finds at least as large a share of the missed parts as of the covered ones. This results in a labelled sample of 3,761 article chunks, of which 509 are positive.

For France, Germany and the US, we use the same process as for the UK, but use the UK labelled data as the initial seed. For France and Germany, we translate the UK labelled data into French and German using GPT-4o-mini, with the prompt given in table S2 of the online supplement. For the US and France, we ran seven Active Learning iterations. For Germany, we ran four, as its pool of articles is smaller. For all three countries, we then labelled a further 300 chunks which were predicted as on topic by an intermediate model, for a total of 1,000 labelled chunks for the US and France, and 700 for Germany.

Labelling rules Labelling was carried out manually by one of the authors and an undergraduate research assistant. Research assistants (and the coauthor) were all native speakers and familiar with the context of the country in question. A sample was labelled by both to check coherence on the task and congruence between labellers. Labels agreed in 97% of

¹These keywords are only used to initiate the Active Learning process, which can select any article from the pool for further labelling. Therefore we are in no way restricting the scope of the classifier to only these keywords.

cases, and the remaining cases were agreed to be labelling mistakes rather than conceptual differences. Any mention of working-age welfare benefits or the welfare system was considered on topic, where a working-age welfare benefit is any money or vouchers paid regularly by the state to an individual or family. The full labelling rules, including what does and does not count as a welfare benefit, are given in section S3 of the online supplement.

Model training We train a different model for each country, allowing for language differences, as well as specificities of the welfare system in each country. Table 3 reports the model, data splits, hyperparameters and test-set performance for each country. The Active Learning process means we have pulled out examples that are especially difficult to classify, so these are lower bounds on classification performance in the full sample.

B.2 ‘Mainly about benefits’ classifier

Within articles that mention welfare benefits, we wish to separate those that have a tangential mention from those where this is a major topic of the article.

Training data and labelling rules For France, Germany and the US, we took 500 random articles from the pool of articles that were predicted as mentioning benefits. For the UK, we took 1,000 articles, upweighting the number of articles from *The Sun* newspaper. As this was a difficult task, each article was labelled independently by two labellers. Differences between labellers were resolved by a coauthor, in order to maintain coherence across the four countries. “Mainly about” is inherently subjective, so to ensure consistency across labellers, we provided them with a set of rules of thumb, given in section S3 of the online supplement. As for the ‘mentions benefits’ classifier, we train a different model for each country, and table 3 reports the models, data splits, hyperparameters and test-set performance.

B.3 GPT classifiers

For extracting the most fine-grained information from the articles, we use a series of GPT classifiers. Table 3 reports the model used for each, the sizes of the development and test sets, and test-set performance, and table S2 of the online supplement gives the full text of every prompt. In all cases, the labelled articles are randomly sampled.

Individual, policy, statistics We classify each article as being focused on a specific individual or family, a policy, or statistical information. An article can be classified as

multiple, or none of these. To identify articles about specific individuals or families, we use a prompt asking whether the article is mainly about one specific individual or family. We follow this with a second prompt, which asks whether the main individual or family is a politician or famous. An article is classified as being about an individual if the first prompt returns ‘yes’ and the second prompt returns ‘no’. To identify articles about policies, we use a prompt asking whether the article is mainly about a public policy. To identify articles about statistics, we use two prompts, which ask whether the article is mainly about data, statistics or a market brief, and whether it is mainly about a research report. An article is classified as being about statistics if either prompt returns ‘yes’.

Slant classification To classify the slant of an article towards welfare benefits, we use a prompt which asks whether the article gives a ‘good’, ‘bad’ or ‘neutral’ impression of people seeking or claiming welfare benefits. We evaluate the performance of this prompt using ‘good’ and ‘neutral’ as a combined category, as these categories are always combined in the causal analysis.

Person characteristics For articles about private individuals, we further extract additional characteristics about those individuals. First we identify the main person. We then extract their gender, employment status, marital status, whether they have children and, if so, how many, whether they are native to the country, whether the article concerns benefit fraud, how long they have been claiming benefits, and whether they have won the lottery.

Race classification To classify the perceived race of individuals mentioned in articles, we use a name-based approach. Unlike the other person attribute classifiers, which receive the full article text, this classifier receives only the extracted name of the main individual. Including the full article usually led to a refusal to make a guess for the race. To validate this approach, as race is rarely directly mentioned in the articles, we took a random sample of articles for each country and searched to find the original article. In the cases where the original article could be found and it included a photo of the individual in question, we noted their race. We used this method to build a set of 200 hand-labelled names for the UK and 162 for the US. We used the first half to develop the prompts and the second half for evaluation, and table 3 reports accuracy and F1 on this held-out set. There, the classifier outperforms name-based ethnicity classification methods such as the `ethnicolr` Wikipedia model for the UK. The classifier estimated the share of non-white individuals in the held-out set at 13.1%, compared to the true share of 15.0%.

For the US, we were unable to find a name-based method that reliably estimated the aggregate share of non-white individuals. We evaluated several approaches: GPT-based classification (using a prompt adapted for US Census race categories), the `ethnicolr` Florida voter registration model, US Census surname distributions (via `surgeo`), and RaceBERT, as well as ensemble methods combining GPT with probabilistic corrections from `ethnicolr`. We therefore do not report race breakdowns for US newspaper coverage.

***Sun* sample filters** For the causal analysis, we drop *Sun* reader communications and articles about immigrants, jihadists or terrorists, other countries or TV programmes, each flagged by a yes/no prompt (table S2 of the online supplement). The immigrant prompt flags an article only if it clearly states that the main claimant immigrated, without inferring this from names or stereotypes. The UK prompt counts Northern Ireland as part of the UK and returns ‘yes’ when unclear, so only articles clearly about another country are removed.

B.4 Negative coverage of benefit recipients

Figure 18 reproduces four articles from the final sample of negative *Sun* articles about individual benefit recipients, the treatment population used in the main event studies.

B.5 Trope classification

We derive the taxonomy of fourteen welfare tropes by applying the agentic clustering procedure of Löwe and Silcock (2026) to summaries of negative articles about individual benefit recipients. We generate these summaries from the headline and body text of the article, with o4-mini, using a prompt that asks for a summary in English in 25 words or fewer, without using any names of people, places or organisations. Requiring the summary in English allows articles from all four countries to be clustered and classified together. Suppressing names of people, places and organisations means that tropes are discovered from the narrative content of the articles rather than from recurring individuals, locations or institutions.

We then run the agentic clustering procedure with a target of between five and twenty clusters and the instruction that clusters should be distinct tropes about benefit recipients. Three proposer agents each independently sample 150 summaries and propose a cluster set; they converge on twelve to thirteen clusters with closely matching content, which a synthesiser agent merges into fourteen. An auditor agent assigns 200 previously unseen summaries to this set, achieving 94% coverage at a mean self-reported confidence of 4.4 out

Disability swindler is busted at fit club (2014-01-11)

A BENEFIT cheat who swindled £21,000 claiming she could barely walk was caught taking part in FITNESS classes. Anne Bird, who has osteoarthritis, was filmed by undercover investigators after a tip-off. The 60-year-old, who said she needed help dressing, did pilates, aerobics and spinning bike classes, a court heard. She was also filmed doing sit-ups at a gym. She told the Department for Work and Pensions she needed the higher rate of disability allowance because she also had problems with her knees and hips. Despite Bird, from Pleck, West Mids, being entitled to some benefits, between October 2008 and August 2012 she was overpaid £21,109. She admitted two counts of making a false declaration on forms at Walsall magistrates. Bird got 12 weeks jail suspended for 12 months.

AWFUL DOLE MODELS (2013-10-04)

I HOPE you all read about the feckless, idle, disgraceful Flisher family boasting of their £30,000 benefits and new four-bedroom council house, and how working is “a mug’s game”. If you did, you’d know why the Government is so right to crack down on abusers of welfare benefits. But it’s Maggie and Gavin Flisher’s six children I most worry about — research proves workshy parents damage their offspring for life, dooming them to an unambitious, pointless existence like their own. Sad.

Benefits boob op (2018-12-03)

A MUM of eight who claims £30,000 a year in benefits is splashing out £2,500 on a boob job. Marie Buchan, dubbed Octomum, will fly to Istanbul for the operation to boost her breasts from 34B to 36DD. The ex-lapdancer, 36, said she saved up for the surgery through car boot sales and part-time work. But Marie, of Selly Oak, Birmingham, works less than 16 hours a week and rakes in more than the UK average salary in benefits. She said: “I’m doing this for me.”

£100K CON ‘AIDS BENEFITS STIGMA’ (2021-11-25)

A BENEFITS cheat who enjoyed five-star foreign holidays after swindling almost £100,000 was told by a judge she helped stigmatise honest claimants. Claire Finney, 41, dishonestly claimed extra child tax credits, universal credit, housing benefit and income support for five years. The mum of four used the £97,028 she was not entitled to on regular, expensive trips with her family to Cyprus. Passing sentence, Judge Nicholas Dean QC told her: “You will know there is sometimes a stigma associated towards people who claim benefits — that is, in large measure, because of people like you. You give those people who legitimately claim benefits a bad name as they are tarred with the same brush.” Finney, of Eccles, near Salford, fiddled the money by saying she was on her own, while living with partner Joseph Perry. She admitted the con but was spared jail at Manchester crown court as her children were deemed vulnerable. Instead, she got a suspended 12-month sentence, with 180 hours’ unpaid work. Finney also has to pay back the money, but at just £11.77 a month it will take her 686 years to do so.

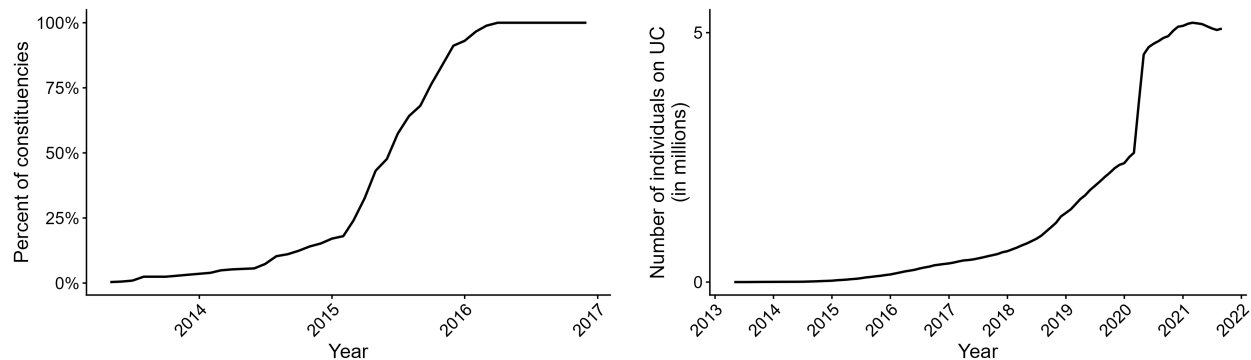
Figure 18: Examples of articles that we classified as containing negative content about benefit recipients.

of 5. An investigator agent and an adversarial critic agent then review the weakest clusters and the taxonomy as a whole, and their recommendations are applied as revisions to nine of the fourteen definitions, chiefly the addition of boundary notes separating confusable tropes. We subsequently hand-edit nine definitions, tightening the wording where it was unclear. Table S1 of the online supplement reports the full definition of each trope, including the boundary notes supplied to the classifier.

After developing the taxonomy, we classify each article summary into one of fourteen trope clusters (or “none of these”) using GPT-5.4 with medium reasoning effort. The input to the classifier is the 25-word article summary (not the full article text). The system prompt contains the cluster definitions and boundary notes (without examples), preceded by a short header, given in table S2 of the online supplement. To reduce non-determinism, we run the classifier three times per article and take the majority vote; table 3 reports the accuracy of this majority vote on the test set. Of the 11 errors on it, 7 involve articles that plausibly fit two tropes (genuine dual-trope cases where the human and model selected different but defensible clusters).

C Universal Credit

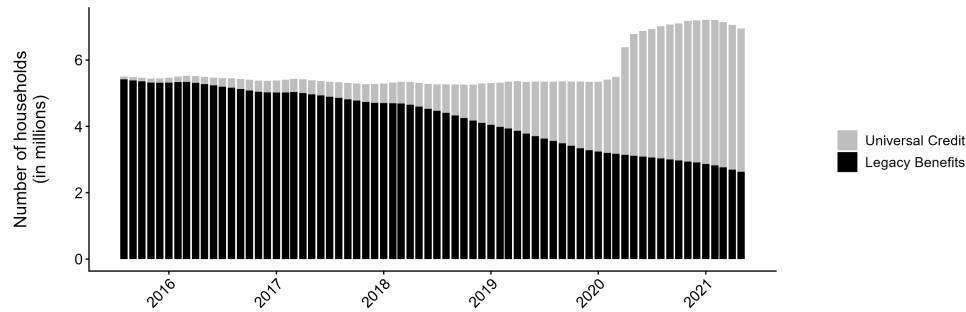
Figures 19, 20 and 21 present extra descriptive data on Universal Credit and the legacy benefits it replaced.



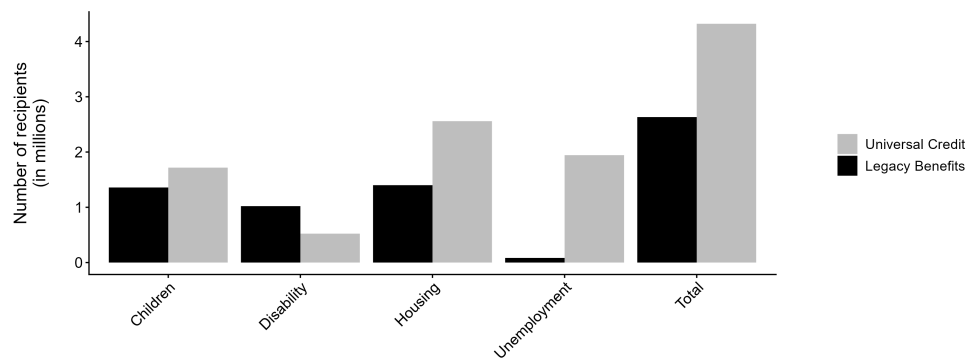
(a) Share of parliamentary constituencies that have rolled out Universal Credit live service. (b) Number of people receiving Universal Credit.

Figure 19: Time series of Universal Credit across constituencies and individuals.

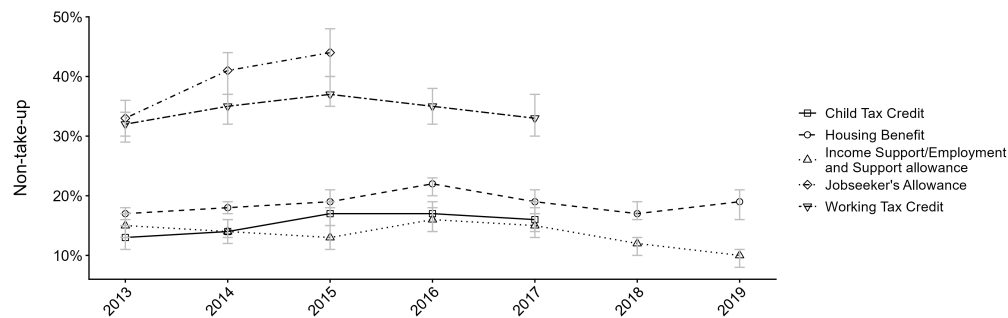
Notes: Sourced from the House of Commons Library. See <https://commonslibrary.parliament.uk/constituency-data-universal-credit-roll-out/>.



(a) Number of households receiving Universal Credit or legacy benefits.



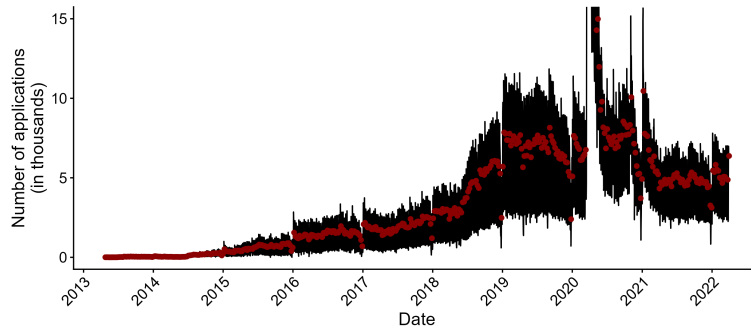
(b) Recipients of Universal Credit or legacy benefits, by type of support received.



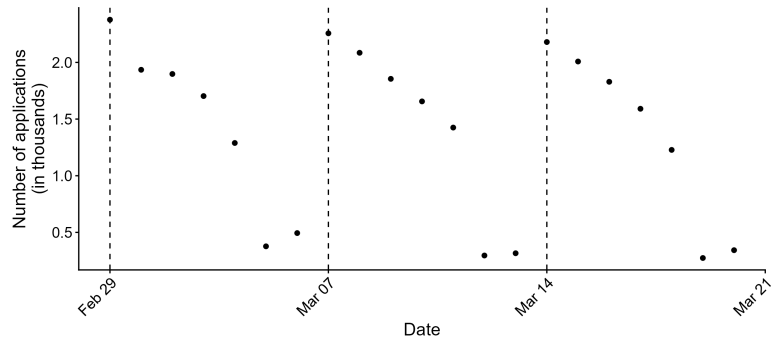
(c) Non-take-up of legacy benefits.

Figure 20: Universal Credit and the legacy benefits it replaced.

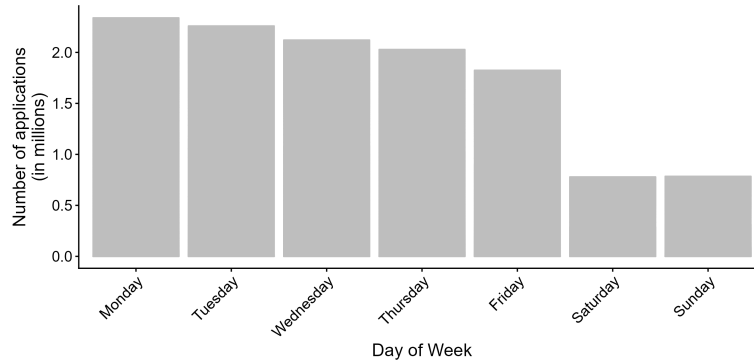
Notes: Panels (a) and (b) are from the House of Commons Library (see figure 19); in panel (b), unemployment benefits are counted in individuals and all others in households, in May 2021. Panel (c) is from the DWP, *Income-related benefits: estimates of take-up*, on a caseload basis (recipients relative to all those entitled), with error bars for the published lower and upper bounds. See <https://www.gov.uk/government/collections/income-related-benefits-estimates-of-take-up--2>.



(a) Daily Universal Credit applications, 2013–2022. Red dots show weekly averages.



(b) Daily Universal Credit applications, 29 February to 20 March 2016. The dashed lines are Mondays.



(c) Total number of Universal Credit applications by day of week, 2013–2022.

Figure 21: Daily Universal Credit applications.

Notes: Panel (a) shows the overall trend in the number of applications. Its vertical axis is truncated at 15,000 applications; the peak of 119,268 applications on 27 March 2020, at the start of the COVID-19 pandemic, lies outside the plotted range. Panel (b) shows a randomly chosen three-week interval. Panel (c) shows the day-of-week pattern of Universal Credit applications.

D Other data

In our descriptive analysis we use data on the demographics of Universal Credit recipients, from the DWP,² as well as data on the demographics of SNAP recipients, from the SNAP Quality Control Public Use Files for fiscal years 1996 to 2022, published by the Food and Nutrition Service of the US Department of Agriculture.³ The one exception is the length of time US households have been claiming, which the Quality Control files do not record; we take this from the 2022 Survey of Income and Program Participation, accessed through the US Census Bureau’s API.⁴ For table 4, we use monthly unemployment data from FRED.⁵

For the heterogeneity analysis, we use data on income by region. We obtain data on income at the parliamentary-constituency level for tax year 2018/19 from HM Revenue & Customs.⁶ Constituencies and postcode areas (the regional decomposition in the Universal Credit applications data) are not uniquely matched.⁷ To map the constituency-level income data to postcode areas we map postcodes to constituencies using the November 2018 Office for National Statistics (ONS) Postcode Directory. Then, for each postcode area, we take a weighted mean of constituency-level mean employment income across all overlapping constituencies, weighting by the number of taxpayers in each constituency.

We also use data on *Sun* readership from the 2015 British Election Study (BES) face-to-face post-election survey,⁸ the most recent wave to record respondents’ Lower Layer Super Output Area (LSOA). A respondent counts as a *Sun* reader if they name *The Sun* as the paper in which they most often read about politics or current affairs; respondents who read no paper count as non-readers. We map LSOAs to postcode areas using the November 2018 ONS Postcode Directory, splitting an LSOA across postcode areas in proportion to

²See <https://stat-xplore.dwp.gov.uk/>.

³See <https://snapqcddata.net/datafiles>, accessed 1 February 2025.

⁴See <https://api.census.gov/data/2022/sipp>.

⁵OECD harmonised unemployment rates, retrieved from FRED, Federal Reserve Bank of St. Louis: series LRUNTTTUSM156S (United States), LRHUTTTTGBM156S (United Kingdom), LRHUTTTTFRM156S (France) and LRHUTTTTDEM156S (Germany). See <https://fred.stlouisfed.org/>.

⁶*Income and Tax by Parliamentary Constituency*, 2018 to 2019 release, Table 3.15 (NS_Table_3_15_1819.ods). See <https://www.gov.uk/government/statistics/income-and-tax-by-parliamentary-constituency-2010-to-2011>; the table itself is at https://assets.publishing.service.gov.uk/media/6082ad9be90e076aaaa26e3d/NS_Table_3_15_1819.ods. Constituencies are the geographical entities that elect a single Member of Parliament. There are 650 constituencies in total, with 533 constituencies in England, 40 in Wales, 59 in Scotland and 18 in Northern Ireland (as of 2019). The number of eligible voters varies roughly between 50,000 and 100,000 per constituency.

⁷37% of English constituencies fall within a single postcode area, 32% span two, 20% span three and 11% span four or more.

⁸See <https://www.britishelectionstudy.com/data-objects/cross-sectional-data/>.

its postcodes where it straddles a boundary, although 95% of LSOAs fall wholly within a single postcode area. Weighting by the BES combined survey weight then gives a *Sun* readership share for each postcode area. The underlying sample is thin. The BES covers 2,721 respondents in England and Wales, drawn from 542 LSOAs; 100 of these respondents read *The Sun*. That leaves a median of 22 respondents and one *Sun* reader per postcode area, and 42 of the postcode areas the survey covers contain no *Sun* reader at all.⁹ We therefore use the readership share only to place each postcode area above or below the median, a coarse binary measure at the postcode-area level, which leaves 94 of the 99 English postcode areas.

We also use data on daily counts of disability-related hate crimes, covering every day from 1 January 2013 to 31 December 2022. These are not published as an official series, so we obtain them by Freedom of Information requests to each of the 39 territorial police forces of England and to the British Transport Police, of which 35 returned daily data.¹⁰ Nottinghamshire Police reported the dates on which a disability-related hate crime was recorded but not how many, so we exclude it. Our series is the daily sum across the remaining 34 forces, averaging 13.5 recorded crimes per day. Forces begin reporting at different points, so the level of the series rises over the period as more of them enter, and 2.1% of days record no crime at all, almost all of these in the early years when few forces are reporting.

As a proxy for job-search effort we use daily Google Trends search volumes in England.¹¹ We use the “Job” topic, a Google-curated entity aggregating the queries Google associates with jobs and job-seeking, rather than a fixed keyword list, so that the measure is not sensitive to changes in the words people search for. Google Trends returns a daily series only for short request periods, and scales each series so that its within-period maximum is 100, so we download the data in three-month windows stepped forward two months at a time and chain them by scaling each window so that its mean over the one-month overlap equals the mean of the already-chained series over the same days. Because the index is computed from a sample of searches, repeated requests for the same window differ slightly, so we download each of the 65 windows fifteen times and average across downloads before chaining. The resulting series is a relative index with no natural units, and covers every day

⁹Five postcode areas in the Universal Credit data (EC, HD, HR, SN and WC) contain no BES respondent and are dropped from the readership splits.

¹⁰The force-by-force returns, with each force’s coverage period, its caveats and links to its requests and responses, are deposited at Harvard Dataverse as “Daily Disability-Related Hate Crimes Recorded by Police Forces in England and Wales, 2013–2023 (Freedom of Information Returns)”, <https://doi.org/10.7910/DVN/KVHQ43>.

¹¹See <https://trends.google.com/trends/>.

from 1 January 2013 to 31 December 2022. It is reported as zero on 55 of those days, all falling between October 2021 and August 2022.

E Additional descriptive figures and tables

This appendix collects the supporting figures and tables referenced in the stylized facts of section 2. Figure 22 repeats the tabloid–broadsheet comparison of figure 4 in the main text with US newspapers excluded. Figure 23 gives the newspaper-by-newspaper breakdown behind the tabloid/broadsheet aggregation of Fact 2 (panel (a) of figure 5), separating articles about policy, statistics and individuals. Figure 24 gives the newspaper-by-newspaper breakdown behind Fact 3 (panel (b) of figure 5), showing positive or neutral and negative coverage of individuals. Table 4 reports the regressions behind the unemployment–coverage relationship discussed under Fact 3 in section 2; a rise in the unemployment rate is associated with more negative coverage of individuals in tabloids, but not in broadsheets.

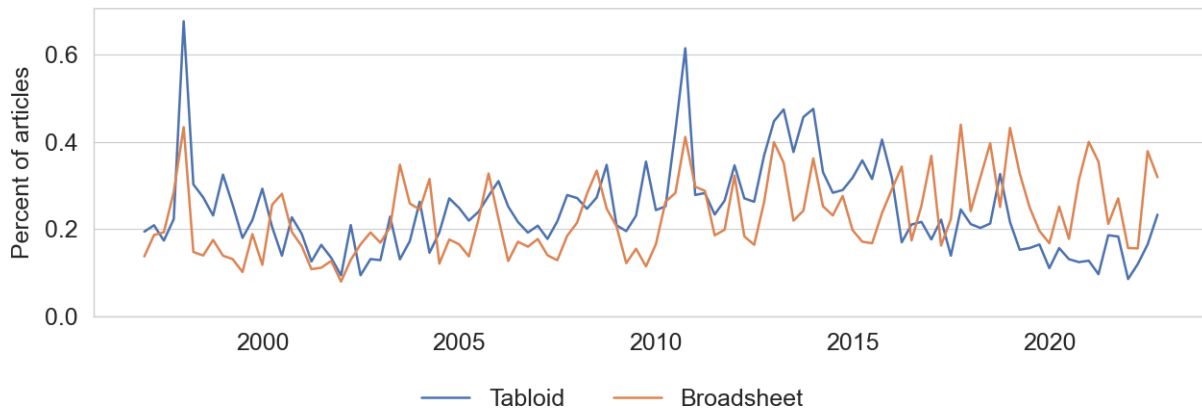


Figure 22: Share of articles mainly about welfare benefits by newspaper type, excluding the United States.

E.1 Variable definitions for figure 7

This subsection gives the coding rules and sources behind the comparison of article and claimant characteristics in figure 7. Claims statistics are computed from administrative data up to 2022, except that UK ethnicity statistics are published only from January 2023, so the non-white row uses the earliest available month. Three rows are shown for the UK only: US SNAP fraud statistics are not published, we have no reliable race measure for US articles,

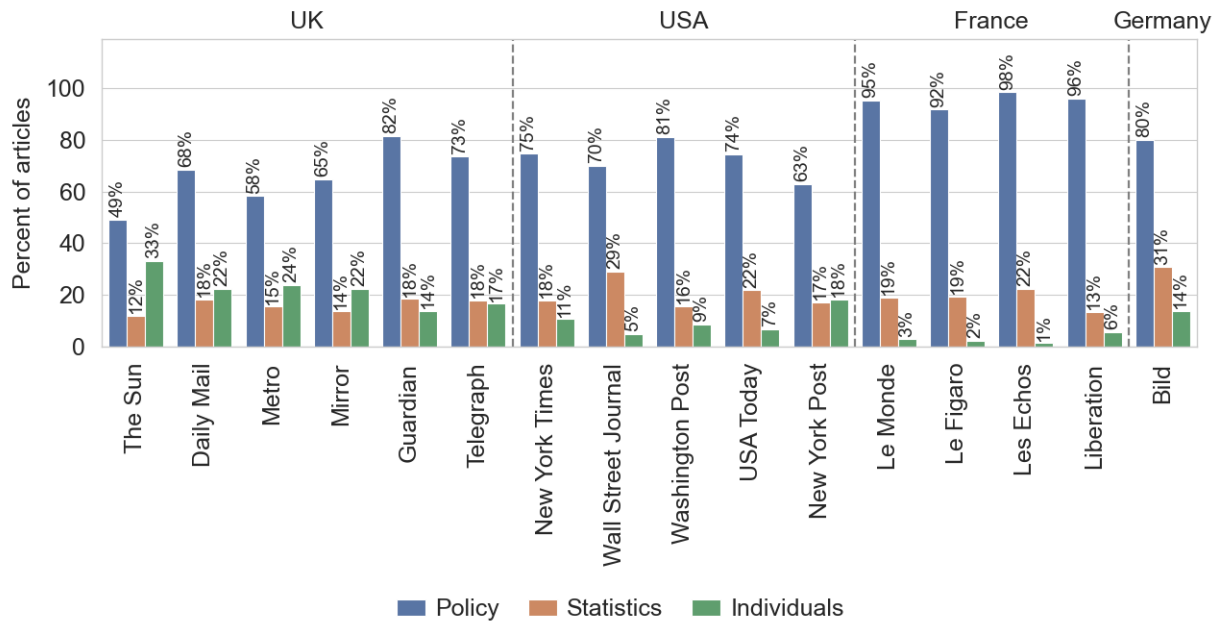


Figure 23: Share of articles about policy, statistics and individuals, by newspaper.

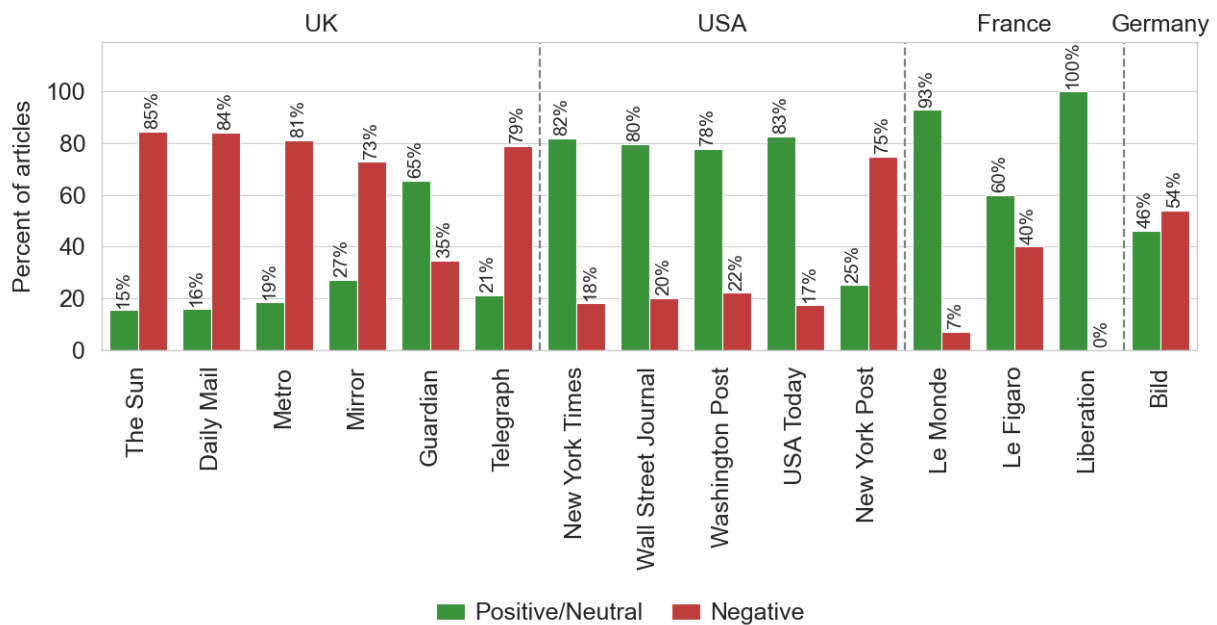


Figure 24: Share of articles about individuals by slant (positive/neutral, negative), by newspaper.

Notes: *Les Echos* is omitted because every one of its articles featuring a private individual is also classified as a policy or statistics article and is therefore excluded from the Fact 3 sample.

Table 4: Relationship between unemployment rate and media coverage of welfare benefits.

Panel A: All newspapers

	Mentions benefits (% total) (1)	Mainly about benefits (% total) (2)	Individuals (% total) (% mainly) (3) (4)		Negative (% total) (% mainly) (5) (6)		(% indiv) (7)
Unemployment rate	0.1016** (0.0413)	0.0370*** (0.0081)	0.0035** (0.0014)	-0.0052 (0.1335)	0.0029** (0.0013)	0.2088 (0.1551)	1.194 (0.7392)
Observations	4,408	4,408	4,408	4,188	4,408	4,188	1,847
R ²	0.41631	0.15569	0.35716	0.34722	0.38259	0.35879	0.34051
Within R ²	0.03069	0.05433	0.04192	4.97×10^{-7}	0.04441	0.00122	0.00299
Baseline mean	1.709	0.2792	0.0189	8.513	0.0127	5.777	58.50
Newspaper fixed effects	✓	✓	✓	✓	✓	✓	✓

Panel B: Tabloid newspapers

	Mentions benefits (% total) (1)	Mainly about benefits (% total) (2)	Individuals (% total) (% mainly) (3) (4)		Negative (% total) (% mainly) (5) (6)		(% indiv) (7)
Unemployment rate	0.1036** (0.0320)	0.0453** (0.0115)	0.0084** (0.0029)	0.0896 (0.4171)	0.0074** (0.0025)	0.5078 (0.3951)	2.544** (0.7045)
Observations	1,542	1,542	1,542	1,471	1,542	1,471	1,088
R ²	0.50675	0.21987	0.24044	0.11258	0.24376	0.11930	0.04278
Within R ²	0.11696	0.16391	0.10687	6.67×10^{-5}	0.11004	0.00284	0.01357
Baseline mean	0.9412	0.2331	0.0389	18.16	0.0303	13.84	75.43
Newspaper fixed effects	✓	✓	✓	✓	✓	✓	✓

Panel C: Broadsheet newspapers

	Mentions benefits (% total) (1)	Mainly about benefits (% total) (2)	Individuals (% total) (% mainly) (3) (4)		Negative (% total) (% mainly) (5) (6)		(% indiv) (7)
Unemployment rate	0.1006 (0.0597)	0.0332** (0.0111)	0.0013 (0.0011)	-0.0483 (0.0697)	0.0009 (0.0009)	0.0728 (0.1087)	-0.1719 (0.8470)
Observations	2,866	2,866	2,866	2,717	2,866	2,717	759
R ²	0.19766	0.12852	0.18136	0.16643	0.20995	0.18626	0.22663
Within R ²	0.02268	0.03573	0.01327	9.64×10^{-5}	0.01483	0.00050	6.21×10^{-5}
Baseline mean	2.122	0.3040	0.0082	3.290	0.0033	1.413	34.23
Newspaper fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: All regressions include newspaper fixed effects. Standard errors clustered by newspaper in parentheses. Columns (1)–(2) show the percentage of total articles that mention or are mainly about welfare benefits. Columns (3)–(4) show the percentage of articles about individuals, as a share of total articles and of articles mainly about welfare benefits respectively. Columns (5)–(6) show the percentage of articles that are about individuals and have a negative slant towards benefit claimants, on the same two bases. Column (7) shows the percentage of articles about individuals that have a negative slant.

and once fraud cases are excluded (see below) too few negative US articles remain for the unemployed row. Note that the article and claims measures of a given characteristic are not always constructed identically, and that the UK and US claims measures differ from one another; the definitions below state each one.

Articles data (both panels): Articles with “unclear” or “unknown” values are excluded from both numerator and denominator. Articles about fraud cases are excluded from the unemployed metric, since in those articles the individual’s employment status is the substance of the allegation rather than an ordinary claimant characteristic. Single parents are defined as individuals who are single or have unclear marital status and have children. Articles state that an individual has children but almost never that they have none, so an unclear value is read as no children and the single parent and 4+ children rows take all individuals as their denominator. Race, used for the UK non-white row, is taken from the GPT article labels described in appendix B.

UK claims data: The women, unemployed, immigrant, non-white and claim duration rows are shares of people on Universal Credit. The single parent, 4+ children and fraudulent claim rows are shares of households, where a household is a single claimant or a couple claiming jointly. Women and unemployed statistics are from DWP Universal Credit statistics by gender and employment status. Ethnicity is from DWP Universal Credit ethnicity statistics. Immigrant statistics are from DWP immigration status statistics, where immigrants are defined as all non-“CTA - UK, Ireland, Right of Abode” categories. Claim duration is the share implied by the spell-length model of appendix H.2, fitted to DWP Universal Credit duration statistics. Children and single parent statistics are from DWP household composition data. Fraud rate is the average from DWP Fraud and Error statistics (FYE 2016–2022). Lottery odds are 1 in 45,057,474 (UK Lotto jackpot).

US claims data: Statistics are computed from SNAP Quality Control (QC) Public Use Files (FY1996–FY2022) using fiscal year weights. Here the claim duration row is a share of people and the remaining rows are shares of households, since women and immigrants are measured on the household head. Women: female head of household. Single parents: households with children and no spouse present. 4+ children: households with four or more children. Immigrants: non-citizen head of household. Claim duration (>2 years) is from the 2022 Survey of Income and Program Participation (SIPP) API, computed as the weighted share of SNAP benefit owners who have received SNAP for more than two years. Lottery odds are 1 in 292,201,338 (US Powerball jackpot).

F Event-study supplementary results

This appendix provides supplementary results to those presented in section 3.

Table 5 accounts stage by stage for the selection of events, from the days with negative coverage of individuals to the estimation sample.

Table 5: From negative coverage of individuals to the estimation sample.

	Days	Change
Days with negative coverage of individuals	603	
Excluding out-groups, reader letters, television and non-UK articles	476	−127
Excluding articles also about policy or statistics	456	−20
Full event window inside the applications sample period	415	−41
Non-overlapping events only (one-sided restriction)	80	−335
Excluding windows containing a public holiday	61	−19
Excluding the 14 March 2020 event	60	−1
Excluding windows containing a zero recorded count	56	−4

Notes: Each row applies one restriction to the row above it. The first row counts the days with negative coverage of individuals from section 1.1.1. Row two drops articles about immigrants or Islamic extremists, reader letters, and articles about television or not about the UK; row three drops articles also about policy or statistics. The fourth row keeps only dates whose full 12-day window falls inside the applications sample period. The fifth applies the one-sided overlap restriction of section 3.1. The sixth removes events whose window contains a public holiday or a day between Christmas Eve and New Year’s Eve, the seventh the 14 March 2020 event, and the last events whose window contains a day with zero recorded applications.

Table 6 reports the coefficients from estimating the event-study specification of equation (1) on four article subsets. Columns (1)–(2) cover articles about individuals and articles about policy or statistics (plotted in figure 9); columns (3)–(4) cover negative and positive-or-neutral articles about individuals (plotted in figure 10). For the day of publication and the day after, the difference between the coefficients in columns (1) and (2) is statistically significant ($p = 0.015$), while the difference between columns (3) and (4) is not.

Columns (5)–(10) of table 6 report the coefficients from estimating the event study on negative articles about individuals, split three ways: into postcode areas with low and high *Sun* readership (columns (5)–(6), plotted in figure 11), by the position of the article in the newspaper (columns (7)–(8), plotted in figure 12), and by whether the article fits one of the fourteen welfare tropes of section 2 (columns (9)–(10), plotted in figure 13). The percentages quoted in the main text for article position refer to publication day ($\tau = 0$). A joint test of equality between the two position groups at $\tau = 0, 1$ does not reject equality ($p = 0.222$).

Table 7 reports the coefficients from estimating the event study on two further outcomes. Columns (1)–(2) use the daily count of disability-related hate crimes, estimated by PPML,

Table 6: Event-study coefficients by article type, *Sun* readership, article position and trope membership.

	Article type				<i>Sun</i> readership		Article position		Trope membership	
	All indiv. (1)	Policy or stats (2)	Negative indiv. (3)	Pos./neutral indiv. (4)	Low (5)	High (6)	Pages 1–5 (7)	Pages 6+ (8)	In trope (9)	Not in trope (10)
$\tau = -6$	-0.0145 (0.0222)	0.0088 (0.0189)	-0.0238 (0.0219)	0.0114 (0.0257)	-0.0164 (0.0214)	-0.0498 (0.0315)	-0.0412 (0.0495)	-0.0194 (0.0219)	-0.0203 (0.0306)	0.1086*** (0.0309)
$\tau = -5$	-0.0145 (0.0280)	0.0184 (0.0180)	-0.0375 (0.0289)	0.0724*** (0.0263)	-0.0300 (0.0306)	-0.0201 (0.0291)	-0.0411 (0.0547)	-0.0352 (0.0278)	-0.0095 (0.0325)	0.0509 (0.0370)
$\tau = -4$	-0.0193 (0.0198)	-0.0030 (0.0174)	-0.0127 (0.0204)	0.0068 (0.0251)	-0.0091 (0.0255)	-0.0121 (0.0242)	0.0252 (0.0469)	-0.0083 (0.0192)	0.0151 (0.0275)	0.0360 (0.0339)
$\tau = -3$	-0.0332 (0.0258)	-0.0255 (0.0177)	-0.0180 (0.0207)	-0.0600 (0.0414)	0.0090 (0.0219)	-0.0190 (0.0207)	-0.0325 (0.0462)	-0.0100 (0.0210)	-0.0087 (0.0280)	0.0245 (0.0367)
$\tau = -2$	-0.0145 (0.0202)	-0.0126 (0.0188)	-0.0164 (0.0212)	-0.0334 (0.0417)	-0.0236 (0.0260)	-0.0123 (0.0243)	-0.0350 (0.0449)	-0.0137 (0.0216)	0.0044 (0.0250)	-0.0450 (0.0414)
$\tau = 0$	-0.0551** (0.0248)	0.0058 (0.0127)	-0.0656*** (0.0236)	-0.0424 (0.0287)	-0.0502** (0.0239)	-0.0839*** (0.0266)	-0.1904** (0.0772)	-0.0604** (0.0234)	-0.0566* (0.0294)	-0.0399 (0.0401)
$\tau = 1$	-0.0769*** (0.0253)	0.0104 (0.0212)	-0.0826*** (0.0242)	-0.0339 (0.0255)	-0.0626*** (0.0213)	-0.0940*** (0.0333)	-0.1903 (0.1179)	-0.0781*** (0.0237)	-0.0559* (0.0322)	-0.0232 (0.0322)
$\tau = 2$	-0.0416 (0.0265)	-0.0157 (0.0210)	-0.0385 (0.0238)	0.0181 (0.0285)	-0.0182 (0.0252)	-0.0467* (0.0272)	-0.0713 (0.0595)	-0.0559* (0.0310)	-0.0068 (0.0306)	-0.0225 (0.0436)
$\tau = 3$	-0.0273 (0.0223)	-0.0408 (0.0286)	-0.0431** (0.0213)	-0.0308 (0.0271)	-0.0341 (0.0254)	-0.0415* (0.0247)	-0.1302* (0.0698)	-0.0438** (0.0204)	-0.0018 (0.0278)	-0.0499 (0.0351)
$\tau = 4$	-0.0189 (0.0209)	-0.0217 (0.0189)	-0.0331 (0.0224)	-0.0220 (0.0243)	-0.0089 (0.0221)	-0.0251 (0.0257)	-0.0328 (0.0494)	-0.0378* (0.0227)	-0.0060 (0.0268)	-0.0284 (0.0371)
$\tau = 5$	-0.0139 (0.0201)	-0.0462** (0.0214)	-0.0340* (0.0205)	-0.0861* (0.0519)	-0.0056 (0.0235)	-0.0522** (0.0242)	-0.0807 (0.0580)	-0.0248 (0.0191)	-0.0060 (0.0269)	-0.0441 (0.0374)
Observations	3,204	3,204	3,204	3,204	3,147	3,175	3,204	3,204	3,204	3,204
R ²	0.94208	0.94837	0.94284	0.94235	0.93498	0.94243	0.94127	0.94286	0.94287	0.94340
Within R ²	0.93688	0.94120	0.93676	0.93788	0.92876	0.93701	0.93785	0.93649	0.93711	0.93441
N. of events	52	65	56	37	55	55	16	57	64	39
Event fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each column reports OLS event-study coefficients from equation (1), where the treatment vector is constructed from a different subset of *Sun* articles. Column (1) uses all articles about individuals; column (2) uses articles about policy or statistics; column (3) uses negative articles about individuals; column (4) uses positive or neutral articles about individuals. Columns (1) and (2) are mutually exclusive: articles classified both as about an individual and as about policy or statistics are excluded from both subsets. Columns (5)–(10) use negative articles about individuals. Column (5) restricts to postcode areas with low *Sun* readership and column (6) to areas with high *Sun* readership. Both are estimated on the same event set; the observation counts differ because each readership group’s daily series has its own days with zero applications, which cannot be logged. Every such day falls outside the event windows, so they enter only the day-of-week and time-trend terms. In columns (7)–(8), each date is assigned by its earliest qualifying article: pages 1–5 in column (7), page 6 or later in column (8). Event-selection restrictions are applied separately within each group. Column (9) restricts to articles fitting one of the fourteen tropes of section 2; column (10) restricts to negative-individual articles not fitting any trope. Newey–West HAC standard errors (bandwidth 7) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Event-study coefficients for disability-related hate crimes and Google Trends search volume.

	Disability hate crimes (PPML)		Google Trends job search (OLS)	
	Fake disability (1)	Other / none (2)	Negative (3)	Pos./neutral (4)
$\tau = -6$	0.1365** (0.0696)	-0.1408** (0.0574)	-0.0060 (0.0155)	0.0192 (0.0127)
$\tau = -5$	0.0518 (0.0951)	-0.1107** (0.0484)	0.0016 (0.0156)	0.0280** (0.0140)
$\tau = -4$	0.0640 (0.0678)	-0.0408 (0.0568)	0.0040 (0.0170)	0.0263** (0.0116)
$\tau = -3$	-0.0188 (0.0831)	-0.0783 (0.0480)	-0.0110 (0.0151)	0.0010 (0.0129)
$\tau = -2$	0.0292 (0.0746)	-0.0152 (0.0534)	0.0080 (0.0103)	0.0158 (0.0097)
$\tau = 0$	0.0130 (0.0707)	-0.0422 (0.0523)	-0.0466** (0.0203)	-0.0107 (0.0145)
$\tau = 1$	0.1506** (0.0743)	-0.0440 (0.0482)	-0.0514** (0.0232)	-0.0028 (0.0111)
$\tau = 2$	0.0895 (0.0738)	-0.0545 (0.0542)	-0.0269 (0.0199)	-5.78×10^{-5} (0.0122)
$\tau = 3$	0.1069 (0.0759)	-0.1325*** (0.0510)	-0.0350* (0.0180)	-0.0121 (0.0116)
$\tau = 4$	0.0298 (0.0760)	-0.1555*** (0.0569)	-0.0457** (0.0222)	-0.0287** (0.0123)
$\tau = 5$	0.1588** (0.0714)	-0.1603*** (0.0501)	-0.0435** (0.0222)	-0.0278* (0.0144)
Observations	3,652	3,652	3,597	3,597
R ²			0.72900	0.71412
Within R ²			0.71229	0.70201
Pseudo R ²	0.51215	0.51348	0.44664	0.42836
N. of events	57	64	61	45
Event fixed effects	✓	✓	✓	✓

Notes: Columns (1)–(2) report PPML event-study coefficients on the daily count of disability-related hate crimes summed across English police forces. Column (1) restricts to negative-individual *Sun* articles carrying the *Fake Disability Claimant* trope; column (2) restricts to negative-individual articles carrying any other or no trope. Columns (3)–(4) report OLS event-study coefficients from equation (1), with log Google Trends search volume in the “Job” topic as the outcome. Column (3) is estimated around negative-individual *Sun* articles; column (4) around positive/neutral-individual articles. Newey–West HAC standard errors (bandwidth 7) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

around articles carrying the *Fake Disability Claimant* trope versus any other or no trope, plotted in figure 14. Columns (3)–(4) use log Google Trends search volume in the “Job” topic, around negative- and positive/neutral-individual articles respectively, plotted in figure 15 (the job-search analysis of section 4.4).

F.1 Additional heterogeneity results

Figure 25 shows the results of estimating the negative-individual event study within each of the four readership $\times$ income cells. Point estimates remain larger in the high-readership cells within both income halves, and the income profile remains flat within both readership halves.

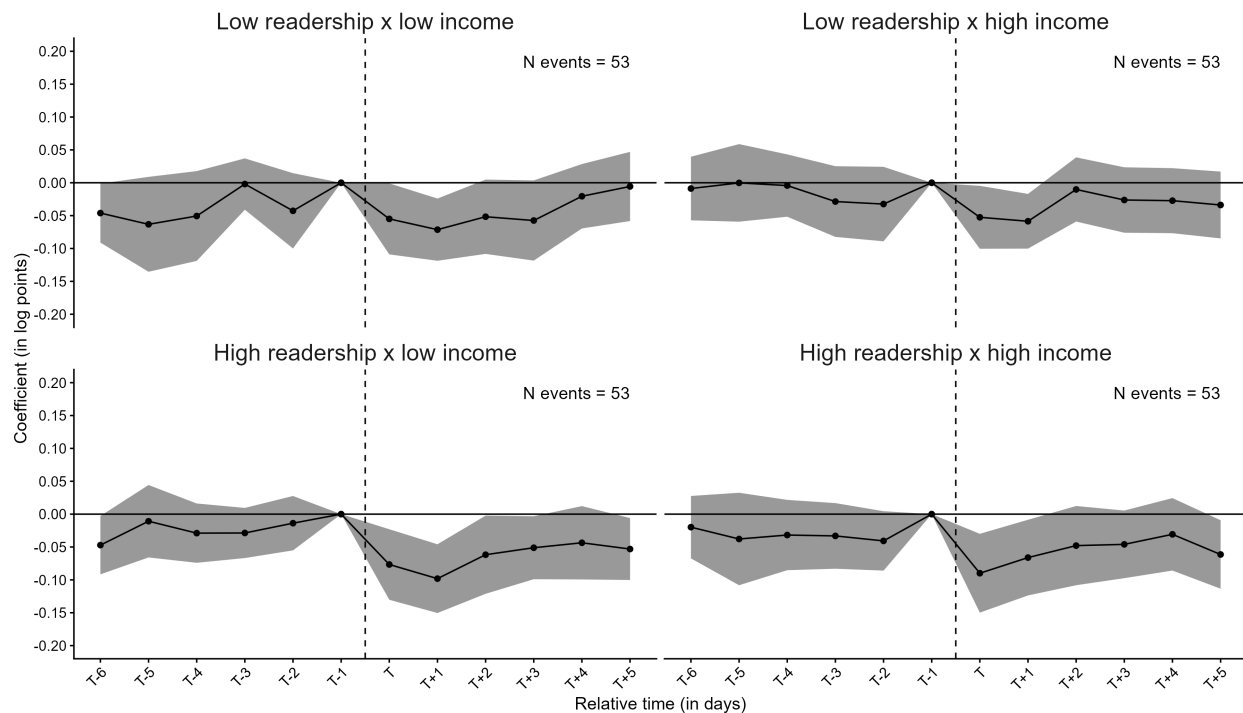


Figure 25: Event study by *Sun* readership and 2018/19 average income, split at the median.

Notes: Each panel re-estimates the negative-individual event study on the subset of postcode areas in the indicated readership $\times$ income cell. Low / high readership is the binary split used in figure 11; low / high income corresponds to the bottom and top halves of the 2018/19 postcode-area income distribution. 3 of the 56 negative-individual events are excluded, as Universal Credit was rolled out to high-income postcode areas later than low-income postcode areas, leading to zero applications in high-income postcode areas for some early events.

G Robustness tests

G.1 Leave-one-event-out

We refit the main specification dropping each event in turn (figure 26). No single event drives the result: every leave-one-out path stays close to the baseline coefficients (black line). The publication-day coefficient ranges from -0.0698 to -0.0469 , and the following-day coefficient from -0.0887 to -0.0707 , in log points. These refits also give a jackknife estimate of the standard errors, which we report alongside the other inference methods in table 10.

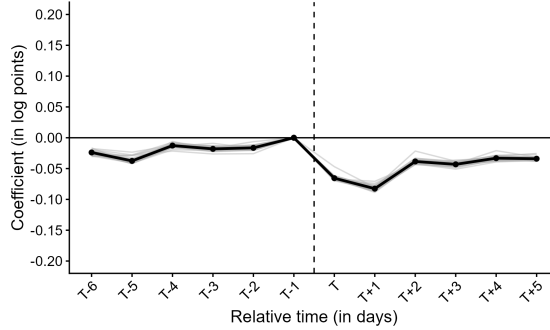


Figure 26: Leave-one-event-out analysis.

Notes: The graph shows the coefficients in the main regression, leaving each event out one at a time. The black line represents the coefficient in the main regression.

G.2 Accounting for day-of-week, time trends and functional form

We consider alternative ways of controlling for day-of-week and the time trend (table 8, columns (1)–(5)). The estimated drop is insensitive to how the weekly and secular trends are absorbed. We also re-estimate the event study by PPML on the daily count of applications (column (6)), rather than OLS on the log, and find a similar pattern, with a significant drop on the two days following publication.

G.3 Event selection

We first vary the window length, shrinking it to 10 days (5+5) or widening it to 14 days (7+7), either side of the 12-day default (table 9, columns (1)–(2)). The coefficients track the baseline at both lengths, though results are noisier for the 14-day window due to the reduction in the number of events.

Table 8: Time trend and functional form robustness: event-study coefficients.

	Poly. 3 (1) OLS	Poly. 7 (2) OLS	Spline (20 df) (3) OLS	DoW $\times$ year (4) OLS	Poly. 5 + DoW $\times$ year (5) OLS	PPML (6) Poisson
$\tau = -6$	-0.0158 (0.0212)	-0.0259 (0.0217)	-0.0236 (0.0222)	-0.0254 (0.0241)	-0.0219 (0.0230)	-0.0216 (0.0170)
$\tau = -5$	-0.0219 (0.0288)	-0.0451 (0.0288)	-0.0316 (0.0330)	-0.0334 (0.0310)	-0.0325 (0.0308)	0.0064 (0.0252)
$\tau = -4$	-0.0062 (0.0198)	-0.0181 (0.0206)	-0.0374 (0.0228)	-0.0073 (0.0230)	-0.0059 (0.0224)	-0.0078 (0.0218)
$\tau = -3$	-0.0195 (0.0210)	-0.0190 (0.0204)	-0.0360* (0.0213)	-0.0274 (0.0241)	-0.0244 (0.0225)	0.0136 (0.0172)
$\tau = -2$	-0.0161 (0.0207)	-0.0116 (0.0206)	-0.0185 (0.0207)	-0.0174 (0.0237)	-0.0166 (0.0226)	-0.0019 (0.0148)
$\tau = 0$	-0.0601*** (0.0230)	-0.0668*** (0.0242)	-0.0664*** (0.0247)	-0.0677*** (0.0237)	-0.0673*** (0.0235)	-0.0418** (0.0171)
$\tau = 1$	-0.0751*** (0.0235)	-0.0852*** (0.0246)	-0.0840*** (0.0254)	-0.0735*** (0.0268)	-0.0754*** (0.0258)	-0.0348** (0.0172)
$\tau = 2$	-0.0235 (0.0228)	-0.0462* (0.0242)	-0.0344 (0.0274)	-0.0240 (0.0258)	-0.0286 (0.0252)	-0.0112 (0.0202)
$\tau = 3$	-0.0374* (0.0199)	-0.0488** (0.0221)	-0.0692*** (0.0232)	-0.0274 (0.0226)	-0.0314 (0.0220)	-0.0049 (0.0206)
$\tau = 4$	-0.0353 (0.0221)	-0.0344* (0.0207)	-0.0507** (0.0233)	-0.0319 (0.0255)	-0.0343 (0.0241)	-0.0146 (0.0173)
$\tau = 5$	-0.0344* (0.0200)	-0.0297 (0.0205)	-0.0360 (0.0225)	-0.0248 (0.0216)	-0.0293 (0.0208)	-0.0219 (0.0196)
Observations	3,204	3,204	3,204	3,204	3,204	3,204
R ²	0.93801	0.95005	0.96194	0.93404	0.95265	
Within R ²	0.93142	0.94474	0.95789	0.00034	0.28235	
Pseudo R ²	0.66739	0.71920	0.78443	0.65249	0.73204	0.78248
Event fixed effects	✓	✓	✓	✓	✓	✓
Day of week $\times$ year fixed effects				✓	✓	

Notes: Event-study coefficients for the negative-individual sample, under alternative functional forms for the time trend and the outcome. Columns (1)–(5) estimate the baseline OLS specification on log daily Universal Credit applications, varying how the secular and day-of-week trends are absorbed: column (1) uses a degree-3 polynomial in day-number interacted with day-of-week; column (2) a degree-7 polynomial interacted with day-of-week; column (3) a natural cubic spline with 20 degrees of freedom interacted with day-of-week; column (4) day-of-week interacted with year fixed effects; and column (5) the same day-of-week $\times$ year fixed effects plus an additive degree-5 polynomial trend in day-number. Column (6) instead estimates the event study by PPML on the daily *count* of applications. All columns are estimated on the same sample: days with a zero recorded count are excluded from every specification, since the applications data are zero edited rather than recording true zeros, so the columns differ in functional form alone. All specifications include event fixed effects. Newey–West HAC standard errors (bandwidth 7) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Next we impose a two-sided overlap restriction, dropping events if any other event window overlaps either the pre- or post-period, rather than only the pre-period (table 9, column (3)). The stricter rule retains fewer events but leaves the same pattern.

We then drop Saturday-published events – the largest weekly application drop falls Friday-to-Saturday, where imperfect day-of-week control would bite hardest – and, separately, add the 14 March 2020 event back into the treatment set despite its post-window overlapping the early COVID-19 surge in applications (table 9, columns (4) and (6)). Neither change moves the estimate.

Finally we vary the sample period (table 9, column (5)). Restricting the daily series to June 2016 onwards, after live service had reached all of England (full service by December 2018), removes the steepest part of the roll-out-driven trend. The estimate is unchanged.

G.4 Standard errors

We re-compute inference under alternative standard errors: Newey–West HAC with a data-driven bandwidth in place of our fixed bandwidth of 7, standard errors clustered at the event level, and the jackknife over events described above. Table 10 reports the standard errors and p -values under each. Standard errors are virtually unchanged.

G.5 Randomisation inference

We conduct Fisher randomisation inference, re-fitting the event study on 10,000 placebo permutations of the event dates to construct an empirical null distribution for each relative-time coefficient (figure 27, table 10). Each replicate draws the same number of candidate event dates as the actual pre-overlap pool of negative-individual *Sun* publications, sampling without replacement from the analysis-window calendar; the day-of-week sampling probabilities are set to match the empirical day-of-week distribution of those publications. Each replicate then applies the same one-sided non-overlap restriction and public holiday drops as the real pipeline before re-fitting the main event-study specification on the resulting placebo events. The reported two-sided p -value is the fraction of replicates whose absolute coefficient estimate is at least as large as the absolute actual coefficient estimate. The actual estimates for $\tau = 0$ and $\tau = 1$ fall in the tails of the placebo distribution, confirming the main-specification inference.

Table 9: Event selection robustness: event-study coefficients.

	5+5 (1)	7+7 (2)	6+6 two-sided (3)	No Sat (4)	Post-rollout (5)	Include COVID (6)
$\tau = -7$		-0.0123 (0.0220)				
$\tau = -6$		-0.0199 (0.0248)	-0.0428 (0.0310)	-0.0181 (0.0257)	-0.0406** (0.0201)	-0.0294 (0.0239)
$\tau = -5$	-0.0200 (0.0227)	-0.0215 (0.0356)	-0.0721 (0.0485)	-0.0623* (0.0334)	-0.0420 (0.0287)	-0.0399 (0.0293)
$\tau = -4$	0.0004 (0.0169)	-0.0050 (0.0256)	-0.0469* (0.0278)	-0.0153 (0.0276)	-0.0435** (0.0195)	-0.0168 (0.0216)
$\tau = -3$	-0.0114 (0.0193)	0.0286 (0.0211)	-0.0079 (0.0234)	-0.0244 (0.0260)	-0.0319 (0.0205)	-0.0221 (0.0217)
$\tau = -2$	-0.0052 (0.0178)	-0.0070 (0.0239)	-0.0154 (0.0233)	-0.0107 (0.0267)	-0.0216 (0.0189)	-0.0188 (0.0214)
$\tau = 0$	-0.0542*** (0.0193)	-0.0726** (0.0335)	-0.0645** (0.0256)	-0.0730** (0.0290)	-0.0865*** (0.0283)	-0.0648*** (0.0234)
$\tau = 1$	-0.0751*** (0.0209)	-0.0507** (0.0254)	-0.1117** (0.0447)	-0.0888*** (0.0289)	-0.0644*** (0.0194)	-0.0778*** (0.0244)
$\tau = 2$	-0.0360 (0.0247)	-0.0446 (0.0317)	-0.0629** (0.0276)	-0.0777** (0.0302)	-0.0544** (0.0266)	-0.0309 (0.0246)
$\tau = 3$	-0.0389** (0.0178)	-0.0358 (0.0218)	-0.0327 (0.0347)	-0.0519* (0.0285)	-0.0539** (0.0210)	-0.0245 (0.0270)
$\tau = 4$	-0.0516** (0.0203)	-0.0108 (0.0213)	-0.0553* (0.0315)	-0.0366 (0.0286)	-0.0390* (0.0210)	-0.0088 (0.0310)
$\tau = 5$		-0.0331 (0.0226)	-0.0706* (0.0366)	-0.0321 (0.0255)	-0.0297 (0.0214)	-0.0054 (0.0325)
$\tau = 6$		-0.0466* (0.0239)				
Observations	3,204	3,204	3,204	3,204	2,130	3,204
R ²	0.94345	0.94118	0.94047	0.94256	0.85159	0.94293
Within R ²	0.93553	0.93646	0.93848	0.93840	0.82827	0.93653
N. of events	79	35	18	43	44	57
Event fixed effects	✓	✓	✓	✓	✓	✓

Notes: Event-study coefficients on log daily Universal Credit applications for the negative-individual sample, under alternative window-length, overlap, and sample choices. Columns (1)–(3) use alternative window-length or overlap rules; column (4) drops Saturday-published events; column (5) restricts the daily series to June 2016 onwards (after live service was rolled out across England); column (6) adds the 14 March 2020 event back into the treatment set. Newey–West HAC standard errors (bandwidth 7) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Inference under alternative methods: standard errors and p -values for the main event-study coefficients.

	$\tau = -6$	$\tau = -5$	$\tau = -4$	$\tau = -3$	$\tau = -2$	$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
Coefficient estimate	-0.024	-0.038	-0.013	-0.018	-0.016	-0.066	-0.083	-0.038	-0.043	-0.033	-0.034
<i>Panel A: Standard errors</i>											
Newey–West, bandwidth 7 (baseline)	0.022	0.029	0.020	0.021	0.021	0.024	0.024	0.024	0.021	0.022	0.021
Newey–West, data-driven bandwidth	0.024	0.029	0.022	0.022	0.022	0.024	0.025	0.022	0.022	0.023	0.021
Clustered by event	0.021	0.028	0.019	0.021	0.021	0.024	0.024	0.024	0.020	0.023	0.019
Jackknife over events	0.019	0.028	0.019	0.020	0.020	0.023	0.023	0.024	0.019	0.022	0.019
<i>Panel B: p-values</i>											
Newey–West, bandwidth 7 (baseline)	0.275	0.194	0.534	0.384	0.441	0.005	0.001	0.107	0.043	0.138	0.098
Newey–West, data-driven bandwidth	0.311	0.202	0.567	0.409	0.448	0.005	0.001	0.087	0.055	0.143	0.109
Clustered by event	0.264	0.189	0.509	0.385	0.435	0.009	0.001	0.122	0.040	0.164	0.085
Jackknife over events	0.225	0.190	0.496	0.360	0.426	0.006	0.001	0.113	0.027	0.141	0.079
Randomisation inference	0.435	0.240	0.670	0.497	0.523	0.025	0.005	0.168	0.140	0.281	0.278

Notes: The first row reports coefficient estimates, in log points, from the main negative-individual event study. Panel A reports standard errors and panel B two-sided p -values under Newey–West HAC standard errors with our baseline bandwidth of 7 and with a data-driven bandwidth, standard errors clustered by event, and a jackknife over events. Panel B also reports p -values from randomisation inference over 10,000 placebo draws of the event dates.

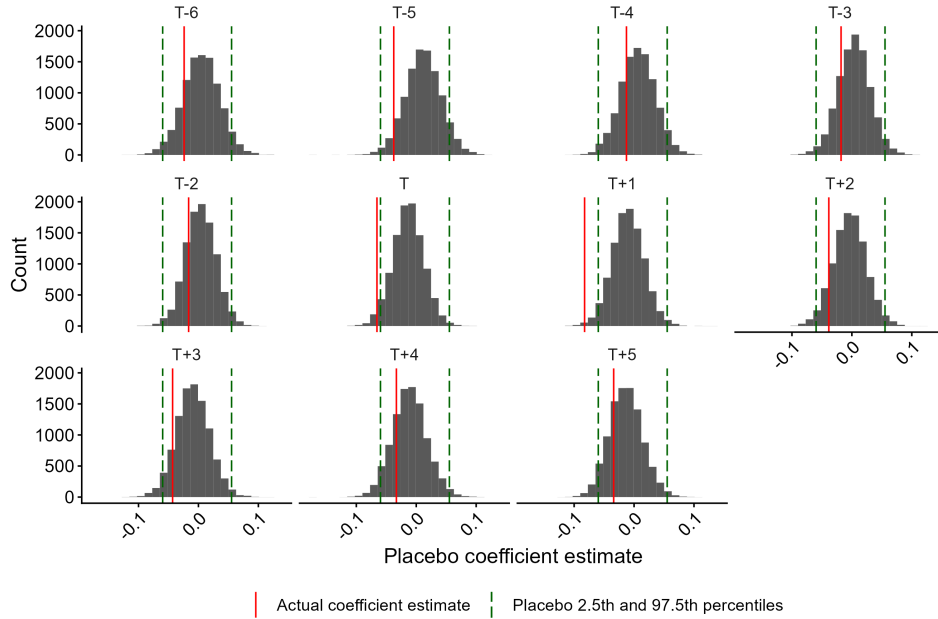


Figure 27: Randomisation inference for the main regression: placebo coefficient estimates.

Notes: The event days are randomly re-drawn 10,000 times. The event-study procedure described above (restricting to non-overlapping windows and estimating equation (1)) is applied to every draw. The coefficient estimates for each draw are then plotted in the histograms for each relative time. The red line is the coefficient estimate from the actual event dates. The dark green lines are the 2.5th and 97.5th percentiles of the placebo coefficient estimates.

H Back-of-the-envelope calculations

This appendix gives the sources and calculations behind the magnitudes reported in section 4.2 and in the discussion of job search in section 4.4. All Universal Credit figures are 2019 values, so that they predate the COVID-19 pandemic, and their euro and dollar equivalents are converted at 2019 annual average exchange rates.

H.1 Cost to individuals

The average monthly award of £660 is the mean award for households where Universal Credit was in payment in February 2019, published by the DWP in *Universal Credit statistics to 11 April 2019*.¹² To compare the cost of a story with household income, we use the DWP’s *Universal Credit Full Service Claimant Survey* (2018)¹³ of recent claimants, in which the median household falls in the £10,000–£20,000 annual income band. We convert this band to 2019 prices using the ONS Consumer Prices Index.

H.2 Expected Universal Credit spell length

The upper bound on the cost to individuals multiplies the average monthly award by the expected length of a Universal Credit claim. We estimate this expected spell length from two Stat-Xplore tables, the number of people on Universal Credit by duration band and the number of starts, both monthly. Everyone in band $[a, b)$ at snapshot t began claiming between a and b months before t , so the expected count in the band is $\sum_{u=a}^{b-1} \text{starts}(t-u) S(u)$, where S is the survivor function. We estimate S by Poisson likelihood on these counts, pooling the seven bands up to five years across the twelve monthly snapshots of 2019. Using the starts in each individual month, rather than a band average, matters because starts grew steeply during roll-out.

We model S as a mixture of two exponentials: the fit places 14% of claims in a short-spell group averaging 1.9 months, consistent with claims abandoned during the five-week wait, and 86% in a long-spell group averaging 28 months, giving an expected spell of 24.2 months. Weibull and negative binomial specifications fit almost as well and, like a geometric, imply a similar expected spell of 23.5 to 26.4 months. Log-normal and log-logistic fit worse than

¹²See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/801411/universal-credit-statistics-to-11-april-2019.pdf.

¹³See <https://assets.publishing.service.gov.uk/media/5b1a4f9eed915d2cc380163f/universal-credit-full-service-claimant-survey.pdf>.

either and imply implausibly long spells of 38 and 56 months, and a Poisson worse still by two orders of magnitude, since it implies a hazard rising with duration where the data show one falling sharply. Because the bands stop at five years, the tail beyond that point is extrapolated from the fitted model rather than observed. Stat-Xplore records duration for people rather than households, so this is a spell length for individuals, while the award it multiplies is a household mean; the two coincide for the 86% of Universal Credit households that consisted of a single adult in February 2019.

H.3 Counterfactual take-up

Let t^c denote the counterfactual take-up rate, e our estimated proportional reduction in applications on a treated day (the mean of the $\tau = 0$ and $\tau = 1$ coefficients), and s the share of days that are treated. Take-up is $t^c(1 - e)$ on a treated day and t^c otherwise, so the observed average take-up rate is $\bar{t} = t^c(1 - se)$. The counterfactual rate is therefore recovered as

$$t^c = \frac{\bar{t}}{1 - se}.$$

H.4 Entitlement changes equivalent to the discourse effect

For each study, we convert the reported effect of an increase in entitlement on take-up into the share of baseline non-take-up that it closes. We then compute the increase needed to close the 4.7% of non-take-up that we attribute to negative stories.

Adam and Brewer (2005) estimate how take-up responds to greater entitlement, evaluating this response at the median predicted probability of take-up among lone parents in each of two datasets: averaging across years, a 10% increase in Family Credit entitlement raises take-up by 1.68 percentage points on a baseline take-up rate of 78.7% in the Family Resources Survey, and by 2.10 percentage points on a baseline of 82.5% in the Families and Children Survey. Against baseline non-take-up of 21.3% and 17.5%, a 10% increase therefore closes $1.68/21.3 = 7.9\%$ and $2.10/17.5 = 12.0\%$ of non-take-up, so closing 4.7% requires an increase of between $10 \times 4.7 / (100 \times 2.10 / 17.5) = 3.9\%$ and $10 \times 4.7 / (100 \times 1.68 / 21.3) = 6.0\%$. Their comparable estimates under Working Families' Tax Credit, the more generous program that replaced Family Credit, average 1.00 and 1.33 percentage points, respectively. Applying the same calculation gives required entitlement increases of 6.2–10.0%.

Anderson and Meyer (1997) estimate that a 10% increase in the weekly benefit amount raises take-up by 2–2.5 percentage points, and assume a base take-up rate of between 40%

and 60%. Taking the midpoint gives baseline non-take-up of 50%, so a 10% increase closes $2/50 = 4.0\%$ to $2.5/50 = 5.0\%$ of non-take-up, and closing 4.7% requires an increase of between $10 \times 4.7/5.0 = 9.4\%$ and $10 \times 4.7/4.0 = 11.8\%$.

Whelan (2010) estimates that an extra C\$100 per month raises take-up by 5.2–6.8 percentage points, on a take-up rate of 34% in his estimation sample and hence baseline non-take-up of 66%. An extra C\$100 therefore closes $5.2/66 = 7.9\%$ to $6.8/66 = 10.3\%$ of non-take-up, and closing 4.7% requires between $100 \times 4.7/(100 \times 6.8/66) = \text{C\$}46$ and $100 \times 4.7/(100 \times 5.2/66) = \text{C\$}60$ per month; against his reported mean entitlement of C\$608 per month this is an increase of 8–10%.

Bruckmeier and Wiemers (2012) estimate that an extra €100 per month raises take-up by 5.8 percentage points. Their rate of persistent non-take-up runs from 48.9% in 2005 to 41.3% in 2007, and taking the midpoint of 45.1%, an extra €100 closes $5.8/45.1 = 12.9\%$ of non-take-up, so closing 4.7% requires €37 per month. They do not report a mean entitlement, so this cannot be expressed as a proportional increase.

Currency equivalents of these estimates are converted at annual average rates from the Federal Reserve H.10 series, for the year each estimate refers to: 1997 for Whelan (2010) and 2006, the midpoint of the 2005–2007 sample, for Bruckmeier and Wiemers (2012). The euro did not exist in 1997, so the euro figures for Canada use the 1999 rate, the earliest available. No adjustment is made for inflation, so these amounts are not comparable in real terms with the Universal Credit figures in section 4.2.

H.5 Fiscal cost of reduced job search

We value the additional days that households remain on Universal Credit after a negative shock at the average daily Universal Credit amount per household of £22 (section 4.2). Google Trends measures search in the population as a whole, so we assume that job search among claimants responds in the same proportion as job search among everyone else. We further assume that the rate at which claimants leave Universal Credit moves one-for-one with search intensity.