# Prolific Poll: Methodology

## How does Prolific collect its polling data?

### Who are the Prolific participants?

The Prolific platform houses over 140,000 active participants from over 30 countries worldwide. Participants sign up to the platform mainly through word-of-mouth or, in some cases, due to referrals from adverts or marketing.

Prolific participants are highly vetted, including ID, phone, location, and bank account verification. They're tracked for key response quality metrics throughout their time on the platform. We remove low-quality participants to keep the overall pool quality high.

We also collect over 300 pieces of demographic information about every participant in its pool, giving us a holistic view of each one. This is in stark comparison to other polling and panel agencies, which do no form of vetting and collect very little in the way of demographic or quality criteria.

All participants on Prolific receive payment for completing studies in cash and at a fair hourly rate (at least £6/hr). In comparison, other polling and panel agencies pay their participants very little or not at all. Our participants are therefore highly motivated to stay engaged and provide truthful and attentive responses to survey questions. For evidence of how we compare to other providers, please see here and here.

### Poll release schedule

We release our polls every two weeks on a Monday at midday. The poll is left open and available until Wednesday morning or until all available spaces in the poll are filled.

Polls specifically ask about the previous two weeks in a respondent's life, hence why we keep the release schedule fortnightly.

### How do we collect the poll sample?

For polling, we recruit a Prolific representative sample of 2,000 UK-based participants, stratified by their age, gender, and ethnicity (according to the UK census). Poll spaces

are open to any UK participants that fit the required demographic criteria. They're filled on a first-come first-served basis.

Our backend survey distribution ensures that the final sample corresponds with the most recent census-level national splits of age, gender, and ethnicity.

## What information do we capture?

The Prolific poll asks participants 18 long-term trend questions about mood, political discourse, media, AI, environment, and cost of living. We also track participants' current approval rate of the government through a binary approve vs disapprove question ("Do you approve or disapprove of the current Government's record to date?").

We ask participants about voting intention by presenting the seven major parties expected to run in the upcoming election (Conservatives, Labour, Liberal Democrats, Reform UK, Greens, SNP, Plaid Cymru) along with an 'Other' option and an option for "I do not intend to vote". Participants are asked, "Which party do you intend to vote for in the upcoming general election?".

We also collect extra demographic information to enable our weighting process. We ask participants which party they voted for in the 2019 election (same 7 party options, with the Brexit party in the place of Reform UK) and the highest education level they've completed (census-matched options). Through the Prolific platform, we also have access to a participant's age, gender, and ethnicity by default.

For the full set of questions asked in this poll please see here.

# How do we analyse the data?

We take the data that we collect from our poll and run it through our custom post-stratification pipeline. This produces results that are highly representative of the UK population, and predict responses to poll items on a national level or for any combination of our four weighting factors:

- Gender (Male/Female)
- Age Bracket (18-24/24-49/50-64/65+)
- Ethnicity (White/Black/Asian/Mixed/Other)

- 2019 vote (Conservatives/Labour/Liberal Democrats/Reform UK/Greens/SNP/Other)

To produce the nationally representative data, we first construct a ‘population file’. This records the proportion of the UK population that fits within each combination of the four factors above. We construct these proportions by using a combination of 2021/2022 census data for England, Wales, Northern Ireland, and Scotland*, and information on average voting behaviours by age group from the British Election Study. We use this file in subsequent steps to post-stratify the raw data and make it more representative.

We process the raw survey data using a Multilevel Regression and Post-stratification (MRP) approach. MRP is a statistical technique used in survey research and political science to estimate population-level parameters, such as voting preferences or opinions, based on a combination of survey data and existing demographic information. This approach provides highly accurate election forecasts even from non-representative groups (for more detail on this method please see here).

The output is a set of responses for each polling question that represent the estimated responses to the question, on both a national level and for two-way combinations of demographic groups. This data can be used to provide insights into how different demographic segments within the population might respond to the poll questions asked.

# Understanding the output

Poll results are released weekly on OSF (here). Poll results are displayed as *predicted* response values for each poll item. We present these predictions three different levels.

Our primary result is the national level (Demographic Factors = “National’). These values represent predicted responses to the poll item for the UK as a whole (accounting for the age, gender, ethnic, and 2019 vote splits of the countries population). In other words, these results tell us what the country thinks once we’ve weighted our data by those factors.

The next level of results are single demographic predictions (Demographic Factors = ‘Gender’, ‘Age’, ‘Ethnicity’, ‘Party2019’). These values represent predicted responses for that specific demographic once we have weighted by the other three factors. These results are further split by each level of the demographic factor. For example, when Demographic Factors = ‘Gender’ and Demographic Level = ‘Female’, the results

represent how our model predicts females only would respond to the poll item (after accounting for age, ethnicity, and 2019 vote).

The final level of results are interacted demographic predictions (Demographic Factors = ‘Gender Age’, ‘Gender Ethnicity’, ‘Gender Party2019’, ‘Age Ethnicity’, ‘Age Party2019’, ‘Ethnicity Party2019’). These results represent predicated responses for interactions of two different demographics once we have weighted by the other two factors. These results are further split by each level of the demographic factors involved. For example, when Demographic Factors = ‘Gender Age’ and Demographic Level = ‘Female 18-24’, the results represent how our model predicts females aged 18-24 only would respond to the poll item (after accounting for ethnicity and 2019 vote).**

# Notes

** Scotland’s census was delayed due to the Covid pandemic and only high-level data on counts of the population by gender and age has been made available. Accordingly, we made the following assumptions when integrating this data into the UK-wide population file:*

- *Proportions of different ethnicities mirrored those observed within the data for England, Wales, and Northern Ireland.*

** At the level of multiple demographic interactions the data we are using to predict response becomes more sparse (i.e., we do not have many participants that fit those exact demographic combinations). The result of this is that predicted responses may weight more heavily on to one or other response option.