Internet Coverage Bias in Web Surveys in Europe

Alessandra Gaia[1,2] (iD) · Emanuela Sala[2] (iD) · Chiara Respi[2] (iD)

[1]University College London, Centre for Longitudinal Studies

[2]University of Milano-Bicocca

The use of web surveys has increased over the last decades in an attempt to reduce survey costs and maximise response rates; this trend has accelerated during the Covid-19 pandemic, when social distancing measures impede face-to-face data collection. Despite the widespread use of web surveys, the uncomplete Internet coverage may still pose a threat to data quality. Using large scale probability-based Eurobarometer data from 2010-2019 we: i) describe the trend in Internet coverage rate across Europe, ii) investigate demographic and socio-economic differences between the Internet and non-Internet population, iii) explore variation over time and across countries in Internet coverage bias, and iv) assess whether countries' socio-economic context is associated with Internet coverage bias. We find that a non-negligible share of the population does not use the Internet and Internet coverage varies widely across Europe. In addition, we document that coverage bias: decreases over time for most of the variables considered; seems more pronounced in age, education, and life satisfaction, and negligible in other variables; and is associated with countries' socio-economic context.

Keywords: web surveys; internet coverage; coverage bias; total survey error; data quality

1 Introduction

Over the last decades, web surveys have increasingly been used both as stand-alone probabilistic surveys, as well as opt-in non-probabilistic online panels (Tourangeau et al., 2013); also, web interviewing has been widely adopted as one of the modes of data collection in mixed-mode studies (e.g. *Understanding Society*: the UK Household Longitudinal Study; the National Child Development Study; The European Social Survey). Recently, the Covid-19 pandemic, and the associated social distancing measures, further accelerated the transition to mixed-modes web-first designs—e.g. in *Understanding Society*: The UK Household Longitudinal Study (Burton et al., 2020), or led to a switch from face-to-face to unimode web designs—e.g. in some

countries participating in the Standard Eurobarometer 96 (European Union, 2023a, b, c). Furthermore, web surveys have gained popularity not only when conducting different kinds of studies (e.g. opt-in panels, stand-alone surveys, mixed-mode studies), but also in a variety of different contexts, including academic research, journalism, market research, and official statistics.

While web surveys have been valued for their lower costs, faster data collection and privacy protection compared to interviewer-administered surveys, this mode of data collection has also drawbacks (Couper, 2017). For example, one issue associated with web surveys is the lack of coverage of the non-Internet population, i.e. members of the target population who do not have Internet access or lack the digital skills to participate (Mohorko et al., 2013; Sterrett et al., 2017), which may lead to coverage bias, when the Internet and non-Internet populations differ on the variables of interest.

On a more general level, coverage bias depends on two factors: the proportion of the population that is not covered by the sampling frame and the difference in the variable(s) of interest between those who are covered and those who are not covered (Eckman, 2016; Groves et al., 2009; Sterrett et al., 2017). Importantly, coverage bias is a property of a sampling frame and target population *on a specific statis-*

Corresponding author: Alessandra Gaia, Centre for Longitudinal Studies, University College London, 55–59 Gordon Square, London, WC1H 0NU, United Kingdom (Email: a.gaia@ucl.ac.uk)

tic (Groves et al., 2009). This means that bias is concept-dependent, i.e. it varies on a question-by-question basis.

Although research on coverage bias in web surveys has flourished in the last decade (e.g. Blom et al., 2017; Cornesse & Schauer, 2021; Dutwin & Buskirk, 2023; Eckman, 2016; Toepoel & Hendriks, 2016), the interest on this topic may partially fade out as larger shares of the population have access to the Internet. Indeed, Internet coverage has risen worldwide from 31% in 2011 to 63% in 2021, and from 72 to 90% in high income countries (World Bank, 2023). However, research on coverage bias in web surveys in Europe remains pivotal for four main reasons. First, while in Europe the non-Internet population is decreasing over time, some countries may still have high levels of non-coverage and cross-country differences may complicate comparative analysis (Vicente & Reis, 2012). Second, while the percentage of the non-Internet population is constantly shrinking, it has been argued that it is becoming different from the rest of the population, and, thus, bias may not decrease (Eckman, 2016). Third, as Internet coverage varies over time, coverage bias in web-only surveys complicates the analysis of trends: indeed, researchers may confound increases (or decreases) in coverage bias as changes over time in the concepts of interest (Sterrett et al., 2017). Finally, considerations around Internet coverage bias are important in deciding whether the inclusion of the offline population is worth the effort—e.g. whether to equip non-Internet users with devices/training to participate in a web survey or whether to administer a mixed-mode survey *versus* a unimode web design (for a discussion, see: Kocar & Biddle, 2023).

Against this background, the overall aim of this paper is the analysis of coverage bias in Europe over time. In doing so we extend, using more recent data, a study by Mohorko and colleagues (2013), which, to our knowledge, is the only study which analysed Internet coverage bias across most European countries and over an extended time period; the authors found that coverage bias varies across countries, is associated with country-level characteristics, and diminishes over the time period 2005–2009. While their work has been highly influential, it refers to the situation of over a decade ago: we believe that the fast changes in Internet take-up which occurred in the last decade require a new assessment of Internet coverage and coverage bias. Thus, we extend Mohorko et al.'s (2013) work by analysing the evolution of bias over the subsequent decade (i.e. 2010–2019), comparing and contrasting the different European countries.

Specifically, we answer the following research questions:

RQ1 What is the trend in Internet coverage across the different European countries?

RQ2 Are there demographic and socio-economic differences between the Internet and non-Internet population?

RQ3 Does Internet coverage bias change over time and across countries?

RQ4 What is the role of countries' socio-economic context?

In order to reach our research aims we use data from the 2010–2019 Eurobarometer survey series, which is deemed to have a very good coverage of the Internet and non-Internet population (e.g. Fuchs & Busse, 2009); these are the most recent Eurobarometer data which are collected through a unimode face-to-face design and, hence, contain information on both Internet users and non-users (while from 2020 onwards data are collected through a web unimode design for some selected countries).

We believe this study is an important contribution to the literature. First, as it provides a novel assessment of coverage bias across Europe; second, because the analysis of the trend in coverage bias over an extended time period gives indications for the forecast of future developments in bias in Europe, and, thus, provides important information for survey methodologists collecting data through web-interviewing in multi-country European studies as well as for substantive researchers conducting comparative research.

2   Literature Review

Since the early days of web surveys, survey methodologists have been concerned about the potential coverage bias due to part of the population not having access to the Internet. Indeed, from the early 2000s, the survey methodology literature documented marked demographic, socio-economic, attitudinal, and behavioural differences between the Internet and non-Internet population (Tourangeau et al., 2013), with implications for coverage bias in web surveys.

These differences were generally consistent across studies, showing how the most disadvantaged groups in society often lack the access to the Internet and/or the ability to use it efficiently. Not surprisingly, the phenomenon has obtained attention also from the sociological/media studies literature, with the term *digital divide*, coined in the 1990s by Lloyd Morrisett (Hoffman et al., 2000). Both fields of research (survey methodology and media studies) provide useful insights on coverage bias in web surveys (for a discussion, see: Dutwin & Buskirk, 2023).

2.1 Coverage Bias in Socio-demographic Variables

With reference to Europe the literature shows that Internet users are on average younger (Blom et al., 2017; Bosnjak et al., 2013; Leenheer & Scherpenzeel, 2013; Lipps & Pekari, 2016), more highly educated (Blom et al., 2015, 2017; Bosnjak et al., 2013), living in larger households (Blom et al., 2017; Leenheer & Scherpenzeel, 2013) and in urban areas (Blom et al., 2017). With reference to immigration status, in The Netherlands, the Internet population under-represents non-western immigrants (Leenheer & Scherpenzeel, 2013).

Similarly to research conducted in European countries, also research carried out in non-European countries, such as in Israel (Mesch & Talmud, 2011) and the United States (Antoun, 2015; Couper et al., 2007; Schonlau et al., 2009), found significant differences in demographic and socio-economic factors between the Internet and non-Internet population. For example, in the US, compared to non-users, Internet users are more likely to be younger—Antoun (2015) and, among the old age population: Couper et al. (2007) and Schonlau et al. (2009)—highly educated (Antoun, 2015; Couper et al., 2007; Schonlau et al., 2009), married (Couper et al., 2007; Schonlau et al., 2009), have higher income (Antoun, 2015; Schonlau et al., 2009), are more often home-owners (Schonlau et al., 2009) and live in urban/suburban areas (*versus* rural) (Couper et al., 2007). Also, ethnic minorities are under-represented in the Internet population, both in Israel (Mesch & Talmud, 2011) and in the US (Antoun, 2015; Couper et al., 2007; Mesch & Talmud, 2011; Schonlau et al., 2009). Overall, with only few exceptions, differences observed in Europe are consistent with the evidence arising from the US, both in terms of which variables are subject to bias as well as in the direction of bias.

2.2 Coverage Bias in Attitudinal and Behavioural Variables

In many cases, differences in sample composition are also accompanied by differences in attitudinal and behavioural variables. In Germany, for example, the Internet population is more interested in politics than the non-Internet population (Blom et al., 2017). Findings from the Netherlands shows that the Internet population is more likely to be liberal (e.g. less likely to be religious and supporter of Christian parties) than the non-Internet population, although there are no statistically significant differences when comparing the personality traits, social integration indicators, and leisure time activities of the two populations (Toepoel & Hendriks, 2016). Also, in comparison with non-Internet users, the Internet population has a higher level of civic and political activity (Zhang et al., 2009), community partici-

pation (Stern & Dillman, 2006; Zhang et al., 2009), more positive opinions and attitudes on the economic outlook (Valliant & Lee, 2005), and higher levels of openness and tolerance (Robinson et al., 2002); however, the Internet and non-Internet populations do not seem to differ with regard to other variables, such as social trust and quality of life (Antoun, 2015).

Differences between the Internet and non-Internet population are also marked when looking at population subgroups, such as, for example, the old age population—which, on average, is characterised by a lower use of Information Communication Technologies (ICTs) (Sala et al., 2022). Indeed, in the US, among older adults, Internet users report better health (Couper et al., 2007; Schonlau et al., 2009).

Differences between the Internet and non-Internet population does not seem to be eroded over time: a recent study from the US population (Dutwin & Buskirk, 2023) finds that, out of 542 variables considered, for 122 the estimate for the Internet population is at least double that for the non-Internet population, or *vice versa*, the estimate for the non-Internet population is at least double that for the Internet population. The authors observe that even if the non-Internet population is found to be rare (under 10 % of the total population), a survey which includes only Internet users would obtain biases of as much as 9% on several measures and more than 100 variables would show biases of at least 5% (Dutwin & Buskirk, 2023). Similarly, recent research conducted in Australia (Kocar & Biddle, 2023) found statistically significant differences between the Internet and the non-Internet population for more than half of all studied variables; however, in this case, the magnitude of differences, considered together with the size of the non-Internet population, led authors to conclude that differences (on attitudinal, behavioural, and factual measures) are, on average, minor.

2.3 Cross-country Analysis of Coverage Bias

Although several studies investigated Internet coverage bias in some European countries, little research has focused on multiple European countries simultaneously (exceptions are, for example: Mohorko et al., 2013; Schnell et al., 2017; Vicente & Reis, 2012). Adopting a cross-country approach, Schnell et al. (2017) documented that the Internet and non-Internet population in Europe and in the US differ with respect to general subjective health, with healthier people being more likely to be over-represented in the Internet population; also, Vicente and Reis (2012) have shown that Internet households in Europe are more likely to be more satisfied with different aspects of their life and less concerned about their financial situation. Also, the authors found that

Internet households have significantly different opinions on which are the most serious problems that the world is facing (Vicente & Reis, 2012). In addition, Brandtzæg and colleagues (2011), focussing on five European countries simultaneously (i.e. Norway, Austria, Sweden, UK, and Spain), profiled the Internet population according to its level of Internet use; the authors documented that being women, of old age, and belonging to a larger household increases the probability of belonging to the "non-users" type.

As mentioned in the previous section, we are aware of only one cross-country study that analyses the trend in Internet coverage bias in Europe over a prolonged time period, i.e. the study by Mohorko et al. (2013). Using data from the 2005–2009 Eurobarometer surveys, the authors described changes in Internet coverage bias focusing on several demographic (sex and age), socio-economic (education) and substantive variables (life satisfaction and political left-right self-placement). In addition, they also explored if and how these changes were related to the countries' socio-economic context, using several country-level indicators from Eurostat, the World Bank, and the Human Development Report. Performing multilevel analysis, the authors found: (i) a high level of cross-country variation in Internet coverage rate and bias, (ii) a negative trend between Internet coverage rate and bias (i.e. Internet coverage rate increases over time, while Internet coverage bias decreases), (iii) marked cross-country variation in the rate of change in Internet coverage rate and bias, and (iv) variability in the extent to which the different country-level indicators affect coverage bias for the variables under analysis (e.g., the Gini coefficient is related with Internet coverage bias for age, whereas it is unrelated with Internet coverage bias for life satisfaction). Following the publication of Mohorko et al.'s article, a similar study was conducted by Sterrett and colleagues in the US (Sterrett et al., 2017), documenting both similarities and differences between the European and the US case. In short, "the declines in coverage bias related to education and age were similar to those observed in Europe from 2005 to 2009 by Mohorko and colleagues (2013), although declines in bias associated with gender and life satisfaction were observed in Europe and not the United States" (p. 349).

3 Data

We use data from the 2010–2019 Standard Eurobarometer surveys in conjunction with country-level socio-economic indicators (e.g. the Gini coefficient) from supranational statistical bodies (e.g. Eurostat). Similar to Mohorko et al. (2013), we use the Eurobarometer surveys because this data source includes the non-Internet population (key to assess coverage bias) and collects information for all European (EU27) countries, over a long time period.

3.1 The Standard Eurobarometer Surveys

The Standard Eurobarometer surveys are part of the Eurobarometer Series, i.e. a series of surveys that the European Commission (EC) has conducted since the early 1970s on European Union (EU) member states and candidate countries. The Standard Eurobarometer surveys are probability-based repeated cross-sectional surveys conducted twice a year (in spring and autumn) on individuals aged 15 years old or older. They cover a wide range of topics, including life satisfaction, political attitudes, and Internet access.

The 2010–2019 Standard Eurobarometer surveys are individual level face-to-face surveys. The Standard Eurobarometer surveys considered in this analysis (covering years 2010–2019) adopt an address-based sampling design. This sampling design allows a high coverage of the general population (higher than telephone-list sampling frames, Groves et al., 2009). Indeed, as mentioned in Sterrett et al. (2017) (quoting Groves et al., 2009), "in-person surveys with address-based sampling designs offer the greatest coverage of the general population and can produce samples that are highly representative of both those with and without Internet" (p. 341).

Standard Eurobarometer surveys are based on a multistage, probability sampling design, with Primary Sampling Units (PSU) being selected randomly (from each of the administrative regions of each country) after stratification according to the distribution of the national, resident population in metropolitan, urban, and rural areas. Further information on the survey designs is available at GESIS (2021b) and European Union (2021).

Data from subsequent years—i.e. the 2020–2022 Standard Eurobarometer surveys, namely 93.1, 95.3, 96.3 (European Commission, 2020, 2021, 2022)—have been excluded from the analysis. This is because, for some countries, data is collected through a unimode web design, hampering the comparison between the Internet and non-Internet population.

Response rates vary by country and year; systematic information on response rates for each year under analysis and each country is not currently available. In 2019, response rates vary from 17% in Finland to 78% in The Netherlands (see Table A1 in the supplementary material)[1].

---

[1] No empirical data is available to assess whether there is a difference between the Internet population and the non-Internet population in their propensity to participate in the Eurobarometer face-to-face survey (i.e. nonresponse bias in Internet access/use). However, it should be noted that, while higher response rates are desirable, lower response rates are not *per se* an indication of higher non-response bias (Groves and Peytcheva 2008). While we did consider including non-response rates as covariates in our analysis to control for cross-country heterogeneity, this was not possible due to response rates not being available for most country-year.

The analysis presented in this paper is conducted on individuals aged 15 years or older living in 27 European Union Member States (i.e. Austria, Belgium, Bulgaria, Croatia, Cyprus[2], Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, The Netherlands) and the United Kingdom. Sample size is approximately 1000 observations for each country, except for small countries (e.g. Malta and Luxembourg) which have a sample size of approximately 500 observations. In this research we analyse the following datasets: Eurobarometer 92.3, 90.3, 88.3, 86.2, 84.3, 82.3, 80.1, 78.1, 76.3, 74.2 (European Commission, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019a, b). Data can be accessed at GESIS (2022).

3.2 Additional Country-level Data

Following Mohorko et al. (2013), we include in our analysis country-level data from the Eurostat and World Bank data warehouses. The Eurostat database (Eurostat, 2021) gathers harmonised data from various surveys or administrative sources, collected from the European countries on different topics. In this study, we use data from: (i) the national statistical institutes (for life expectancy), (ii) the Statistics on Income and Living Conditions (EU-SILC) study (for the Gini coefficient), and (iii) national surveys on consumer price and household expenditure (for inflation). The Eurostat database does not include information on the other variables of interest (e.g. education); we, therefore, also use data from the World Bank DataBank (World Bank, 2021), containing collections of time series data on a variety of topics from all the countries of the world. Specifically, we use four different data sources, namely (i) the UNESCO Institute for Statistics (for education), (ii) the United Nations World Urbanization Prospects (for urbanisation), (iii) the OECD National Accounts (for GDP growth), and (iv) the International Labour Organization (for employment).

4 Methods

To address our research questions, we replicated Mohorko et al. (2013)'s analytical approach.

4.1 RQ1: What is the trend in Internet coverage across the different European countries?

To describe the trend in Internet coverage, we use an indicator of coverage rate, calculated as the share of respondents who have access to and use the Internet (see section *Variables* for the variable description). When computing the coverage rate, we weighted the data using the Eurobarometer population size weights (provided with the datasets). These weights aim at ensuring that each country and each sub-sample (e.g. East and West Germany, Great Britain and Northern Ireland) are represented in proportion to the country/sub-sample population size and they also include the post-stratification weighting factors. For more information on the weighting procedure, please see the data documentation available at GESIS (2021a) and European Union (2021).

4.2 RQ2: Are there demographic and socio-economic differences between the Internet and non-Internet population?

To investigate the demographic and socio-economic differences between the Internet and non-Internet population, we compute the relative coverage bias (also used by Lessler & Kalsbeek, 1992, p. 60), which is described as follows:

$$\text{relative coverage bias} = \frac{\overline{y}_{\text{Int}} - \overline{y}_{\text{EB}}}{\overline{y}_{\text{EB}}}$$

Where EB is the total Eurobarometer survey sample (used in this analysis as a proxy for the target population), and *Int* represents the Internet sub-sample. Thus, for a given variable *y*, $\overline{y}_{\text{EB}}$ represents the mean of the Eurobarometer sample, and $\overline{y}_{\text{Int}}$ the mean of the Internet sub-sample (see section *Variables*, for the description of the variables used in the analysis). A positive/negative sign of relative coverage bias indicates an over-representation/under-representation of a given *y* variable in the Internet sub-sample. For example, a negative relative coverage bias on the continuous variable age indicates that older respondents are *underrepresented* in the Internet sub-sample. For categorical dichotomous variables (e.g. sex) a positive/negative sign indicates an over-representation/under-representation (in the Internet sub-sample) of the category coded with the highest value.

---

[2] Data from Cyprus do not include the Turkish Republic of Northern Cyprus.

4.3 RQ3: Does Internet coverage bias change over time and across countries?

When exploring the trend in the Internet coverage bias and the role played by country-level contextual factors, we turned to multilevel modelling, that is the ideal modelling strategy when dealing with multilevel datasets, such as ours. Specifically, our units of analysis are 280 country-years, nested into 10 years (first level) which are nested into 28 countries (second level). For example, the country-years Austria 2010, ..., Austria 2019 are nested into the years 2010, ..., 2019 and nested into the country Austria.

Our dependent variable is the absolute relative coverage bias (also adopted by Groves & Peytcheva, 2008), calculated for each of the five variables of interest (see section *Variables*) and computed as:

$$\text{absolute relative coverage bias} = \left| \frac{\overline{y}_{\text{Int}} - \overline{y}_{\text{EB}}}{\overline{y}_{\text{EB}}} \right|$$

We used absolute relative coverage bias to avoid that positive and negative values of relative coverage bias cancel out (and falsely give the impression of a small or null coverage bias) when analysing changes over time and across countries. Following Mohorko et al. (2013), in the dataset used for the regression analysis, the absolute coverage biases are expressed as percentage points for ease of interpretation.

Our independent variables are: time (years) and a number of socio-economic country-level indicators, described in the section *Variables*. For each dependent variable, we run two models. To assess variation in Internet coverage bias over time and across countries (RQ3), we run a model (Model 1) including as independent variables only time and time squared; time squared is included to capture non-linear relationships in the association between bias and time.

4.4 RQ4: What is the role of countries' socio-economic context?

To explore the role of countries' socio-economic context, as well as the differences between countries in the rate of change over time (RQ4), we run a second set of models (Model 2) which includes: (i) the variables included in Model 1 (i.e. time and time squared), (ii) the country-level variables, (iii) the interactions between the country-level variables and time. The country-level indicators capture whether changes in bias are associated with the socio-economic context of countries, whereas interaction terms indicate whether country-level contextual factors moderate the effect of time on bias.

Following Mohorko and colleagues (2013), we removed from the analysis any variable not resulting significantly associated with the dependent variable in all models. We perform the analysis using the software STATA, version 16.

5 Variables

In this section, we describe the variables used in the empirical analysis.

5.1 RQ1: What is the trend in Internet coverage across the different European countries?

There are different ways to conceptualise and measure Internet coverage, as documented in Tourangeau et al. (2013). As a proxy for Internet coverage, we use the variable "Internet use". This is a derived variable taking value 1 if the respondent reported having access *and* using the Internet, and 0 if the respondent reported either not using or not having access to the Internet or both. The questions used to create the derived variable is a battery of items, aiming at quantifying the frequency of Internet use in different contexts. Specifically, the question text reads: "Could you tell me if ...?", followed by the three items: "You use the Internet at home, in your home", "You use the Internet at your place of work", and "You use the Internet somewhere else (school, university, cyber-café, etc.)". To reflect changes in Internet use over time, in 2018 the following item was added to the original (three-item) battery question: "You use the Internet on your mobile device (laptop, smartphone, tablet, etc.)". For the four items, response categories are: "Every day or almost every day", "Two or three times a week", "About once a week", "Two or three times a month", "Less often", "Never". Respondents' spontaneous reports of not having access to the Internet were recorded by the interviewer as "no access". The derived variable Internet use takes value 1 for respondents reporting using the Internet in at least one of the three/four items (depending on the number of items included in the battery) and value 0 if respondents reported either that they have never used the Internet or that they have no access to the Internet *in all items*. The proxy for Internet coverage adopted in our study differs from the one used by Mohorko and colleagues (2013), i.e. Internet access at home. This is to reflect changes in the pattern of Internet use over time, as the authors suggested: "the use of mobile Internet on telephones and tablet devices is likely to increase further in the near future, which will necessitate a change in the measurement of Internet access" (p. 619). A more "inclusive" measure of Internet access was also adopted in Sterrett et al. (2017)'s study on Internet coverage bias in the United States; indeed, the au-

thors included Internet access *via* mobile devices in their definition of Internet access.

5.2 RQ2: Are there demographic and socio-economic differences between the Internet and non-Internet population?

Coverage bias is assessed with respect to three demographic and socio-economic variables and two substantive variables. For the demographic and socio-economic variables, we used the same variables adopted by Mohorko and colleagues (2013), i.e. sex, age, and length of education. These variables are important correlates of many substantive variables used in social and economic research. Sex takes value 1 if the respondent is a male and 2 if the respondent is a female, age is a continuous variable taking values 15–98, and length of education is a continuous variable measuring respondents' age when completing full-time education. For respondents who are still in full-time education at the time of the interview (e.g. 6% in 2019), we have imputed their age at the time of the interview.

As substantive variables, Mohorko et al. (2013) considered life satisfaction and political left-right self-placement, two important socio-political indicators. For life satisfaction, we use the same question adopted by Mohorko et al. (2013). Life satisfaction is measured through a 4 items scale, using the following question wording "On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?". Response categories also include the "Don't know" answer. Because of changes in the Eurobarometer questionnaire content, we could not use political left-right self-placement in our study (the variable is not available in three of the years under analysis, i.e. 2011–2013). We used instead the variable *assessment of the national economic situation or condition*. This is measured through the question: "How would you judge the current situation in each of the following?" where one of the items is: "The situation of the (NATIONALITY) economy" (response categories are: "Very good", "Rather good", "Rather bad", "Very bad", and "Don't know"). When computing the relative coverage bias, both variables are considered as quasi-cardinal variables and "Don't know" values (e.g. respectively 0 and 2% in 2019) are excluded from the analysis.

5.3 RQ3: Does Internet coverage bias change over time and across countries? RQ4: What is the role of countries' socio-economic context?

To assess variation over time and across countries in Internet coverage bias, we use time (i.e. years, coded: 2010 =

0 … 2019 = 10) and nine demographic and country-level socio-economic indicators: life expectancy at birth, Gini coefficient, inflation, duration of primary and secondary education, gross enrolment ratio (GER), urbanisation, Gross Domestic Product (GDP) growth, and labour force rate. Data are not available for some of the country-years considered (i.e. Gini Coefficient was missing for the United Kingdom in 2019 and GER was missing for 15% of country-years considered). Thus, we have imputed missing data estimating the average between the year before and the subsequent year for which the information is available (typically, the year after). If missing values occur in the last year (2019), and thus by definition subsequent observations are not available, data are imputed through the Last Information Carried Forward method (LICF, also called Last Observation Carried Forward), i.e. an imputation technique that imputes missing data with the most recent available information (Salkind, 2010).

As an indicator of the level of economic development, we use GDP growth, measured as the annual percentage growth rate at market prices based on constant local currency. Gini coefficient of equivalised disposable income is a standard economic measure of income inequality, measured on a scale that ranges from 0–100. Inflation is measured using the annual average rate of change of the Harmonised Index of Consumer Prices (HICP). The labour force participation rate is the proportion of the population aged 15 and older that is economically active (i.e. employed and unemployed but actively searching for a job).

As proxies for countries' human capital, we use three different variables, i.e., the theoretical duration in years of primary and secondary education and the yearly GER, measuring the total enrolment in primary, secondary, and tertiary education. This indicator is expressed as a percentage of the total population of primary school age, secondary school age, and the five-year age group following on from secondary school leaving. In some instances, the GER may exceed 100% due to the inclusion of over-aged and under-aged students (e.g. early or late school entrance and grade repetition). We also consider urbanisation, measured as the percentage of people living in urban areas as defined by national statistical offices.

6 Results

In this section, we shall discuss the findings from our analysis, focussing on each research question separately.

Table 1

Coverage rate by European country (2010–2019)

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Delta
Austria	72	74	75	78	77	78	82	85	85	86	14
Belgium	73	76	83	83	84	84	86	89	90	92	20
Bulgaria	52	52	55	59	60	64	62	69	71	74	22
Croatia	55	56	61	63	71	71	68	74	77	82	27
Cyprus	52	50	57	64	69	68	71	74	75	80	28
Czech Republic	67	71	75	78	78	76	81	82	82	86	18
Denmark	88	91	93	92	94	95	93	95	95	96	7
Estonia	72	76	78	79	80	78	79	80	83	84	13
Finland	82	80	85	86	86	87	87	88	90	90	8
France	75	76	78	82	81	81	84	84	86	87	12
Germany	75	72	79	78	80	83	82	86	86	88	13
Greece	50	54	59	60	60	62	65	66	71	72	22
Hungary	55	59	64	66	65	68	70	70	76	80	24
Ireland	77	76	85	85	84	84	87	87	89	90	13
Italy	64	68	70	70	72	71	71	76	80	81	17
Latvia	72	78	78	81	76	74	77	79	83	82	11
Lithuania	61	66	68	70	70	71	71	71	75	78	17
Luxembourg	80	85	85	83	88	85	89	91	93	95	15
Malta	65	64	66	71	72	74	76	78	74	78	13
Poland	61	64	66	67	76	73	72	75	75	77	15
Portugal	43	45	45	51	59	65	67	72	72	74	31
Romania	48	50	51	56	59	57	59	60	64	70	22
Slovakia	72	72	73	76	72	75	77	73	75	79	7
Slovenia	69	72	74	75	74	75	78	79	79	80	11
Spain	61	62	68	69	71	69	73	77	79	82	21
Sweden	90	94	95	96	93	95	94	96	97	98	8
The Netherlands	92	94	92	96	97	96	97	99	99	99	7
United Kingdom	77	80	82	82	86	87	90	89	90	90	13
Europe	69	70	74	75	73	77	79	81	83	85	16

Data are weighted with the Eurobarometer population size weights
Population size weights also include the post-stratification weighting factors
"Delta" represents the difference between the share of adults using the Internet in 2019 and 2010
"Europe" includes the 27 European Union Member States and the United Kingdom
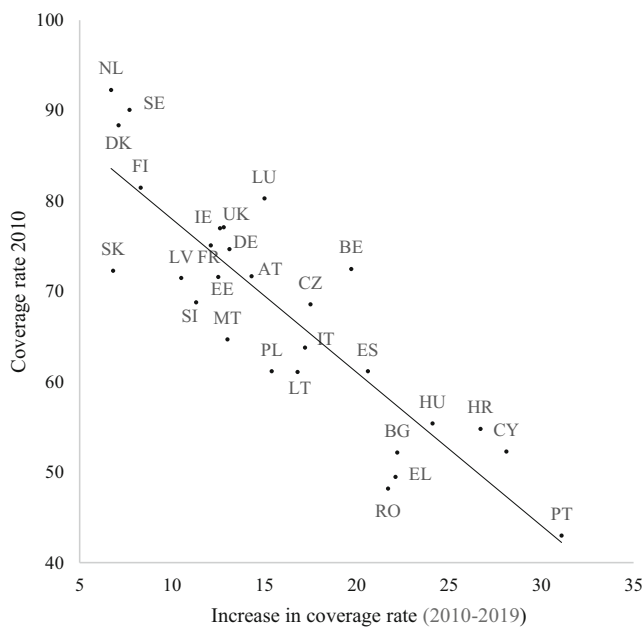Source: authors' elaboration on Eurobarometer data

6.1 RQ1: What is the trend in Internet coverage across the different European countries?

Table 1 shows the Internet coverage rate over the period 2010–2019, across Europe. At the European level, there is a 15.9 percentage point increase in the Internet coverage rate, rising from 69% in 2010 to 85% in 2019 (see the column "delta" in Table 1). This result also signals that a non-negligible share of adults (i.e. 16%) in Europe still does not have access to/do not use the Internet, in 2019.

The trend of Internet coverage varies across European countries. Indeed, there are marked cross-country differences both in Internet coverage as well as in the growth in Internet coverage rates over the period 2010–2019. For example, in 2019, Internet coverage rate varies between 99% in The Netherlands and 70% in Romania. Also, Internet coverage rate has grown by as much as 28.1 percentage points in Cyprus, while, in the same time span, it has risen by only 6.7 percentage points in The Netherlands. Despite these cross-country differences, we also notice very strong
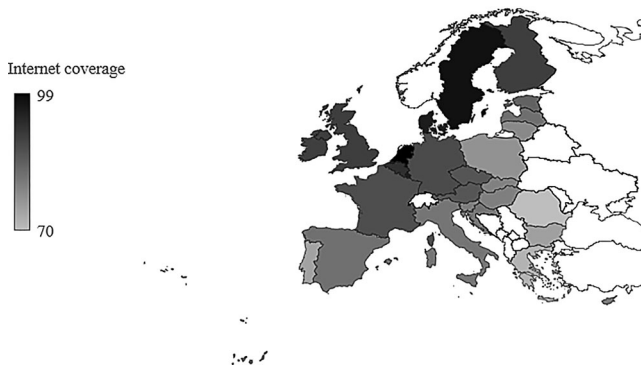
Fig. 1

Share of the population using the Internet in 2019. (Countries in white colour are excluded from the analysis, for more information see paragraph Data. Source: authors' elaboration on Eurobarometer data)

Fig. 2

Coverage rate and increase in coverage rate. (Source: authors' elaboration on Eurobarometer data)

similarities within macro-geographical areas (see Figs. 1 and 2). The South/South-Eastern European countries are those with both the lowest share of Internet coverage[3] and the highest growth in Internet coverage rate, whereas the

---

[3] As Internet use presupposes access to the Internet, respondents who do not have any source of access to the Internet are considered non-Internet users. For a description of the concept operativisation see the section *Variables*.

Nordic countries are those with both the highest share of Internet coverage and the lowest growth in Internet coverage rate. For example, The Netherlands, Finland, Denmark, and Sweden are placed on the top left corner of Fig. 2 (high coverage rate in 2010 and low increase over the period 2010–2019), while Portugal, Cyprus, Croatia, Bulgaria and Romania on the bottom right (low coverage in 2010 and high increase over the period 2010–2019). The finding that countries with higher baseline Internet coverage show a lower Internet coverage growth (and, *vice versa*, that countries with lower baseline Internet coverage experience a higher growth) is not surprising: indeed, countries with over 90% coverage at baseline (i.e. close to saturation rates at baseline) have little space for further increases in coverage, while countries with lower baseline level of coverage have more possibility for Internet coverage growth. This also signals that European countries tend to converge towards higher levels of Internet coverage.

6.2 RQ2: Are there demographic and socio-economic differences between the Internet and non-Internet population?

Fig. 3 and Tables A2–6 (see supplementary materials) show the relative coverage bias for the demographic and socio-economic variables and the substantive variables, at the European level. When focussing on the demographic and socio-economic variables, we find evidence for bias for the variables age and education; in Europe, the Internet population under-represents older people and over-represents highly educated individuals. Indeed, in 2019 the relative coverage bias for these variables corresponds to –0.086, and 0.035, respectively. However, for the variable sex, we find that bias is almost negligible (in 2019, the relative coverage bias is –0.003). When focussing on the substantive variables, we find mixed results. Similar to age and education, there are differences between the Internet and non-Internet population in the variable life satisfaction, with the Internet population over-representing more satisfied respondents (relative coverage bias: 0.014). However, if we consider the variable assessment of the economic situation, we find little evidence for bias (relative coverage bias: 0.008). Although the magnitude of the bias varies amongst the variables of interest, Fig. 3 and Tables A2–6 clearly show that, for all five variables, there is a decreasing trend in relative coverage bias. For example, when considering sex, we find a slight over-representation of men at baseline, which however decreases until becoming very small. Specifically, the relative coverage bias decreased from –0.017 to –0.003 between 2010 and 2019. However, there is also evidence suggesting that the rate of decrease is variable-specific, varying from
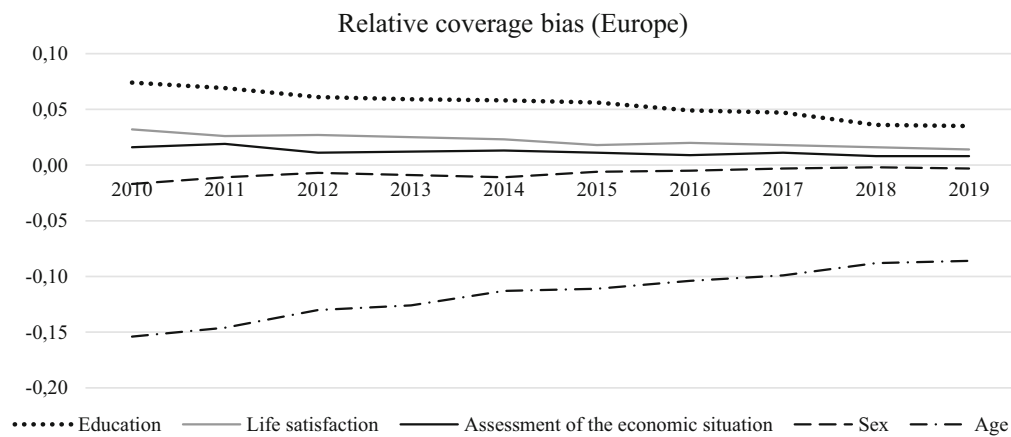
Fig. 3

Trend in relative coverage bias in Europe for selected demographic and socio-economic variables. (Source: authors' elaboration on Eurobarometer data)
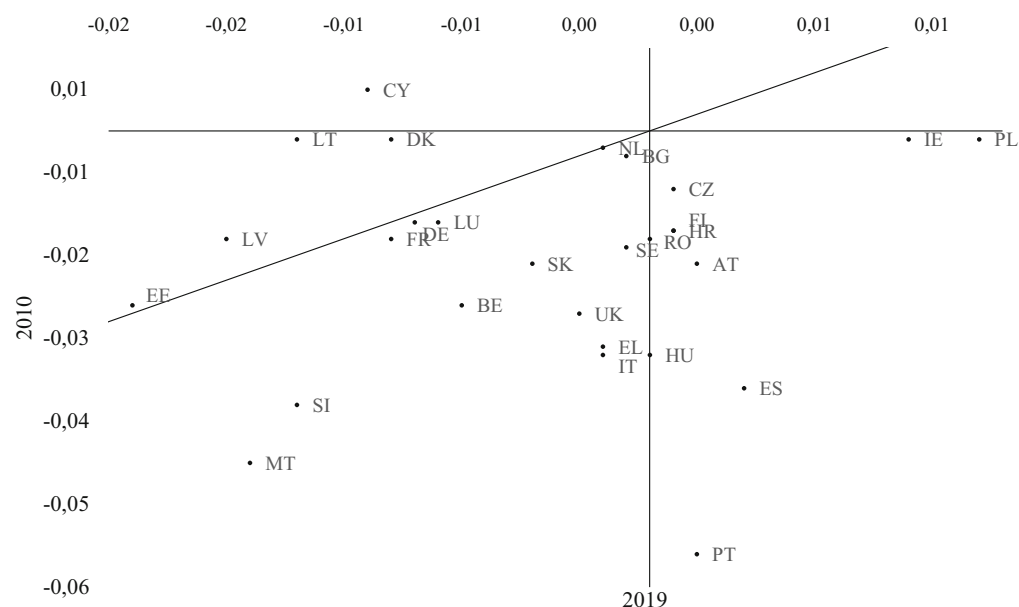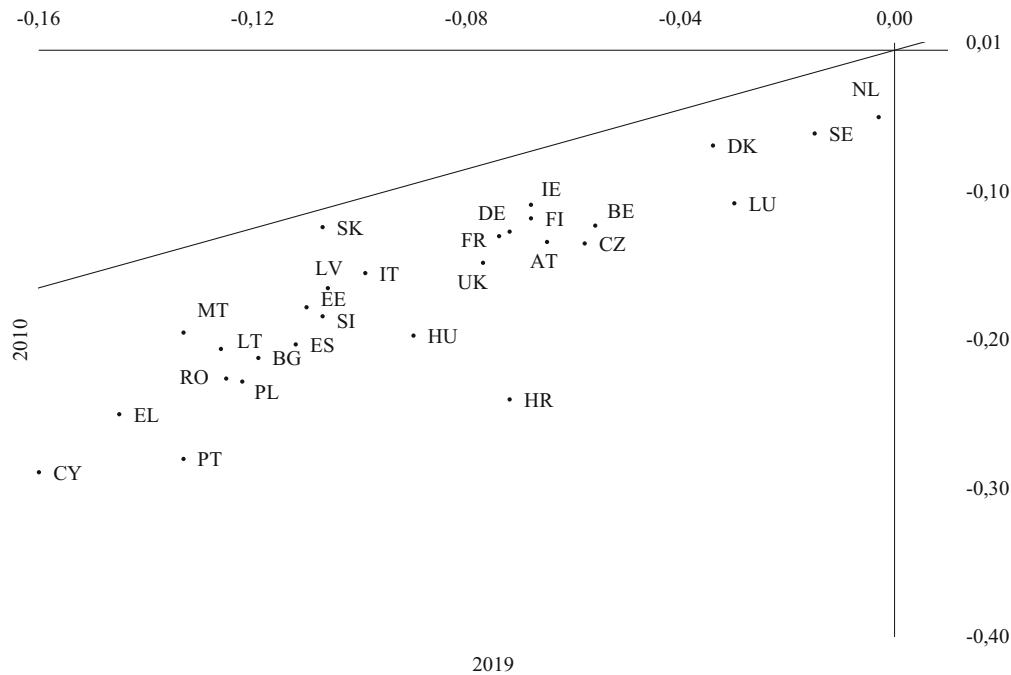

Fig. 4

Relative Coverage Bias for the variable sex, 2010–2019, European countries. (The black line represents the quadrant bisector (i.e. x = y). Source: authors' elaboration on Eurobarometer data)

0.086 for age to 0.008 for assessment of the economic situation.

To analyse the trend in relative coverage bias across the different countries, we plot, for each variable, the relative coverage bias in 2010 and 2019 (Figs. 4, 5, 6, 7, 8 and 9). We draw the quadrant bisector lines (i.e. the lines defined by the equations: y = x, i.e. bias in year *2019* = bias in year *2010,* and y = –x, i.e. bias in year *2019* = – bias in year *2010*). Countries that lie on the y = x line (i.e. bias in

year *2019* = bias in year *2010*) do not report any change in relative coverage bias between 2010 and 2019; countries that lie on the y = –x line (i.e. bias in year *2019* = – bias in year *2010*) report a change of equal size and opposite sign, between the period 2010 and 2019 (e.g. bias equals 0.05 in 2010 and equals –0.05 in 2019).

The interpretation of Figs. 4, 5, 6, 7 and 8 varies depending on which quadrant of the Cartesian plane we are observing. In the first quadrant, countries that lie above the

Fig. 5

Relative Coverage Bias for the variable age, 2010–2019, European countries. (The black line represents the quadrant bisector (i.e. x = y). Source: authors' elaboration on Eurobarometer data)

x = y line report a decrease in bias over time, and countries below the line show an increase. In the third quadrant, countries that lie above the y = x line show an increase in absolute bias (i.e. bias in 2019 is more negative than bias in 2010), while countries that lie below the line report a decrease (i.e. bias in 2019 is less negative than bias in 2010). With reference to the line y = –x (bias in year *2019* = – bias in year *2010*), in the second quadrant, countries above the line show an increase in bias—i.e. positive bias in 2019 is larger than the absolute value of the negative bias in 2010—and countries situated below the y = –x line report a decrease, i.e. the positive bias in 2019 is smaller than the absolute value of the negative bias in 2010. Conversely, in the fourth quadrant countries above the y = –x line report a decrease in bias and countries below show an increase—in this quadrant, bias is negative both in 2010 and 2019 and, above the y = –x line, bias is smaller in absolute value in 2019 (*versus* 2010), while below the line bias is larger, in absolute value, in 2019 (*versus* 2010).

For example, when looking at Fig. 6 (which depicts the first quadrant) we notice that in Portugal (PT) the differences in education levels between the Internet and non-Internet population have decreased between 2010 and 2019 whereas in Denmark they have slightly increased. Fig. 9 shows coverage bias for all countries considered in a single graph—this allows to compare bias in different variables: we notice that the trend in relative coverage bias varies across countries and that the variability seems to be more pronounced for the variables age and education. However, despite the cross-country variability in the trend, Figs. 4, 5, 6, 7, 8 and 9 also document that for all variables considered in the analysis, the relative coverage bias is decreasing or remaining stable in nearly all European countries.

6.3 RQ3: Does Internet coverage bias change over time and across countries?

Table 2 and Table 3 show the results of the five multi-level regression models that explore variation over time and across European countries in Internet coverage bias. In these models, time and time squared are the independent variables whereas absolute relative coverage bias in sex, age, education, life satisfaction, and country situation are the dependent variables. Except for the bias in the evaluation of the country situation, there is a significant decrease in absolute relative coverage bias over time for all variables considered. Indeed, the coefficients of time vary between –0.189 (for sex) to –1.054 (for age). The largest coefficient of time is observed for the variable age; this result is largely
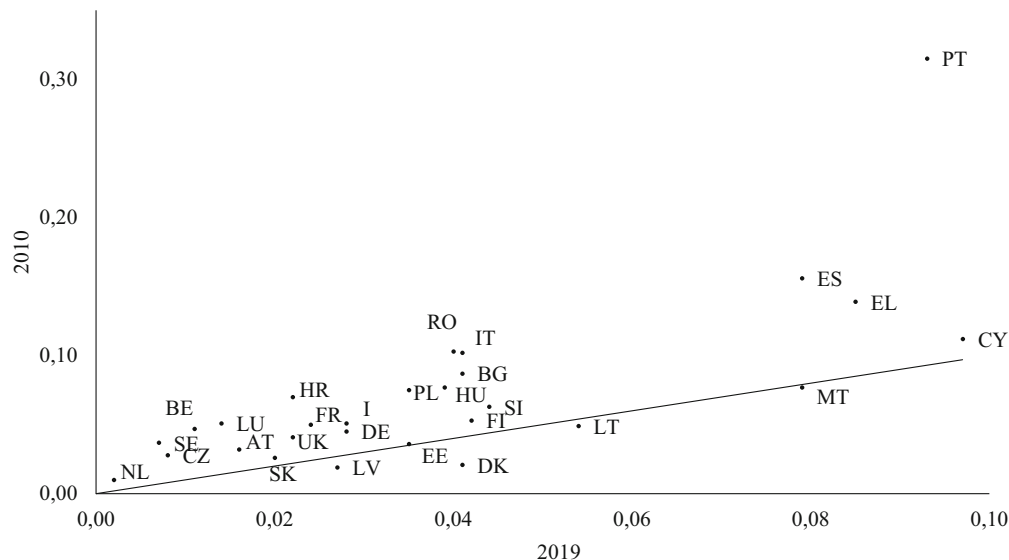
Fig. 6

Relative Coverage Bias for the variable education, 2010–2019, European countries. (The black line represents the quadrant bisector (i.e. x = y). Source: authors' elaboration on Eurobarometer data)

expected as over time cohorts which have been exposed to the Internet in youth and early adulthood become older, and, thus, the difference in Internet use across age groups shrinks.

Analysing the trend in bias (the non-linear effect of time) we find mixed results. There are only two time-squared coefficients that are significantly (and positively) associated with absolute relative coverage bias, i.e. age (0.033) and life satisfaction (i.e. 0.012, note, however, that the size of the coefficients is small). Considering that the coefficients of the variable time are negative, we conclude that for these two variables the decrease in coverage bias is reducing over time (for the other three variables, the decrease in coverage bias remains stable). We also find mixed findings in the variation of bias across countries. Only for the variables age and education the analysis shows large differences in bias between countries (country variance). The time slope variance is small, signalling that bias decreases at similar rates across countries.

6.4 RQ4: What is the role of countries' socio-economic context?

To analyse whether and how the European countries' socio-economic context affects Internet coverage bias and its changes over time, we include in the regression models

country-level characteristics and their interaction with time[4] (Models 2 in Table 2 and 3). Three important findings stand out from our analysis. First, the countries' socio-economic context plays a role in predicting coverage bias. Indeed, all five coverage bias indicators are associated (to a certain extent) to different subsets of the country-level variables. Second, there is evidence suggesting that higher levels of socio-economic development are associated with lower bias: indeed, labour force participation, Gross Enrolment Ratio, urbanicity and life expectancy are all negatively associated with coverage bias indicators. For example, labour force participation is significantly and negatively associated with bias in sex, age, life satisfaction and assessment of the country situation (coefficients are, respectively: –0.106, –0.270, –0.232, and –0.113); this result shows that countries with higher labour force participation are characterised by lower coverage bias. Also, as expected, urbanicity is significantly and negatively correlated with bias in age (b = –0.203), education (b = –0.173), and life satisfaction (b = –0.067), indicating that higher levels of urban population in a country are associated with lower bias in age, education and life satisfaction. An exception to the finding that higher levels of socio-economic development are associated with lower

---

[4] A preliminary analysis has shown that the Gini coefficient (and its interaction with time) is not associated with absolute relative coverage bias (in any of the five regressions considered) at 5% significance level; thus, this variable is omitted from all the models, following the approach adopted by Mohorko et al. (2013).
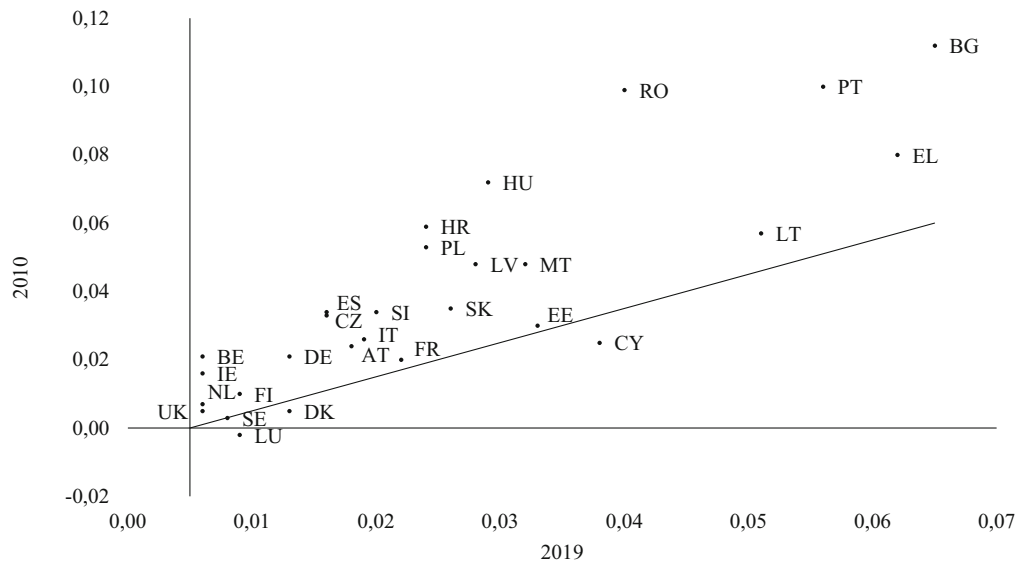
Fig. 7

Relative Coverage Bias for the variable life satisfaction, 2010–2019, European countries. (The black line represents the quadrant bisector (i.e. x = y). Source: authors' elaboration on Eurobarometer data)

bias is the positive and significant association between life expectancy and bias in sex (b = 0.168). This finding, however, is not surprising as higher levels of life expectancy are usually associated with a larger share of the older population in the country, and, given that (a) older adults are more frequently non-Internet users than younger adults and (b) women are over-represented in older age groups (due to longer female life expectancy), then countries with higher life expectancy are likely to have higher biases in the variable sex.

Third, in some cases, country-level contextual variables have a moderating effect on the relationship between time and bias. This suggests that the differences in socio-demographic context influence the change in bias over time. Indeed, coverage bias indicators of sex, age, life satisfaction and the assessment of the country situation are predicted by interactions of country-level indicators and the time indicator. For example, the interaction between labour force participation and time is positively and significantly associated with bias in sex (coeff. = 0.014), age (coeff. = 0.028), and life satisfaction (coeff. = 0.021). This means that as the labour force participation increases, the effect of time on bias also increases (or, conversely, that as time passes, the effect of labour force participation on bias increases).

7 Conclusions and Discussion

Web surveys are widely used in social science research. The popularity of this mode of data collection has further increased over the last years, when the social distancing measures associated with the Covid-19 pandemic impeded face-to-face data collection. However, lack of coverage of the non-Internet population in web surveys may still pose a serious threat to data quality; in this paper, we, examine variations in Internet coverage over time and analyse trends in Internet coverage bias across different countries. Drawing on Mohorko and colleagues (2013)' work, we analyse data from a large-scale probability-based survey—conducted using an address-based sampling and face-to-face interviewing which includes both the Internet and non-Internet population. Our study advances the knowledge in the field by providing an assessment of coverage bias across European countries over the decade 2010–2019.

A number of key findings stand out from the empirical analysis. First, despite the persistence of strong differences across countries (in Internet coverage), there is a clear increase in Internet coverage across Europe between 2010 and 2019. However, we also find that a non-negligible share of residents (i.e. 16% in 2019), especially those living in the Southern and South-eastern European countries, do not use the Internet. This suggests that in some European countries, we are far from reaching the full coverage of the general population when conducting web surveys, and this may lead to bias. Especially in Southern and Eastern Europe, most
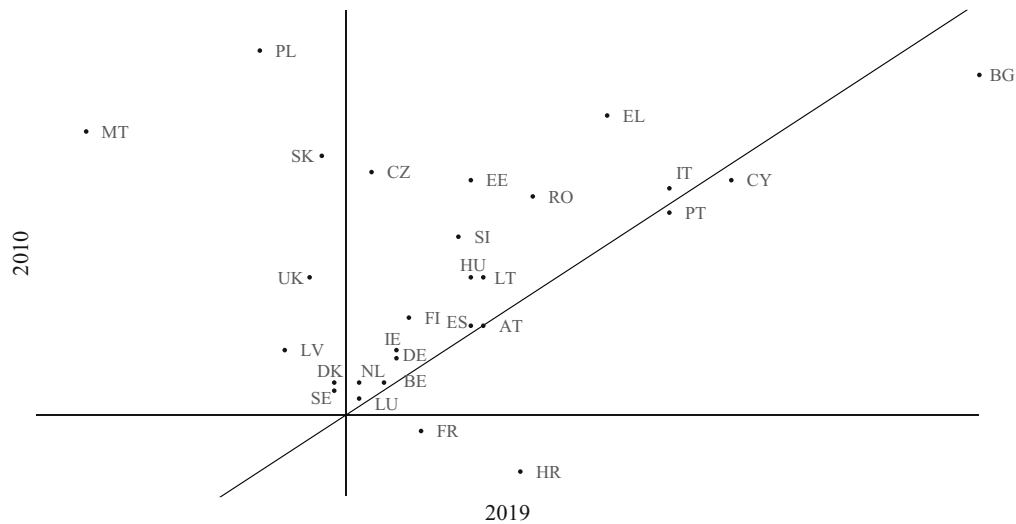
Fig. 8

Relative Coverage Bias for the variable assessment of the economic situation, 2010–2019, European countries. (The black line represents the quadrant bisector (i.e. x = y). Source: authors' elaboration on Eurobarometer data)

◆ Sex   ■ Age   × Education   ▲ Life satisfaction   ○ Assessment of the economic situation
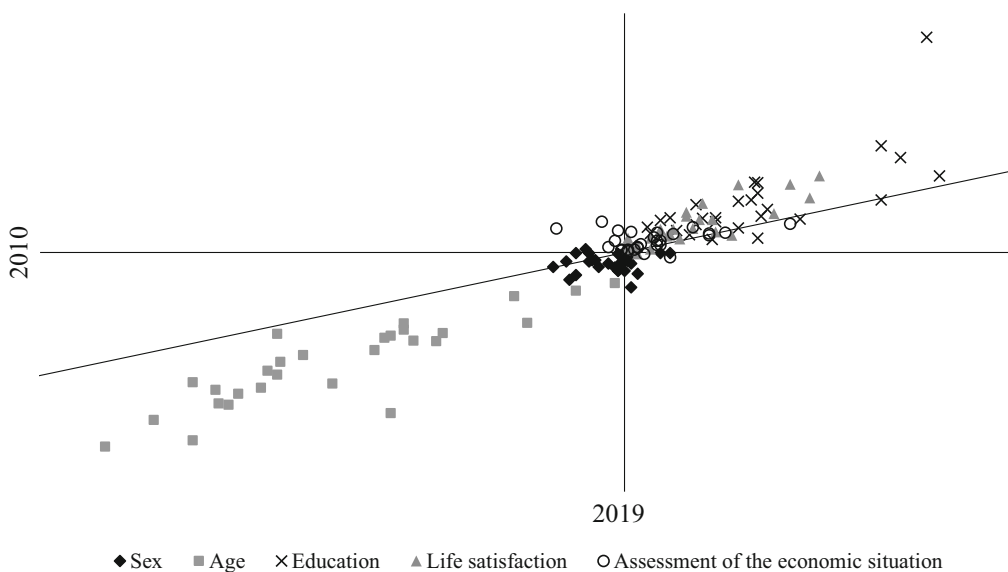
Fig. 9

Relative Coverage Bias for variables sex, age, education, life satisfaction and assessment of the Economic situation, 2010–2019, European countries. (The black line represents the quadrant bisector (i.e. x = y). Source: authors' elaboration on Eurobarometer data)

recent data from Eurostat provide little reassurance on obtaining a representative sample of the general population while excluding those without Internet access: despite an increase in Internet usage during the last years, in 2023, on average, still 6% of the European Union population (EU27) never used the Internet (Eurostat, 2024), with this share in-

creasing to double digit figures in Croatia (14%), Greece (13%), Portugal (12%), and Bulgaria (12%). This evidence (coupled with the results of our analysis on the trend of internet coverage) is particularly important for cross-national European-level online panels, such as for example the CROss-National Online Survey 2 (CRONOS-2) (Euro-

Table 2

Multilevel models for absolute relative coverage bias for selected variables predicted by year and country level variables

Fixed part	Bias Sex Composition				Bias Age Composition				Bias Education Composition			
	b	se	b	se	b	se	b	se	b	se	b	se
Time	-0.189***	0.056	ns	–	-1.054***	0.119	-6.963***	2.075	-0.356**	0.111	ns	–
Time-squared	–	–	0.017*	0.007	0.033**	0.010	ns	–	–	–	ns	–
GDP	–	–	ns	–	ns	–	ns	–	ns	–	ns	–
Inflation	–	–	ns	–	ns	–	ns	–	–	–	ns	–
Gross Enrolment Ratio	–	–	ns	–	–	–	-0.122*	0.048	–	–	ns	–
Primary education duration	–	–	ns	–	–	–	ns	–	–	–	ns	–
Secondary education duration	–	–	ns	–	–	–	ns	–	–	–	ns	–
Urban	–	–	ns	–	–	–	-0.203**	0.077	–	–	-0.173*	0.085
Labour Force Rate	–	–	-0.106***	0.031	–	–	-0.270*	0.135	–	–	ns	–
Life expectancy	–	–	0.168**	0.060	–	–	ns	–	–	–	ns	–
Time*GDP	–	–	0.016*	0.008	–	–	ns	–	–	–	ns	–
Time*Inflation	–	–	ns	–	–	–	ns	–	–	–	ns	–
Time*Gross Enrolment Ratio	–	–	ns	–	–	–	0.017**	0.006	–	–	ns	–
Time*Primary education duration	–	–	ns	–	–	–	0.312**	0.116	–	–	ns	–
Time*Secondary education duration	–	–	ns	–	–	–	0.419***	0.121	–	–	ns	–
Time*Urban	–	–	ns	–	–	–	ns	–	–	–	ns	–
Time*Labour Force Rate	–	–	0.014***	0.004	–	–	0.028*	0.012	–	–	ns	–
Time*Life expectancy	–	–	-0.026***	0.008	–	–	ns	–	–	–	ns	–
Constant	1.604***	0.187	ns	–	15.933***	1.183	95.853***	25.702	6.933***	1.071	ns	–
Residual variance	0.468	0.044	0.459	0.041	1.563	0.148	1.497	0.143	0.777	0.073	0.741	0.071
Country variance	0.690	0.228	0.375	0.144	38.244	10.365	21.709	6.739	31.657	8.532	28.245	8.507
Time slope variance	0.010	0.004	0.002	0.002	0.138	0.042	0.065	0.025	0.214	0.06	0.190	0.06

Gini coefficient, tested but removed from all models because not significant

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3

Multilevel models for absolute relative coverage bias for selected variables predicted by year and country level variables

Fixed part	Bias Life Satisfaction Composition				Bias Country Situation Composition			
	b	se	b	se	b	se	b	se
Time	–0.324***	0.068	–4.313***	0.979	ns	–	ns	–
Time-squared	0.012*	0.006	ns	–	ns	–	ns	–
GDP	–	–	ns	–	–	–	ns	–
Inflation	–	–	–0.219**	0.072	–	–	ns	–
Gross Enrolment Ratio	–	–	ns	–	–	–	–0.042*	0.020
Primary education duration	–	–	ns	–	–	–	ns	–
Secondary education duration	–	–	ns	–	–	–	ns	–
Urban	–	–	–0.067*	0.033	–	–	ns	–
Labour Force Rate	–	–	–0.232***	0.065	–	–	–0.113**	0.043
Life expectancy	–	–	–0.276*	0.137	–	–	ns	–
*Time*GDP*	–	–	ns	–	–	–	ns	–
*Time*Inflation*	–	–	0.040*	0.017	–	–	ns	–
*Time*Gross Enrolment Ratio*	–	–	ns	–	–	–	ns	–
*Time*Primary education duration*	–	–	ns	–	–	–	–0.141**	0.054
*Time*Secondary education duration*	–	–	ns	–	–	–	ns	–
*Time*Urban*	–	–	ns	–	–	–	ns	–
*Time*Labour Force Rate*	–	–	0.021***	0.006	–	–	ns	–
*Time*Life expectancy*	–	–	ns	–	–	–	ns	–
Constant	3.897***	0.590	58.859***	12.322	1.864***	0.273	21.644**	7.693
Residual variance	0.549	0.052	0.521	0.050	0.467	0.044	0.445	0.042
Country variance	9.401	2.563	3.856	1.156	1.793	0.522	1.030	0.333
Time slope variance	0.039	0.012	0.014	0.006	0.014	0.005	0.009	0.004

Gini coefficient, tested but removed from all models because not significant
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

pean Social Survey, 2024a); while the CRONOS 2 target population are only adults who have Internet access[5], the inclusion in the panel of countries with low Internet coverage (e.g. Portugal, Italy, Hungary and Czechia) prompts reflection on the extent of generalisability of results based on Internet users only to the country's population (a level of generalisability which is likely of interest for researchers, notwithstanding that the survey target population is limited to Internet users).

Furthermore, the evidence of low Internet coverage in some countries is highly relevant to surveys which incorporate (or plan to incorporate) web as one of the modes of data collection in mixed-mode designs. An example is

the European Social Survey (ESS) (European Social Survey, 2024b). At the time of writing, plans are in place for the ESS transition, by 2027, to a mixed-mode (web and postal) design. While this design would allow coverage of the non-Internet population through self-completion postal questionnaires, a crucial aspect to consider, both in the ESS and in analogous mixed-mode surveys, is the impact of mode sequence on non-response. In web-first mixed-mode designs, it is important to be mindful that offering online data collection (as a first mode of completion) in European countries with low Internet coverage may adversely impact on participation (in any mode). Indeed, non-Internet users who receive an invitation to participate in a web survey may initially refuse and remain unwilling to reconsider, even if other modes of data collection are offered later—especially if the invitation is sent by postal mail without the presence of interviewers or fieldwork staff to encourage participation. While recent research on a highly

[5] CRONOS is a probability base online panel. CRONOS 2 participants were recruited after participating in the European Social Survey Round 10 (2020–22). ESS respondents who did not have Internet access were not invited to join the online panel.

digitally savvy population—specifically, residents in Germany aged 18–49 years—offers encouraging results, showing no differences in response rates between unimode web designs and mixed-mode (web and postal) concurrent and sequential designs (Christmann et al., 2024), these findings may not hold if a significant portion of the target population has limited or no access to the Internet.

Furthermore, the evidence of a low level of Internet coverage in South and South-East Europe is important to inform survey practice in public opinion and marketing research, which are typically based on online panels, even in European countries with a relatively low Internet penetration (e.g. see Pospíšilová, 2023); in these countries, while online panels provide a rapid data turnaround at contained per-interview cost, the share of non-Internet population remains significant; to which extent this aspect is negligible depends on considerations regarding bias, which are discussed in the following paragraphs, with reference to our results on bias.

Second, in this work, we document differences between the Internet and the non-Internet population, with the former being younger and highly educated (while bias in sex seems to be negligible). In addition, we also find evidence suggesting bias in one of the substantive variables tested: life satisfaction. Indeed, respondents who are more satisfied with their life seem to be over-represented in the Internet population, while little evidence for bias emerges in the variable "assessment of the national economic situation". This signals that, on average in Europe—at least for some of the concepts considered—excluding the non-Internet population may lead to some bias in the estimates; while results are small in size, for some of the considered demographic and socio-economic variables, we observe marked variation in bias across countries; these country-level differences seem to signal that bias is driven by some European countries which have stronger differences between the Internet and the non-Internet population. This interpretation is consistent with very recent findings for specific European countries: with reference to Hungary, a simulation study finds Internet coverage bias in attitudinal, substantive and socio-demographic variables such as interest in politics, religiosity, health and marital status (Szeitl et al., 2023). Conversely, recent research indicates that in countries with higher levels of Internet penetration, such as Germany, the opposite appears to be true: minimal differences in survey estimates are observed when the non-Internet population is provided with online access (Bach et al., 2024).

Overall, due to the variability across European countries, we recommend that researchers planning a web survey—especially in Southern, Eastern, and Southeastern Europe—conduct secondary data analysis using external sources to assess coverage bias in relation to their variables of interest and target populations.

Third, we document a significant decrease in bias for most of the considered demographic and substantive variables, indicating that differences in sex, age, education, and life satisfaction between the Internet and non-Internet population are decreasing over time. The decrease in bias over time is reassuring and may suggests that, over time (as Internet coverage increases) bias may reduce further; however, the empirical evidence at hand does not allow to forecast the evolution of coverage bias in the coming years or decade. On one hand, we speculate that bias may reduce at a decreasing rate over the coming years, as Internet penetration has reached a level which is close to saturation (and further increases in Internet penetration and technology adoption may be slower). On the other hand, the recent Covid-19 pandemic has accelerated the adoption of digital technologies, reducing at a faster rate the share of the population not using the Internet. However, the implications of an increasing Internet penetration on bias are not straightforward, as argued by Eckman (2016).

Also, as the currently working-age population enters old age, Internet coverage might be higher within this group (as this cohort had access to digital technologies during their adulthood and employment). However, recent evidence suggests that older people who may have used the Internet throughout their lives can face significant challenges in maintaining their level of usage of digital technologies in old age due to disabilities and cognitive impairments brought about by the natural process of aging (Scherpenzeel & Bottenheft, 2024).

Our fourth research result indicates that bias is associated with: (i) countries' socio-economic context and (ii) the interaction between time and the countries' socio-economic context. This suggests that bias may be explained by country characteristics (e.g. urbanicity, labour force participation, and level of education) and that these moderate the effect of time on bias. Broadly speaking, countries with higher levels of economic and social development show lower levels of bias. This result suggests that if countries socio-economic development further increases, bias may decrease further (and differences across countries in bias may also reduce).

Our results are in line with the findings for the period 2005–2009 by Mohorko and colleagues (2013): similarly to the authors, we find that bias is decreasing over time, varies (at least for some variables) across countries, and is associated with higher socio-economic development of the countries. The comparability of our research with the work by Mohorko and colleagues (2013) is however limited by a different operationalization of the Internet population and a different choice (driven by data availability) of one of the substantive variables on which bias is computed. Indeed, while the authors restricted their analysis to Internet access at home, we have followed a more inclusive approach (also

adopted by Sterrett et al., 2017) and included Internet access from other locations (e.g. workplace, school, libraries, and on mobile devices). This decision was driven by two considerations: to take into account the contemporary usage of the Internet, as well as to extend our results to web surveys optimised for mobile devices such as smartphones and tablets[6]. Also, due to lack of available data, we were not able to compute bias on respondents' political left-right self-placement and instead calculated bias on the assessment of the national economic situation.

The observed changes in coverage bias over time observed in our study are somewhat comparable to the evidence from the US by Sterrett et al. (2017), though notable differences exist, likely due to the different time spans considered and geographical variations. Indeed, similarly to Sterrett et al. (2017), we find a decline in the overrepresentation of the older segments of the population and of highly educated people—these results are also in line with the findings by Mohorko et al. (2013) for the quinquennial period 2005–2009 in Europe. Additionally, in line with Mohorko et al. (2013), we find decreases in gender bias and bias in life satisfaction—trends not observed in the U.S. data, possibly because bias in these variables was already relatively low during the period analysed (2006–2014).

While we believe the results of this study have important implications, we can not derive conclusive indications on the appropriateness of using web-only data collection in a general population survey. Indeed, as also noted by Vicente and Reis (2012), the level of accuracy of web-only surveys is not universal but varies by country and outcome under analysis. However, given the persistence, in some European countries (e.g. South-East Europe), of both low coverage rate and coverage bias (in key variables that are often associated with substantive outcomes), researchers conducting pan-European surveys nowadays may consider alternatives to web-only data collection. For example, survey practitioners may equip the non-Internet population with Internet connection/devices, adopt mixed-mode survey designs, and/or implement post-collection adjustments (e.g. weighting and modelling strategies, see Dever et al., 2008; Schonlau et al., 2009). These possibilities should be evaluated taking into account survey costs and the trade-offs between the sources of survey error arising when using different protocols (e.g. measurement error due to mode effects in mixed-mode studies). The caution in adopting web-only data collection in European countries may be eased in the future if Internet coverage further increases and bias

further decreases (especially in countries which currently have low Internet coverage, e.g. South-East Europe).

It should also be noticed that the empirical evidence presented here focuses solely on coverage bias. However, adopting the Total Survey Error framework (TSE), coverage error should be considered in conjunction with other sources of survey error (e.g. sampling, non-response, specification, and measurement) and under specific budget constraints. For example, the lack of interviewer presence to motivate sample members to take part in web surveys often leads to lower response rates in web only surveys, as confirmed by a recent meta-analysis by Daikeler et al. (2022) (see also: Bosch et al., 2024; Lozar Manfreda et al., 2008; Wengrzik et al., 2016). Ultimately, mode choice depends on different trade-offs, which are out of scope for this research which focuses specifically on coverage error.

While we believe our study provides important insights regarding data quality in web surveys, it also has some limitations. First, our operativisation of the concept of Internet population is not ideal. Following a consolidated research approach, we define Internet population as individuals who use the Internet; however, as also pointed out by Fuchs and Busse (2009), among the Internet population not everyone has the digital skills and cognitive abilities to complete surveys online. Thus, our results may underestimate coverage bias, as we assume that all individuals having access and using the Internet also have the digital skills and the adequate facilities to participate in a web survey. However, a different operativisation of the Internet population was not possible due to lack of data in the Eurobarometer dataset on respondents' digital skills. Further research should define coverage bias not only with respect to Internet use, but also with reference to actual abilities to fill in a web survey.

Second, this study uses the Eurobarometer sample as a proxy for the general population: however, similarly to other surveys, also the Eurobarometer can not provide a complete representation of the general population, due to non-response or sampling error (as also noted by Sterrett et al., 2017). Non-response may influence our results: indeed, if segments of the Internet population have a different response propensity compared with the non-Internet population (e.g. female Internet users are more likely to participate in the Eurobarometer survey than female non-Internet users), then coverage bias would be confounded with non-response bias. Lack of available data on non-respondents did not allow us to perform more comprehensive analysis. Further studies could be conducted in the future, should researchers have access to data on non-respondents, to better disentangle coverage bias from non-response bias.

Third, we were unable to extend our analysis beyond 2019 due to changes in the Eurobarometer survey design. Starting in 2020, the survey was administered as a unimode

---

[6] It should be noticed that it was not possible to extend our analysis retrospectively to cover the years considered by Mohorko and colleagues (2013), due to lack of available data on Internet use outside of home, for most years in the 5-year period 2005–2009.

web survey in certain countries and no longer collected information on non-Internet users for some country-years.

Finally, we could not extend our analysis after 2019 due to a change in the Eurobarometer survey design which, from year 2020, in some country years is administered as a uni-mode web survey and does not collect information on non-Internet users, for some countries. As mentioned, however, while the Internet coverage rate has increased during the Covid-19 pandemic, still a non negligible share (7%) of the EU population does not use the Internet in 2022 (Eurostat, 2024).

In addition to the previously mentioned line of research, further studies may also expand our work by considering not only bias in selected socio-demographic and substantive variables, but also whether the inclusion/exclusion of the non-Internet population influences the results of multivariate analysis. For example, putting forward the work by Eckman (2016), existing European-level substantive studies may be replicated to assess whether research results are different depending on whether the non-Internet population is included/excluded from the analysis (i.e. whether bias leads to different results in substantive research).

Also, further research may consider not only coverage bias, but also other sources of survey error associated with web surveys (e.g. mode effects or non-response bias due to respondents having different response propensities in different survey modes). Indeed, as already proposed by some authors (e.g. Braekman et al., 2020), data quality in web surveys may be analysed by comparing estimates obtained through surveys administered adopting different modes or mixes of modes. While these works do not have the desired feature of allowing to disentangle between different sources of bias, they may provide a comprehensive overview of data quality in web surveys.

Finally, the current trend towards the use of web as a mode of data collection in official statistics and large-scale surveys complicates the analysis of Internet coverage bias—i.e. it reduces the opportunities to use general population studies to compare the Internet and the non-Internet population and hence to compute coverage bias. Non-Internet users are generally excluded from the sample (unless they are provided with the devices/training to participate) and, even when provided with adequate equipment, have lower propensity to participate than Internet users (Cornesse & Schauer, 2021).

References

Antoun, C. (2015). Who are the internet users, mobile internet users, and mobile-mostly internet users?: demographic differences across internet-use subgroups in the US. In D. Toninelli, R. Pinter & P. de Pe-draza (Eds.), *Mobile research methods: opportunities and challenges of mobile research methodologies* (pp. 99–117). London: Ubiquity Press. https://doi.org/10.5334/bar.g.

Bach, R. L., Cornesse, C., & Daikeler, J. (2024). Equipping the offline population with internet access in an online panel: Does it make a difference? *Journal of Survey Statistics and Methodology*, *12*(1), 80–93. https://doi.org/10.1093/jssam/smad003

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting up an online panel representative of the general population: the German Internet panel. *Field Methods*, *27*(4), 391–408. https://doi.org/10.1177/1525822X15574494.

Blom, A. G., Herzing, J. M. E., Cornesse, C., Sakshaug, J. W., Krieger, U., & Bossert, D. (2017). Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German Internet panel. *Social Science Computer Review*, *35*(4), 498–520. https://doi.org/10.1177/0894439316651584.

Bosch, O. J., Calderwood, L., & Gaia, A. (2024). Genpopweb2: strategies to improve response rates in probability-based online surveys: a systematic literature review. https://www.ncrm.ac.uk/documents/GenPopWeb2_LiteratureReview2.pdf

Bosnjak, M., Haas, I., Galesic, M., Kaczmirek, L., Bandilla, W., & Couper, M. P. (2013). Sample composition discrepancies in different stages of a probability-based online panel. *Field Methods*, *25*(4), 339–360. https://doi.org/10.1177/1525822X12472951.

Braekman, E., Charafeddine, R., Demarest, S., Drieskens, S., Tafforeau, J., Van der Heyden, J., & Van Hal, G. (2020). Is the European health interview survey online yet? Response and net sample composition of a web-based data collection. *European Journal of Public Health*, *30*(3), 595–601.

Brandtzæg, P. B., Heim, J., & Karahasanović, A. (2011). Understanding the new digital divide—A typology of Internet users in europe. *International Journal of Human-Computer Studies*, *69*(3), 123–138.

Burton, J., Lynn, P., & Benzeval, M. (2020). How understanding society: the UK household longitudinal study adapted to the COVID-19 pandemic. *Survey Research Methods*, *14*(2), 235–239. https://doi.org/10.18148/srm/2020.v14i2.7746.

Christmann, P., Gummer, T., Häring, A., Kunz, T., Oehrlein, A.-S., Ruland, M., & Schmid, L. (2024). Concurrent, web-first, or web-only? How different mode sequences perform in recruiting participants for a self-administered mixed-mode panel study. *Journal of Survey Statistics and Methodology*, *12*(3), 532–557. https://doi.org/10.1093/jssam/smae008

Cornesse, C., & Schaurer, I. (2021). The long-term impact of different offline population inclusion strategies in probability-based online panels: evidence from the German Internet panel and the GESIS panel. *Social Science Computer Review*, *39*(4), 687–704.

Couper, M. P. (2017). New developments in survey data collection. *Annual Review of Sociology*, *43*(1), 121–145.

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, *36*(1), 131–148.

Daikeler, J., Silber, H., & Bošnjak, M. (2022). A meta-analysis of how country-level factors affect web survey response rates. *International Journal of Market Research*, *64*(3), 306–333. https://doi.org/10.1177/147078532110509.

Dever, J. A., Rafferty, A., & Valliant, R. (2008). Internet surveys: can statistical adjustments eliminate coverage bias? *Survey Research Methods*, *2*(2), 47–62.

Dutwin, D., & Buskirk, T. D. (2023). A deeper dive into the digital divide: reducing coverage bias in Internet surveys. *Social Science Computer Review*, *41*(5), 1902–1920. https://doi.org/10.1177/08944393221093467.

Eckman, S. (2016). Does the inclusion of non-internet households in a web panel reduce coverage bias? *Social Science Computer Review*, *34*(1), 41–58.

European Commission (2010). *Eurobarometer 74.2, november-december 2010*. Milan: UniData—Bicocca Data Archive. SI295; Version 2.2.0

European Commission (2011). *Eurobarometer 76.3, november 2011*. Milan: UniData—Bicocca Data Archive. SI307; Version 1.0.0

European Commission (2012). *Eurobarometer 78.1, november 2012*. Milan: UniData—Bicocca Data Archive. SI318; Version 1.0.0

European Commission (2013). *Eurobarometer 80.1: europe 2020, economic crisis, European citizenship and media use*. Milan: UniData—Bicocca Data Archive. SI331; Version 2.0

European Commission (2014). *Eurobarometer 82.3: living conditions, trust in institutions, impact of the economic crisis, European citizenship, europe 2020*. Milan: UniData—Bicocca Data Archive. SI342; Version 4.0

European Commission (2015). *Eurobarometer 84.3: life in European Union, europe 2020, economic crisis, European citizenship, media use, and political participation*. Milan: UniData—Bicocca Data Archive. SI353; Version 4.0

European Commission (2016). *Eurobarometer 86.2: standard EU and trend questions, the europe 2020 strategy and policy priorities, the financial and economic crisis and related EU policies, European citizenship, media use and political information*. Milan: UniData—Bicocca Data Archive. SI359; Version 4.0

European Commission (2017). *Eurobarometer 88.3: attitudes towards the EU, europe 2020, European citizenship and media use*. Milan: UniData—Bicocca Data Archive. SI365; Version 1.0

European Commission (2018). *Eurobarometer 90.3: attitudes towards UE, economic crisis impact, european citizenship, use of media, UE balance*. Milan: UniData—Bicocca Data Archive. SI374; Version 1.0

European Commission (2019a). *Eurobarometer 92.2: Parlemeter 2019, europeans attitudes towards cyber security*. Milan: UniData—Bicocca Data Archive. SI387; Version 1.0

European Commission (2019b). *Eurobarometer 92.3: standard Eurobarometecrer, European citizenship, media use, EU budget, artificial intelligence and food safety*. Milan: UniData—Bicocca Data Archive. SI383; Version 1.0

European Commission (2020). *Eurobarometer 93.1: standard Eurobarometer and COVID-19 pandemic*. Milan: UniData—Bicocca Data Archive. SI386; Version 3.0

European Commission (2021). *Eurobarometer 95.3: standard Eurobarometer and COVID 19 pandemic*. Milan: UniData Bicocca Data Archive. SI396; Version 1.0

European Commission (2022). *Eurobarometer 96.3: standard Eurobarometer and COVID 19 pandemic*. Milan: UniData Bicocca Data Archive. SI400; Version 1.0

European Social Survey (2024a). CROss-national Online survey 2 (CRONOS-2) panel. https://www.europeansocialsurvey.org/methodology/methodological-research/modes-data-collection/cronos. Accessed 12 Aug 2024.

European Social Survey (2024b). The road to a sustainable self-completion future. European social survey ERIC strategic plan 2024–2029. https://www.europeansocialsurvey.org/sites/default/files/2024-04/strategic-plan-2024-29.pdf

European Union (2021). Eurobarometer. Public opinion in the European Union. https://europa.eu/eurobarometer/screen/home. Accessed 8 Aug 2024.

European Union (2023a). Standard Eurobarometer 93. Summer 2020. Public opinion in the European Union. First results. https://europa.eu/eurobarometer/surveys/detail/2262. Accessed 7 July 2023.

European Union (2023b). Standard Eurobarometer 94. Winter 2020–2021. Public opinion in the European Union. First results. https://europa.eu/eurobarometer/surveys/detail/2355. Accessed 7 July 2023.

European Union (2023c). Standard Eurobarometer 96. Winter 2021–2022. Public opinion in the European Union. First results. https://europa.eu/eurobarometer/surveys/detail/2553. Accessed 7 July 2023.

Eurostat (2021). Database. https://ec.europa.eu/eurostat/en/web/main/data/database. Accessed 17 Sept 2021.

Eurostat (2024). Individuals—internet use (ISOC_CI_IFP_IU, I_IUX). https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ifp_iu/default/table?lang=en. Accessed 8 Aug 2024.

Fuchs, M., & Busse, B. (2009). The coverage bias of mobile web surveys across European countries. *International Journal of Internet Science*, *4*(1), 21–33.

GESIS (2021a). Eurobarometer data service. https://www.gesis.org/en/eurobarometer-data-service. Accessed 11 Nov 2024.

GESIS (2021b). Sampling and fieldwork. https://www.gesis.org/en/eurobarometer-data-service/data-and-documentation/standard-special-eb/sampling-and-fieldwork. Accessed 5 Oct 2021.

GESIS (2022). Standard and special topic Eurobarometer. https://www.gesis.org/en/eurobarometer-data-service/data-and-documentation/standard-special-eb. Accessed 11 Aug 2022.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Groves, R. M., Fowler, F. J. Jr, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd edn.). Hoboken: John Wiley & Sons, Inc..

Hoffman, D. L., Novak, T. P., & Schlosser, A. (2000). The evolution of the digital divide: how gaps in Internet access may impact electronic commerce. *Journal of Computer-Mediated Communication*. https://doi.org/10.1111/j.1083-6101.2000.tb00341.x.

Kocar, S., & Biddle, N. (2023). Do we have to mix modes in probability-based online panel research to obtain more accurate results? *Methods, Data, Analyses*, *17*(1), 93–120. https://doi.org/10.12758/mda.2022.11.

Leenheer, J., & Scherpenzeel, A. C. (2013). Does it pay off to include non-internet households in an internet panel? *International Journal of Internet Science*, *8*(1), 17–29.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. Chichester: Wiley.

Lipps, O., & Pekari, N. (2016). Sample representation and substantive outcomes using web with and without incentives compared to telephone in an election survey. *Journal of Official Statistics*, *32*(1), 165–186.

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: a meta-analysis comparing response rates. *International Journal of Market Research*, *50*(1), 79–104.

Mesch, G. S., & Talmud, I. (2011). Ethnic differences in Internet access: the role of occupation and exposure. *Information, Communication & Society*, *14*(4), 445–471. https://doi.org/10.1080/1369118X.2011.562218.

Mohorko, A., de Leeuw, E., & Hox, J. (2013). Internet coverage and coverage bias in europe: developments across countries and over time. *Journal of Official Statistics*, *29*(4), 609–622. https://doi.org/10.2478/jos-2013-0042.

Pospíšilová, J. (2023). Challenges in building the first probability-based online panel in the Czech Republic—CRONOS-2 experience. 10th Conference of the European Survey Research Association, Milan, Italy (2023, July 17–21). https://www.europeansurveyresearch.org/conf2023/progGlance.php?sess=22#main.

Robinson, J. P., Neustadtl, A., & Kestnbaum, M. (2002). The online "diversity divide": public opinion differences among internet users and nonusers. *IT & Society*, *1*(1), 284–302.

Sala, E., Gaia, A., & Cerati, G. (2022). The gray digital divide in social networking site use in Europe: Results from a quantitative study. *Social Science Computer Review*, *40*(2), 328–345.

Salkind, N. J. (2010). Last observation carried forward. In N. J. Salkind (Ed.), *Encyclopedia of research design* Vol. 1. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412961288.

Scherpenzeel, A., & Bottenheft, E. (2024). *Online panel participation in an ageing society: insights from the Dutch national panel for the chronically ill and disabled (NPCD)*. Handout paper for the 2024 Panel Survey Methods Workshop, Utrecht, The Netherlands

Schnell, R., Noack, M., & Torregoza, S. (2017). Differences in general health of internet users and nonusers and implications for the use of web surveys. *Survey Research Methods*, *11*(2), 105–123.

Schonlau, M., Van Soest, A., Kapteyn, A., & Couper, M. (2009). Selection bias in web surveys and the use of propensity scores. *Sociological Methods & Research*, *37*(3), 291–318.

Stern, M. J., & Dillman, D. A. (2006). Community participation, social ties, and use of the internet. *City & Community*, *5*(4), 409–424.

Sterrett, D., Malato, D., Benz, J., Tompson, T., & English, N. (2017). Assessing changes in coverage bias of web surveys in the United States. *Public Opinion Quarterly*, *81*(S1), 338–356.

Szeitl, B., Fellner, Z., & Tátrai, A. (2023). From minor data gaps to major errors—simulation study to demonstrate potential bias of online surveys. 10th Conference of the European Survey Research Association, Milan, Italy (2023, July 17–21). https://www.europeansurveyresearch.org/conf2023/progGlance.php?sess=141#main.

Toepoel, V., & Hendriks, Y. (2016). The impact of non-coverage in web surveys in a country with high internet penetration: is it (still) useful to provide equipment to non-internet households in the Netherlands? *International Journal of Internet Science*, *11*(1), 33–50.

Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The science of web surveys*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199747047.001.0001.

Valliant, R., & Lee, S. (2005). Economic characteristics of Internet and non–Internet users and implications for web–based surveys. https://www.researchgate.net/publication/259497193_Economic_Characteristics_of_Internet_and_Non-Internet_Users_and_Implications_for_Web-based_Surveys

Vicente, P., & Reis, E. (2012). Coverage error in internet surveys: can fixed phones fix it? *International Journal of Market Research*, *54*(3), 323–345.

Wengrzik, J., Bosnjak, M., & Lozar Manfreda, K. (2016). *Web surveys versus other survey modes—A meta-analysis comparing response rates*. Paper presented at the General Online Research conference, Dresden, Germany (2016, March 2–4).

World Bank (2021). DataBank. https://databank.worldbank.org/home. Accessed 17 Sept 2021.

World Bank (2023). Individuals using the Internet (% of population) international telecommunication union (ITU) world telecommunication/ICT indicators database. https://data.worldbank.org/indicator/IT.NET.USER.ZS. Accessed 7 July 2023.

Zhang, C., Callegaro, M., Thomas, M., & Di Sogra, C. (2009). Do we hear different voices? Investigating the differences between Internet and non-Internet users on attitudes and behaviors. In *JSM Proceedings*. Alexandria: American Statistical Association.