



MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2021-006 | April 2021
<https://doi.org/10.4054/MPIDR-WP-2021-006>

How Reliable is Facebook's Advertising Data for Use in Social Science Research? Insights from a Cross-National Online Survey

André Grow | grow@demogr.mpg.de

Daniela Perrotta

Emanuele Del Fava

Jorge Cimentada

Francesco Rampazzo

Sofia Gil-Clavel

Emilio Zagheni

René D. Flores

Ilana Ventura

Ingmar Weber

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

How Reliable is Facebook's Advertising Data for Use in Social Science Research?

Insights from a Cross-National Online Survey

André Grow¹, Daniela Perrotta¹, Emanuele Del Fava¹, Jorge Cimentada¹, Francesco Rampazzo², Sofia Gil-Clavel¹, Emilio Zagheni¹, René D. Flores³, Ilana Ventura³, Ingmar Weber⁴

Abstract

Social scientists increasingly use Facebook's advertising platform for research, either in the form of conducting digital censuses of the general population, or for recruiting participants for survey research. Both approaches depend on the reliability of the data that Facebook provides about its users, but little is known about how reliable these data are. We address this gap in a large-scale, cross-national online survey ($N = 137,224$), in which we compare self-reported and Facebook-classified demographic information (sex, age, and region of residence). Our results suggest that Facebook's advertising platform can be fruitfully used for conducting social science research if additional steps are taken to assess the reliability of the characteristics under consideration.

Keywords

Digital censuses, Facebook, online surveys, targeted advertising

✉ André Grow
Max Planck Institute for Demographic Research
Konrad-Zuse-Str. 1, 18057 Rostock, Germany
grow@demogr.mpg.de

¹ Max Planck Institute for Demographic Research

² University of Oxford

³ University of Chicago

⁴ Qatar Computing Research Institute

1 Introduction

Facebook's advertising platform provides aggregated information about the characteristics of the network's users (e.g., gender, age, and interests) and advertisers can use this information for targeted advertising. An increasing number of scholars advocate for the use of these facilities in social science research, either for conducting digital censuses that aim to measure characteristics of the general population, and for recruiting participants for survey research (e.g., Alburez-Gutierrez *et al.*, 2019; Alexander *et al.*, 2019; Cesare *et al.*, 2018; Pötzschke and Braun, 2017; Ribeiro *et al.*, 2020; Rosenzweig *et al.*, 2020; Zagheni *et al.*, 2017). One reason is that traditional probability-based sampling methods, such as address-based sampling and random digit dialing, have proven increasingly costly and inadequate in response-rates and coverage for many segments of the general population (Stern *et al.*, 2014). Against this backdrop, social media and big data—and Facebook's advertising platform in particular—are an attractive supplement for traditional survey research methods. They offer a potentially less expensive and more timely alternative (Amaya *et al.*, 2020) and make it possible to generate samples of geographic or demographic subpopulations that would otherwise be difficult to reach (Zhang *et al.*, 2020).

The feasibility of using Facebook for conducting digital censuses and generating samples of specific subpopulations depends on the reliability of the data that underlies its advertising platform. Systematic misclassification of individual traits like gender and age could significantly bias scholarly research. However, Facebook does not offer much information on how reliable the information about its userbase is, or how it determines characteristics and interests that are partially or completely inferred from user behavior on the network. In this paper, we address this issue by comparing individuals' self-reported information in an online survey, where respondents are recruited using the Facebook advertising platform, with the way Facebook classified the same people for the purposes of targeted advertising. While information collected via surveys has its own limitations, our study sheds light on the extent to which data from Facebook's advertising platform, often considered a 'black box' (Araujo *et al.*, 2017), can be trusted for research, as well as the extent to which the targeting features can be leveraged.

Our assessment is based on a large-scale, cross-national online survey. The survey was conducted in seven European countries (Belgium, France, Germany, Italy, the Netherlands, Spain, and the United Kingdom) and in the United States, with the goal to collect information about people's behaviors and attitudes in response to the COVID-19 pandemic. Recruitment took place daily via targeted Facebook advertising campaigns that were stratified by users' sex, age, and subnational region of residence (such as the "West" of the United States, as defined

by the U.S. Census Bureau). In the questionnaire, respondents were asked to report these characteristics themselves. By comparing their answers with information about the specific ads through which respondents arrived at the survey, it becomes possible to indirectly assess Facebook users' classification. The closer the match between Facebook's categorization and participants' answers, the more reliably Facebook's advertising platform can be used for research purposes. Given that sex, age, and region of residence are commonly used stratification variables in social science research, and are known to relate to a large range of attitudes, behaviors, and demographic outcomes (Geary, 2020; Lutz *et al.*, 1998; Ribeiro *et al.*, 2020), our work is relevant for many researchers who seek to use Facebook for social science research.

We are not the first to assess the reliability of Facebook's advertising data (see, e.g., Pötzschke and Braun, 2017; Rosenzweig *et al.*, 2020; Sances, 2021). However, our study goes beyond earlier work on this topic, by (1) taking a cross-national perspective, (2) assessing classification mismatches across the entire Facebook user population in the respective countries, and (3) assessing the directionality of mismatches (e.g., did those incorrectly classified as 25–44 years old report to be younger or older?). In what follows, we first describe Facebook's advertising platform and discuss how it has been used in earlier research. We then describe the survey and how participants were recruited. After this, we discuss our sample and analytical approach to assess the reliability of Facebook users' classification, and present results. Data was collected between March 13 and August 12, 2020, resulting in a total of $N = 137,224$ questionnaires with complete information on respondents' sex, age, and region of residence. We close with an outlook and recommendations for future research.

To preview results, we find that across countries, for most respondents (99%) the survey answers matched with Facebook's categorization on at least two out of the three characteristics that we considered. At the level of individual characteristics, the reliability of Facebook's categorization was highest for sex (between 98%–99% matches) and lowest for region of residence (between 91%–98% matches). Based on these findings, we suggest that Facebook's advertising platform can be fruitfully used for conducting social science research, if additional steps are taken to assess the reliability of the specific user characteristics that are in the focus of a given study.

2 Facebook's Advertising Platform and its Use in Earlier Research

Facebook is the largest social media platform, with 2.45 billion monthly active users worldwide, as of fall 2019 (Facebook Inc., 2019). Its business model centers on revenue from

online advertising (Zagheni *et al.*, 2017), which is technically implemented through the Facebook Ads Manager (FAM). The FAM allows advertisers to create ad campaigns that can have various goals, such as creating salience for a given service or product among Facebook users, or generating traffic to an external website. Each advertising campaign can target specific user groups, which can be defined based on several self-reported demographic and personal characteristics (e.g., gender and age), and a set of characteristics that Facebook infers from the users' behavior on the network (e.g., political orientation). Campaigns have three levels. At the highest level, the goals of the campaign are defined (e.g., generating awareness or generating traffic). The second level is the ad set level, at which the target audience, budget, and ad delivery schedule are defined. The third level includes the advertisements themselves, which can consist of multiple visual materials (e.g., images, videos), multiple texts, and the URL of a web page to which users should be directed when they click on the ad. Prior to launching a campaign, the FAM provides an estimate of the expected audience size (i.e., the number of daily or monthly active users who are eligible to be shown an ad) given the selected combination of user characteristics. This allows advertisers to optimize their definition of target groups (Cesare *et al.*, 2018).

Earlier social science research has used the FAM mostly in one of two ways. A first set of studies have employed the audience estimates that the FAM provides prior to launching a campaign for obtaining digital censuses of the user population across geographic regions. The resulting information was then used to make inferences about specific social groups and the general population (e.g., Alexander *et al.*, 2019; Kashyap *et al.*, 2020; Rama *et al.*, 2020; Rampazzo *et al.*, 2018; Ribeiro *et al.*, 2020; Zagheni *et al.*, 2017). For example, Zagheni *et al.* (2017) used audience estimates to assess the share of foreign-born people living in the United States, comparing these numbers with data from the 2014 round of the American Community Survey (ACS). Their results showed that the Facebook audience estimates were qualitatively similar to the number of migrants observed in the ACS, which suggests that the FAM data can be used to study compositional population properties. One benefit of this approach is that the information that the FAM provides is updated continuously and can be collected programmatically through Facebook's application programming interface (API). This makes it possible to collect population data in a more continuous and more timely manner than is possible with traditional censuses or register data (Ribeiro *et al.*, 2020).

A second set of studies have used the targeted advertising facilities that the FAM offers to recruit participants for survey research (e.g., Guillory *et al.*, 2018; Kühne and Zindel, 2020; Pötzschke and Braun, 2017; Rinken *et al.*, 2020; Rosenzweig *et al.*, 2020; Sances, 2021;

Schneider and Harknett, 2019; Zhang *et al.*, 2020). With this approach, researchers define one or more Facebook user groups whose members could be shown an ad that invites them to participate in an online survey. This ad will then be displayed, e.g., in the users' timelines, and when they click on it, they are directed to an external webpage where they can participate in the survey. Pötzschke and Braun (2017) used this approach for recruiting Polish migrants in four European countries (Austria, Ireland, Switzerland, and the United Kingdom) for a survey that queried them about their use of social networking sites, their migration experience, and their socioeconomic background. Given Facebook's reach, this approach is particularly attractive when the goal is to recruit members of sub-populations that account only for a small share of the overall population and that are difficult to identify in existing sampling plans (such as migrants or workers in specific industries). More recently, Zhang *et al.* (2020) have shown that targeted advertisements can also be used to collect representative samples of the general population, if the target groups in the advertising campaign are sufficiently fine-grained.

Some of the studies that have used the FAM for survey recruitment have assessed the reliability of the advertising information in reaching the targeted demographic groups. Pötzschke and Braun (2017), for example, reported for their survey of Polish migrants that about 98% of those who arrived at the survey via a Facebook ad lived in one of the countries that were targeted. Ultimately, about 96% of the participants were Polish migrants, lived in one of the four targeted countries, and were at least 18 years old. By contrast, focusing on two countries in the global South (Mexico and Kenya), Rosenzweig *et al.* (2020) reported more variation in the observed matches. While they reported a nearly 100% agreement between respondents' gender and Facebook's advertising data in Mexico, they only found about 13% matches for educational attainment in Kenya. Similarly, drawing on six studies in the US, Sances (2021) reported that almost 100% of respondents who were classified by Facebook as 25 years and older also reported to be older than 24, whereas only about 23% of those who were classified as Black reported to be Black.

While insightful, these earlier studies have in common that they applied their recruitment criteria either to a single country or focused on a small subset of the larger population (circumscribed by demographic and social characteristics, and/or by place of residence). In this paper, we add to this body of literature by taking cross-national perspective in which we assess classification mismatches across the entire Facebook user population in the targeted countries and assessing the directionality of mismatches. This provides additional insights into which users are more likely to correctly or incorrectly classified. For example, our approach enables us to explore whether members of certain age groups are more likely to be misclassified than

members of other age groups, and to explore the age groups to which they are incorrectly assigned.

3 Data and Methods

3.1 Survey and Facebook advertising campaigns

This study uses data from the COVID-19 Health Behavior Survey (CHBS) (Del Fava *et al.*, 2020; Grow *et al.*, 2020; Perrotta *et al.*, 2020). The CHBS is an anonymous, cross-national online survey that was conducted in Belgium, France, Germany, Italy, the Netherlands, Spain, the United Kingdom, and the United States. Participation was voluntary and not incentivized. Data collection began on March 13, 2020 in Italy, the United Kingdom, and the United States. Subsequent countries were added continuously, with Belgium joining last, on April 4, 2020. The data collection ended in all countries on August 12, 2020. The questionnaire had four sections, encompassing questions about respondents' socio-demographic characteristics, health indicators, behaviors and attitudes related to COVID-19, and social contacts. Our focus here is exclusively on respondents' demographic characteristics.

Participant recruitment occurred via targeted Facebook advertising campaigns. The CHBS ran one ad campaign per country with the goal to generate traffic to the survey's webpage (there was one separate webpage per country). Facebook's ad delivery algorithms aimed to optimize ad delivery to increase the likelihood that users who were shown an ad clicked on it. Each campaign was stratified at the ad-set level by users' gender (man or woman), age group (18–24, 25–44, 45–64, and 65+ years), and region of residence (see details in the next subsection), resulting in 24 to 56 strata per country. This stratification approach ensured a balance in central demographic characteristics of the resulting respondent samples, to which post-stratification techniques could be applied to improve representativeness (Grow *et al.*, 2020). Figure 1 illustrates the structure of the campaigns for the United States, including an example of the ads that were used.

–Figure 1 about here–

3.2 Inferring Facebook users' classification

Given the stratified nature of the advertising campaigns, we could infer how Facebook had classified the sex, age, and region of residence of users from the ad through which they arrived at the survey. For example, a participant who arrived at the survey via an ad that targeted 25–44 years old men in the western United States should have reported a matching age, sex, and

region of residence in the survey. If his survey answers deviate from this, this might point to an error in Facebook's user classification, but it might also stem from reporting errors on the side of survey participants, either in the questionnaires or on Facebook. Accordingly, we interpret any difference between participants' answers and their classification by Facebook as bias, regardless of the exact cause of such differences. Note that Facebook users may see and click on ads that are not targeted at them. This can happen, e.g., when a Facebook friend of a non-targeted user comments on an ad, which then may appear as organic content in the non-targeted user's timeline. We do not consider participants who arrived at the survey in this way in our analysis.

The FAM allows advertisers to select from two *genders*, 'men' and 'women', which is based on user self-reported information (Facebook Inc., 2020), and which we used for stratifying the advertising campaigns. By contrast, in the CHBS questionnaire, respondents were asked to report their *sex*, with the options 'male' and 'female'. Prior scholarship has shown that biological sex and gender are not necessarily equivalent (West and Zimmerman, 1987; Westbrook and Saperstein, 2015). Hence, it is an open question whether these terms would overlap in this specific context.

Facebook usage is restricted to individuals age 13 years and older, and advertisers can use single-year age categories to define their target, up to the age of 64 years. Older users are aggregated in the category 65+. In the CHBS questionnaire, respondents were asked to report their age in years, which makes it possible to map their answers onto the four age categories used for stratifying the advertisements. Note that participation in the CHBS was restricted to individuals of age 18 and older, which is the lower age boundary in the advertising campaigns and in the survey data. Facebook employs users' self-reported age in its categorization (cf. Facebook Inc., 2020; United States Securities and Exchange Commission, 2019).

Facebook offers several means for geographic targeting. For example, advertisers can draw on pre-defined regions, such as the state of California in the United States, or advertisers can define their own regions by selecting a geographic point of reference (defined by its latitude and longitude) together with a radius around this point (in miles). User locations are estimated based on several pieces of information, such as information from mobile devices, IP address, and self-reported information (United States Securities and Exchange Commission, 2019). The CHBS advertising campaigns divided each targeted country into three to seven subnational regions (here also called macro regions), which were composed of smaller micro regions. The micro regions were based on pre-defined regions offered by Facebook, largely following the NUTS-1 classification in Europe and the census regions in the United States (see Table 1 for

an overview). The region-related answer categories in the CHBS questionnaire were largely identical to the micro regions that were used in the advertising campaigns.

–Table 1 about here–

The only notable exceptions from the region classification approach described above occurred in the United Kingdom and Spain. To minimize the possibility that the large metropolitan area of London dominated the daily recruitment efforts in England, two separate groups of ad sets were created. The first group focused on England while excluding London, whereas the second group only focused on London. This was achieved by defining a custom region centered on London with a radius of 17 miles, that was selectively included in or excluded from the ad sets. In the case of Spain, the cities of Ceuta and Melilla in northern Africa were not included in the targeting. The reason is that targeting these cities by defining a radius around a geographic reference point would have led to the inclusion of parts of the surrounding African countries, which were not in the focus of the CHBS. These Spanish cities were therefore not included in the ad targeting, but respondents could select them from the set of answers in the CHBS questionnaire.

3.3 Sample selection

Data was collected between March 13 and August 12, 2020. Over this period, 144,034 individuals completed the CHBS questionnaire, but we only considered the subset of respondents who arrived at the survey's page by clicking on an ad that was targeted at them and who reported their sex, age, and region of residence in the survey. For consistency, in the Spanish data we also excluded respondents who reported to live in the cities Ceuta and Melilla in northern Africa, as these areas were not part of the ad targeting (<1% of the sample for Spain). The final sample consisted of 137,224 individuals (95% of the original sample; about 1% of the original sample reported to live in a country that was not in focus of the respective advertising campaign). Tables 2 and 3 show the distribution of respondents across countries, sex, age, and regions. Compared to their respective national populations, female and older individuals were over-represented in the survey. As discussed in Grow et al. (2020), this bias can be addressed with post-stratification weighting to make the data more representative of the respective national populations (see also Perrotta *et al.*, 2020), but in the analysis reported here, we use unweighted data, as we are not aiming to make statistical inferences about national populations.

–Tables 2 and 3 about here–

3.4 Analytical approach

We used standard classification-evaluation metrics to assess the reliability of Facebook’s user classification, namely classification *accuracy*, *precision*, *recall*, and the F_1 score (Tharwat, 2020). All four measures are calculated based on a so-called confusion matrix, that cross-tabulates the actual category to which an object belongs (in our case respondents’ self-reported sex, age, and region of residence) and the class to which it has been assigned by a prediction model (in our case Facebook’s user classification). Table 4 provides an example of such a confusion matrix, assuming that there is one characteristic with three categories. Each cell reports the number of respondents (n_{ij}) who were observed for each combination of actual (i) and predicted category (j). Cells along the main diagonal (n_{11} , n_{22} , and n_{33}) report the numbers of respondents who were correctly classified, whereas all other cells represent incorrect classifications. Note that there is one such matrix for each characteristic and country.

–Table 4 about here–

Given this matrix, *accuracy* is defined as the fraction of respondents who were categorized correctly. This measure is calculated as

$$accuracy = \frac{n_{11} + n_{22} + n_{33}}{\sum_i \sum_j n_{ij}}. \quad (1)$$

Hence, the larger *accuracy*, the more likely that, for a given demographic characteristic, the answer of a randomly selected respondent matches with Facebook’s user classification.

The *accuracy* measure provides a general assessment of the overall quality of the classification, but it has two shortcomings. First, it does not consider that the distribution of correct and incorrect classifications may differ between different categories of the same characteristic (e.g., in the case of sex, there might be more correct classifications for male than for female respondents). Second, if the number of observations across categories are imbalanced, the results tend to be biased towards the dominant category (e.g., if there were more male than female respondents in the sample, the correct and incorrect classifications of male respondents may dominate the results) (Chawla, 2010). The measures *precision* and *recall* address these issues by looking at each category separately. In more detail, *precision* is calculated as the fraction of the predictions for a given category i that were correct. This measure is calculated as

$$precision_i = \frac{n_{ij, i=j}}{\sum_j n_{ij}}. \quad (2)$$

By contrast, *recall* is the fraction of actual instances of category i that were predicted correctly.

It is calculated for a given class i as

$$recall_i = \frac{n_{ij,i=j}}{\sum_j n_{ij}}. \quad (3)$$

Hence, *precision* indicates how many of the observations that were predicted to belong to category i actually belonged to category i (e.g., how many of the individuals who were predicted to be male actually reported to be male?), whereas *recall* indicates how many of the observations that actually belonged to category i were correctly predicted to belong to this category (e.g., how many of the individuals who reported to be male were correctly predicted to be male?).

The measures *precision* and *recall* assess different aspects of the confusion matrix, but they do not provide an overall assessment of the classification per category. The F_1 score (also simply F_1 from here on) provides such an assessment, and is calculated as the harmonic mean of *precision* and *recall* for a given category i as

$$F_{1,i} = 2 \frac{precision_i \times recall_i}{precision_i + recall_i}. \quad (3)$$

Hence, $F_{1,i}$ will be close to one when both $precision_i$ and $recall_i$ are close to one, but $F_{1,i}$ will be lower when $precision_i$ and/or $recall_i$ are lower. In the discussion of our results, we focus on F_1 as a summary measure, and refer to *precision* and *recall* if there are marked differences between them for a given characteristic.

4 Results

4.1 Overall reliability

Table 5 reports the shares of respondents who were classified correctly on zero, one, two, or three characteristics. Across countries, between 86% and 93% of respondents were correctly classified on all three characteristics, meaning that the sex, age, and region of residence that Facebook used for targeting respondents in the advertising campaigns matched with their answers in the CHBS questionnaire. The share of completely correct classifications was lowest in Belgium and France, and highest in the Netherlands. Among those respondents who did not have a perfect match on all three characteristics, typically only one characteristic was incorrect, and very few respondents had only one or no matching characteristics (<2%).

–Table 5 about here–

Table 6 assesses for which characteristics misclassifications were most likely to occur by reporting the respective *accuracy* values across countries. As the table shows, classification

accuracy was highest for sex, ranging from .980 in France and the Netherlands to .987 in Italy and the United States. This means that between 98–99% of all classifications were correct. For age, classification accuracy was somewhat lower, ranging from .925 in France to .963 in the Netherlands. Classification accuracy was lowest for region of residence, and there was somewhat more variation across countries, with values ranging from .909 in Belgium to .981 in the United States.

–Table 6 about here–

4.2 Reliability of sex categories

Table 7 shows *precision*, *recall*, and F_1 measures for male and female respondents across countries. Generally, the combined measure of F_1 was high for both male and female respondents, but it was consistently higher for female than for male respondents by a margin of about .008 to .014 points across countries. At the same time, *precision* was typically higher among female respondents, whereas *recall* was higher among male respondents. For example, in France, *precision* was .994 for female and .950 for male respondents, whereas *recall* was .977 for female and .986 for male respondents. Hence, classification of respondents as women by Facebook were more likely to match with respondents' answers on their sex than classifications as male (*precision*), whereas those who reported to be male were more likely to be classified correctly than those who reported to be female (*recall*). However, while consistent across countries, these differences were relatively small.

–Table 7 about here–

4.3 Reliability of age categories

Compared to sex, we found more variability by country in the match between Facebook's classification of age and respondents' answers. As Table 8 shows, the overall classification quality (as indicated by F_1) was highest for the age category 25–44 years (average F_1 = .958 across countries), and lowest for the age categories 18–24 years and 65+ years (average F_1 = .925 and F_1 = .929, respectively). We observed the lowest single value of F_1 for the category 18–24 years in the United Kingdom (F_1 = .855), and the highest value for the category 25–44 years in Belgium (F_1 = .971). Furthermore, there were systematic differences in *precision* and *recall* across the age groups. Those who were classified as 18–24 years had a comparatively low likelihood to report to belong this age group (average *precision* = .871 across countries), whereas they were more likely to be correctly classified as such when they reported to be 18–

24 years old (average *recall* = .987 across countries). The opposite was the case for the age category 45–65 years (average *precision* = .977 and *recall* = .917, respectively). For example, in the United Kingdom, only about 75% of those who were classified as 18–24 years old reported an age in this range, whereas about 99% of those who reported to be 18–24 years old were also classified as such. By contrast, about 98% of those who were classified as 45–64 years old reported an age in this range, whereas only about 92% of those who reported to be 45–64 years old were also classified as such. For the other age groups, the differences between *precision* and *recall* were less strong and less systematic across countries.

–Table 8 about here–

The fact that respondents reported their age in single years allows us to assess the directionality of misclassifications. Figure 2 shows respondents’ reported age and the age category to which Facebook assigned respondents (data pooled from all countries). The red vertical lines demarcate the boundaries of the different age groups. Congruent with the fact that *precision* was typically above 90% across countries and age groups, the mass of the age distributions fell within the boundaries of the respective age groups to which respondents had been assigned by Facebook. Yet, within these age groups, there was a marked skew towards the upper boundary, except for the oldest age group (65+ years), which had a skew towards the lower boundary.

–Figure 2 about here–

4.4 Reliability of region of residence

Focusing next on regions, Table 9 shows the *precision*, *recall*, and F_1 measures for each of the different regions across countries. Overall, the classification quality was high, with somewhat less variation across categories than was the case for age. The F_1 score mostly varied between .925 (for the “England” region within the United Kingdom) and .993 (for the “Northern Ireland” region also in the U.K.). The only outliers were the regions of Brussels in Belgium and London in the U.K., with F_1 scores of .787 and .791, respectively. Furthermore, also the values of *precision* and *recall* were generally high and did not differ systematically across countries. This indicates that across countries, individuals who were classified as living in a given region by Facebook often also reported living in the same region (*precision*). At the same time, most respondents who reported living in a given region were also correctly classified by Facebook (*recall*). Again, the only marked exceptions were Brussels and London, for which *precision* tended to be lower than *recall* (.654 vs. .987 for Brussels and .662 vs. .983 for London,

respectively). Hence a large share of respondents who were classified as living in Brussels or London by Facebook reported in the survey to live in other regions (*precision*), whereas those who reported to live in Brussels or London were usually correctly classified by Facebook (*recall*).

–Table 9 about here–

The fact that respondents reported detailed regions of residence (micro regions) enables us to assess the directionality of misclassifications. In Figures 3 and 4, we look closer at the misclassifications that occurred in Belgium and the United Kingdom, respectively. Focusing first on Belgium, Figure 4 shows that the low precision for the region of Brussels was largely due to respondents who reported living in the Flemish Brabant and Walloon Brabant regions, but who were classified by Facebook as living in the nearby region of Brussels. Correspondingly, also the *recall* values for Flanders and Wallonia in Table 9 were somewhat lower than the corresponding *precision* values. Focusing next on the United Kingdom, Figure 5 shows that most misclassifications for the region of London concerned respondents who reported living in the East and the South East of England, which are the two regions that geographically surround London. Notably, a substantive share of respondents who reported living in South West England were misclassified as living in the adjacent region of Wales.

–Figures 3 and 4 about here–

5 Discussion and Conclusion

In this paper, we examined whether the information that Facebook’s Ads Manager (FAM) provides about its user database can be used reliably in social science research. We compared the sex, age, and region of residence that participants of an anonymous online survey reported with the way Facebook had classified the same individuals on these characteristics in its advertising algorithms. We relied on the COVID-19 Health Behavior Survey (CHBS), which recruited its participants via targeted ads on Facebook in eight countries. Our results showed that there was a very good, albeit imperfect, match between respondents’ self-reported characteristics and Facebook’s classification. Across countries, about 86%–93% of respondents’ answers matched Facebook’s categorization on all three characteristics that we considered. Misclassifications were most likely to occur for region of residence and least likely to occur for sex.

Why was the error rate for region of residence higher than for sex and age? One possible explanation is that Facebook’s gender and age classifications are largely based on self-reported

information that is not very likely to change over time (i.e., after signing up on Facebook). By contrast, users' region of residence is partially inferred by Facebook and may change frequently, thereby increasing the chance for erroneous classifications. Interestingly, most of the incorrect region classifications concerned people who reported living in regions that were adjacent to those to which they were incorrectly assigned by Facebook. These latter regions were also comparatively small. In more detail, the largest share of misclassifications concerned respondents who Facebook had classified as living in Brussels and London, but who reported living in the surrounding areas. Both cities are among the most important economic hubs in their respective countries, while living space within the cities' borders is scarce. It seems likely that daily commuting for work from the surrounding suburbs may have contributed to the large number of classification errors that we observed. This result parallels the findings of Sances (2021), who reported that in the US classifications were more likely to be correct in larger regions than in smaller regions. In the case of London in our study, this trend may have been aggravated by the fact that the targeting was based on geographic radius around the center of London, rather than its exact borders as was the case for the other regions that we considered. Given that the actual shape of London is more complex than a simple circle, this approach may have inadvertently included Facebook users who lived close to the border of London, but not in London itself.

With regard to age, we observed distinct misclassification patterns across age groups, as well as distinct participation patterns within age groups. In terms of misclassifications, those who were classified as 18–24 years old were least likely to report an age in this interval, whereas those classified as 45–64 years old were most likely to report an age in this interval. Conversely, those who reported to be 18–24 years old were most likely to be correctly classified by Facebook, whereas those who reported to be 45–64 years old were least likely to be correctly classified. Assuming that respondents' survey answers were truthful, this points to the possibility that among 45–64-year-old Facebook users, there is a substantive share who have misreported their age when registering on the social network. Alternatively, if respondents correctly indicated their age on Facebook, it seems possible that many younger survey participants may have reported to be older than they actually are. Facebook has acknowledged that information on age among younger users may be less accurate (United States Securities and Exchange Commission, 2019), but with our data, we cannot adjudicate between these sources of bias.

Next to the observed mismatches in terms of age, we observed notable skews in the age distributions within age groups. In the younger age groups, there was a skew towards the upper

age boundary, whereas in the oldest age group (65+ years) there was a skew towards the lower age boundary. In the younger groups, these patterns may result from at least two interacting processes. First, the CHBS is a health-related survey and the ads showed health related content (see Figure 1 for an example; see Grow et al. (2020) for all images used in the campaigns). Older adults tend to be more interested in health-topics than younger individuals (Pew Research Center, 2015) and COVID-19 tends to have more negative health outcomes for older individuals (Nikolich-Zugich *et al.*, 2020). Both factors may have increased the interest of older Facebook users in the survey. Hence, within each stratum of the CHBS advertising campaigns, older Facebook users may have been more likely to click on the ads and participate in the survey, thereby leading to a skew in the age distribution within the different strata. Second, Facebook's advertising algorithms are designed to maximize the likelihood that users who are shown an ad click on it. If older users were more likely to engage with the CHBS ads, Facebook's advertising algorithms may have reinforced the resulting skew by preferentially targeting older users. Unfortunately, we cannot determine whether the observed age patterns in our data are (at least partially) the result of Facebook's advertising algorithms. However, our results underscore the importance of stratifying advertising campaigns on important demographic characteristics, such as age, if the goal is to obtain representative samples of the population of Facebook users. The pattern observed in the oldest age group may result from the fact that there are relatively few very old individuals on Facebook (cf. Gil-Clavel and Zagheni, 2019). Hence, the skew towards the lower age boundary in the age group 65+ years may simply reflect the age structure in this segment of the Facebook user population.

When it comes to sex, some of the observed misclassifications may stem from the fact that Facebook offers users' gender rather than biological sex for targeting ads. Hence, some of the mismatches that we observed may be due to users whose gender identity differs from their biological sex, or who have non-binary gender identities. This may reduce the likelihood that Facebook assigns them to a gender category that aligns with their biological sex. We cannot directly assess this potential source of bias, but our results show that even though Facebook's user categorization is based on gender, this information can be used reliably to recruit respondents of a specific sex. Note that trans-gender, gender fluid, or non-binary respondents may have opted for the category "prefer not to answer" when asked for their sex. In this case, they would not be included in the analyses presented in this paper.

Our assessment of Facebook's advertising data improves on earlier work on this topic by taking a cross-national perspective, by studying the entire demographic spectrum of Facebook's user base, and by exploring in detail the directionality of observed mismatches. Yet, there are

also some caveats that should be kept in mind. First, our work is not a direct assessment of the accuracy of Facebook's user classification algorithms. Mismatches between Facebook's classification and participants' self-reported characteristics may stem from a genuine misclassification on Facebook's side, but respondents may also have misreported (either on purpose or by accident) their characteristics in the survey. Conversely, users may have misreported their characteristics on Facebook. Hence, our study provides information about how likely misclassifications are to occur, and which characteristics are particularly affected by it, but it does not provide insights into the definite causes of these misclassifications.

Furthermore, our results apply to Facebook users who are actively using the social network, who are willing to participate in online surveys, and who have an interest in health-related topics. Additionally, the highly educated were somewhat over-represented in our sample (cf. Perrotta *et al.*, 2020), which is congruent with the observation that more educated people generally are more likely to participate in survey research (Spitzer, 2020). These aspects may be problematic for several reasons. For example, the accuracy of Facebook's classification may be lower among people who are less active on the platform, if those who use Facebook less frequently are also less likely to keep their profile information up to date. Furthermore, people who are less inclined to participate in surveys may generally be more concerned about their privacy, and this may be associated with less accurate reporting of personal characteristics to Facebook. As these individuals were less likely to take part in the CHBS, we may have inadvertently overestimated the accuracy of Facebook's advertising data.

These caveats notwithstanding, our work has practical implications for scholars who want to use Facebook's Ads Manager in social science research, especially for those who want to recruit participants for survey research. Our results suggest that the FAM is a valuable and largely reliable tool for research, given that Facebook's user categorization matched the self-reported central demographic characteristics reported in our survey. At the same time, there were some mismatches, and their number varied between countries and between the different categories of the characteristics that we considered. We therefore suggest that scholars who want to use the FAM conduct pre-test surveys among their targeted Facebook sub-population to assess the reliability of the user information that Facebook provides. For example, if the goal is to study Turkish immigrants in Germany, researchers could target this group via Facebook ads and invite them to participate in a short demographic survey, in which their country of birth and immigration status are queried. The observed pattern of matches and mismatches could be used to plan the recruitment efforts and budget accordingly. If the goal is to conduct a digital census, information on matches and mismatches could be used to assess the uncertainty that

surrounds the resulting population parameter estimates.

Our work also has implications for the advancement of big data social science research at large. Big data are increasingly seen as an attractive supplement for survey research, as they offer a potentially “less expensive, less burdensome, and more timely alternative for producing a variety of statistics” (Amaya *et al.*, 2020: 90). At the same time, the use of big data comes with its own methodological challenges. One challenge is the assessment of bias in big data, and an increasing number of scholars are calling for a systematic assessment of such bias (e.g., Amaya *et al.*, 2020; Baker, 2017; Schober *et al.*, 2016; Sen *et al.*, 2019). Most notably, Amaya *et al.* (2020) recently suggested assessing big data in a way similar to the Total Survey Error (TSE) framework, calling this new approach the Total Error Framework (TEF). The TSE has been established to quantify bias in survey research, encompassing all research steps from defining the inferential population to questionnaire design and drawing inferences. The TEF applies this approach to big data, considering error sources that may occur, e.g., during data identification and extraction. The approach that we have presented here can contribute to both, the TSE and the TEF. In terms of the TSE, the FAM has been likened to sampling frames that are often used in survey research; like other sampling frames, FAM too suffers from systematic under-coverage of certain segments of the population (e.g., those who do not have a Facebook account for), and its own biases. Our findings, as well as the approach proposed in this paper, contributes to the assessment of sampling error using FAM that arises when there is discrepancy between Facebook’s user classification and respondents’ actual characteristics. Similarly, in the case of the TEF, our approach provides insights into the biases that may emerge when researchers use the FAM for conducting digital censuses, as described above. These findings also open the door to further studies in multi-mode and multiple sample-frame survey research, and the possibility to target different segments of the population through different sample-frames and modes according to population coverage in each, including across national borders.

6 References

- Alburez-Gutierrez, D., Aref, S., Gil-Clavel, S., et al. (2019) Demography in the digital era: New data sources for population research. In: *Book of Short Papers SIS2019* (eds G Arbia, S Peluso, A Pinna, et al.), 2019, pp. 22–33. Pearson. DOI: DOI:10.31235/osf.io/24jp7.
- Alexander, M., Polimis, K. and Zagheni, E. (2019) The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook Data. *Population and Development Review*, **45**, 617–630. DOI: <https://doi.org/10.1111/padr.12289>.

- Amaya, A., Biemer, P. P. and Kinyon, D. (2020) Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology*, **8**, 89–119. Oxford Academic. DOI: 10.1093/jssam/smz056.
- Araujo, M., Mejova, Y., Weber, I., et al. (2017) Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. In: *Proceedings of the 2017 ACM on Web Science Conference*, Troy New York USA, 25 June 2017, pp. 253–257. ACM. DOI: 10.1145/3091478.3091513.
- Baker, R. (2017) Big data: A survey research perspective. In *Proceedings of Statistics Canada Symposium 2016: Growth in Statistical Information: Challenges and Benefits*, pp. 47–69. DOI: 10.1002/9781119041702.ch3.
- Cesare, N., Lee, H., McCormick, T., et al. (2018) Promises and pitfalls of using digital traces for demographic research. *Demography*, **55**, 1979–1999. DOI: 10.1007/s13524-018-0715-2.
- Chawla, N. V. (2010) Data mining for imbalanced datasets: An overview. In *Data Mining and Knowledge Discovery Handbook* (eds O. Maimon and L. Rokach), pp. 875–886. Boston, MA: Springer US. DOI: 10.1007/978-0-387-09823-4_45.
- Del Fava, E., Cimentada, J., Zagheni, E., et al. (2020) The differential impact of physical distancing strategies on social contacts relevant for the spread of COVID-19. *medRxiv*. DOI: <https://doi.org/10.1101/2020.05.15.20102657>.
- Facebook Inc. (2019) Facebook reports third quarter 2019 results. Available at: <https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-Third-Quarter-2019-Results/default.aspx> (accessed 15 April 2020).
- Facebook Inc. (2020) Age and gender. Available at: <https://business.facebook.com/business/help/717368264947302?id=176276233019487>.
- Geary, D. C. (2020) *Male, Female: The Evolution of Human Sex Differences*. 3rd ed. American Psychological Association.
- Gil-Clavel, S. and Zagheni, E. (2019) Demographic differentials in Facebook usage around the world. *Proceedings of the International AAAI Conference on Web and Social Media*, **13**, 647–650.
- Grow, A., Perrotta, D., Del Fava, E., et al. (2020) Addressing public health emergencies via Facebook surveys: Advantages, challenges, and practical considerations. *Journal of Medical Internet Research*, **22**, e20653. DOI: 10.2196/20653.

- Guillory, J., Wiant, K. F., Farrelly, M., et al. (2018) Recruiting hard-to-reach populations for survey research: Using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. *Journal of Medical Internet Research*, **20**, e197. DOI: 10.2196/jmir.9461.
- Kashyap, R., Fatehkia, M., Tamime, R. A., et al. (2020) Monitoring global digital gender inequality using the online populations of Facebook and Google. *Demographic Research*, **43**, 779–816.
- Kühne, S. and Zindel, Z. (2020) Using Facebook and Instagram to recruit web survey participants: A step-by-step guide and application. *Survey Methods: Insights from the Field*. DOI: 10.13094/SMIF-2020-00017.
- Lutz, W., Goujon, A. and Doblhammer-Reiter, G. (1998) Demographic dimensions in forecasting: Adding education to age and sex. *Population and Development Review*, **Supplement: Frontiers of Population Forecasting**, 42–58.
- Nikolich-Zugich, J., Knox, K. S., Rios, C. T., et al. (2020) SARS-CoV-2 and COVID-19 in older adults: What we may expect regarding pathogenesis, immune responses, and outcomes. *GeroScience*, **42**, 505–514. DOI: 10.1007/s11357-020-00186-0.
- Perrotta, D., Grow, A., Rampazzo, F., et al. (2020) Behaviors and attitudes in response to the COVID-19 pandemic: Insights from a cross-national Facebook survey. *medRxiv*. DOI: <https://doi.org/10.1101/2020.05.09.20096388>.
- Pew Research Center (2015) Public interest in science and health linked to gender, age and personality. *Numbers, Facts and Trends Shaping the World*. DOI: 202.419.4372.
- Pötzschke, S. and Braun, M. (2017) Migrant sampling using Facebook advertisements: A case study of Polish migrants in four European countries. *Social Science Computer Review*, **35**, 633–653.
- Rama, D., Mejova, Y., Tizzoni, M., et al. (2020) Facebook ads as a demographic tool to measure the urban-rural divide. In *Proceedings of The Web Conference 2020*, pp. 327–338. Available at: <https://dl.acm.org/doi/abs/10.1145/3366423.3380118> (accessed 24 February 2021).
- Rampazzo, F., Zagheni, E., Weber, I., et al. (2018) Mater certa est, pater numquam: What can Facebook advertising data tell us about male fertility rates? In *Proceedings of the International AAAI Conference on Web and Social Media*.
- Ribeiro, F. N., Benevenuto, F. and Zagheni, E. (2020) How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. In *WebSci '20: 12th ACM Conference on Web Science*.

- Rinken, S., Domínguez-Álvarez, J.-A., Trujillo, M., et al. (2020) Combined mobile-phone and social-media sampling for web survey on social effects of COVID-19 in Spain. *Survey Research Methods*, **14**, 165–170. 2. DOI: 10.18148/srm/2020.v14i2.7733.
- Rosenzweig, L., Bergquist, P., Pham, K. H., et al. (2020) Survey sampling in the Global South using Facebook advertisements. SocArXiv. DOI: 10.31235/osf.io/dka8f.
- Sances, M. W. (2021) Missing the target? Using surveys to validate social media ad targeting. *Political Science Research and Methods*, **9**, 215–222. Cambridge University Press. DOI: 10.1017/psrm.2018.68.
- Schneider, D. and Harknett, K. (2019) Consequences of routine work-schedule instability for worker health and well-being. *American Sociological Review*, **84**, 82–114. DOI: <https://doi.org/10.1177/0003122418823184>.
- Schober, M. F., Pasek, J., Guggenheim, L., et al. (2016) Social media analyses for social measurement. *Public Opinion Quarterly*, **80**, 180–211. DOI: 10.1093/poq/nfv048.
- Sen, I., Floeck, F., Weller, K., et al. (2019) A total error framework for digital traces of humans. *arXiv:1907.08228 [cs.CY]*. Available at: <http://arxiv.org/abs/1907.08228>.
- Spitzer, S. (2020) Biases in health expectancies due to educational differences in survey participation of older Europeans: It's worth weighting for. *The European Journal of Health Economics*, **21**, 573–605. DOI: 10.1007/s10198-019-01152-0.
- Stern, M. J., Bilgen, I. and Dillman, D. A. (2014) The state of survey methodology: Challenges, dilemmas, and new frontiers in the era of the tailored design. *Field Methods*, **26**, 284–301. DOI: 10.1177/1525822X13519561.
- Tharwat, A. (2020) Classification assessment methods. *Applied Computing and Informatics*, **Online first**. DOI: 10.1016/j.aci.2018.08.003.
- United States Securities and Exchange Commission (2019) Annual report on form 10-K, Facebook Inc. Available at: <https://www.sec.gov/Archives/edgar/data/1326801/000132680119000009/fb-12312018x10k.htm>.
- West, C. and Zimmerman, D. H. (1987) Doing gender. *Gender & Society*, **1**, 125–151. DOI: 10.1177/0891243287001002002.
- Westbrook, L. and Saperstein, A. (2015) New categories Are not enough: Rethinking the measurement of sex and gender in social surveys. *Gender & Society*, **29**, 534–560.
- Zagheni, E., Weber, I. and Gummadi, K. (2017) Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, **43**, 721–734.

Zhang, B., Mildenberger, M., Howe, P. D., et al. (2020) Quota sampling using Facebook advertisements. *Political Science Research and Methods*, **8**, 558–564. Cambridge University Press. DOI: 10.1017/psrm.2018.49.

Tables and Figures

Country	Macro Region	Micro Region
Belgium	Brussels	Brussels
	Flanders	Antwerp, East Flanders, Flemish Brabant, Limburg, West Flanders
	Wallonia	Hainaut, Liege, Luxembourg, Namur, Walloon Brabant
France	Île de France	Île de France
	Northeast	Alsace Champagne-Ardenne Lorraine, Bourgogne - Franche-Comté, Nord-Pas-de-Calais Picardie
	Southeast	Auvergne-Rhône-Alpes, Corse, Provence-Alpes-Côte-d'Azur
	Southwest	Aquitaine Limousin Poitou-Charentes, Languedoc-Roussillon Midi-Pyrénées
	West	Bretagne, Centre Val de Loire, Normandie, Pays de la Loire
Germany	East	Berlin, Brandenburg, Sachsen, Sachsen-Anhalt, Thüringen
	North	Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Schleswig-Holstein
	South	Baden-Württemberg, Bayern
	West	Hessen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland
Italy	Central	Lazio, Marche, Toscana, Umbria
	Insular	Sardegna, Sicilia
	Northeast	Emilia-Romagna, Friuli-Venezia Giulia, Provincia Autonoma di Bolzano/Bozen, Provincia Autonoma di Trento, Veneto
	Northwest	Liguria, Lombardia, Piemonte, Valle d'Aosta/Valleé d'Aoste
	South	Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia
Netherlands	East	Flevoland, Gelderland, Overijssel
	North	Drenthe, Friesland, Groningen
	South	Limburg, Noord-Brabant
	West	Noord-Holland, Utrecht, Zeeland, Zuid-Holland
Spain	Islas Canarias	Islas Canarias
	Center	Castilla - La Mancha, Castilla y León, Extremadura
	East	Cataluña, Comunidad Valenciana, Islas Balears
	Madrid	Madrid
	Northeast	Aragón, Navarra, La Rioja, Pais Vasco
	Northwest	Cantabria, Galicia, Principado de Asturias
	South	Andalucia, Region de Murcia

Continues on next page

Continued

Country	Macro Region	Micro Region
United Kingdom	Other	Ciudad Autónoma de Ceuta, Ciudad Autónoma de Melilla
	England	East Midlands, East of England, North East, North West, South East, South West, West Midlands, Yorkshire and The Humber
	London	London
	Northern Ireland	Northern Ireland
	Scotland	Scotland
United States	Wales	Wales
	Midwest	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
	Northeast	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
	South	Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia, Washington DC
	West	Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming

Table 1. Countries, macro regions, and micro regions in the CHBS/Facebook advertising campaigns

Country	N	Sex (%)		Age (%)			
		Female	Male	18-24	25-44	45-64	65+
Belgium	12,657	65	35	14	29	36	21
France	13,430	69	31	16	29	35	20
Germany	25,707	59	41	17	37	32	15
Italy	15,651	67	33	16	39	31	14
Netherlands	11,280	64	36	11	22	40	27
Spain	13,345	69	31	6	35	43	16
United Kingdom	14,216	65	35	7	21	42	30
United States	30,938	63	37	8	24	36	32
Total	137,224	64	36	12	30	36	22

Table 2. Number of respondents per country and their distribution across sex and age as self-reported in the CHBS questionnaire

Country	Region	%
Belgium	Brussels	16
	Flanders	41
	Wallonia	43
France	Ile de France	15
	North East	22
	South East	22
	South West	20
	West	21
Germany	East	23
	North	23
	South	26
	West	28
Italy	Central	21
	Insular	9
	Northeast	25
	Northwest	33
	South	12
Netherlands	East	24
	North	20
	South	26
	West	31
Spain	Canarias	7
	Center	12
	East	24
	Madrid	17
	North East	11
	North West	13
	South	16
United Kingdom	England (w.o. London)	49
	London	10
	Northern Ireland	7
	Scotland	19
	Wales	14
United States	Midwest	27
	Northeast	22
	South	24
	West	26

Table 3. Distribution of respondents across regions within countries as self-reported in the CHBS questionnaire

		Predicted		
		Cat. 1	Cat. 2	Cat.3
Actual	Cat. 1	n_{11}	n_{12}	n_{13}
	Cat. 2	n_{21}	n_{22}	n_{23}
	Cat. 3	n_{31}	n_{32}	n_{33}

Table 4. Example of confusion matrix

Country	Correct characteristics (%)			
	0	1	2	3
Belgium	<1	1	13	86
France	<1	1	12	86
Germany	<1	1	8	91
Italy	<1	1	7	92
Netherlands	<1	1	6	93
Spain	<1	1	9	90
United Kingdom	<1	1	13	87
United States	<1	1	7	92

Table 5. Share of respondents for which zero, one, two, or three of their reported characteristics (sex, age, and region of residence) matched with Facebook's classification. Cells show row percentages.

Country	Accuracy		
	Sex	Age	Region
Belgium	.982	.959	.909
France	.980	.925	.944
Germany	.984	.948	.970
Italy	.987	.951	.972
Netherlands	.980	.963	.984
Spain	.985	.934	.972
United Kingdom	.986	.941	.929
United States	.987	.942	.981

Table 6. *Accuracy* for sex, age, and region by country

Country	Precision		Recall		F ₁	
	Female	Male	Female	Male	Female	Male
Belgium	.994	.960	.978	.989	.986	.975
France	.994	.950	.977	.986	.985	.968
Germany	.992	.972	.980	.989	.986	.980
Italy	.995	.972	.986	.991	.991	.981
Netherlands	.993	.958	.976	.988	.984	.973
Spain	.989	.974	.988	.977	.989	.975
United Kingdom	.996	.968	.982	.994	.989	.981
United States	.994	.974	.985	.990	.989	.982

Table 7. *Precision, recall, and F₁* for the different categories of sex by country

Country	Precision				Recall				F_1			
	18-24	25-44	45-64	65+	18-24	25-44	45-64	65+	18-24	25-44	45-64	65+
Belgium	.923	.960	.978	.952	.994	.982	.937	.945	.957	.971	.957	.949
France	.863	.931	.968	.906	.988	.966	.882	.890	.921	.948	.923	.898
Germany	.931	.962	.976	.881	.991	.961	.908	.950	.960	.962	.941	.914
Italy	.929	.957	.982	.900	.981	.973	.917	.932	.954	.965	.949	.916
Netherlands	.893	.946	.987	.974	.995	.981	.944	.964	.941	.963	.965	.969
Spain	.832	.941	.973	.873	.974	.973	.903	.918	.897	.956	.936	.895
United Kingdom	.753	.922	.981	.957	.987	.972	.922	.937	.855	.947	.951	.947
United States	.846	.916	.968	.964	.988	.981	.924	.922	.911	.948	.945	.943

Table 8. *Precision, recall, and F_1* for the different categories of age by country

Country	Region	Precision	Recall	F ₁
Belgium	Brussels	.654	.987	.787
	Flanders	.992	.900	.944
	Wallonia	.991	.888	.936
Germany	East	.946	.988	.966
	North	.975	.968	.971
	South	.980	.968	.974
	West	.977	.958	.968
Spain	Canarias	.985	.987	.986
	Centre	.958	.937	.948
	East	.989	.974	.981
	Madrid	.950	.974	.962
	North East	.960	.982	.971
	North West	.981	.978	.979
	South	.980	.977	.978
France	Ile de France	.913	.946	.929
	North East	.966	.936	.951
	South East	.954	.946	.950
	South West	.950	.958	.954
	West	.931	.938	.934
Italy	Central	.973	.977	.975
	Insular	.957	.984	.970
	Northeast	.975	.968	.972
	Northwest	.977	.973	.975
	South	.964	.961	.963
Netherlands	East	.976	.984	.980
	North	.985	.989	.987
	South	.986	.988	.987
	West	.988	.978	.983
United Kingdom	England	.993	.866	.925
	London	.662	.983	.791
	Northern Ireland	.992	.993	.993
	Scotland	.988	.993	.990
	Wales	.910	.988	.947
United States	Midwest	.984	.983	.983
	Northeast	.978	.988	.983
	South	.978	.968	.973
	West	.985	.987	.986

Table 9. *Precision, recall, and F1* for the different categories of region by country

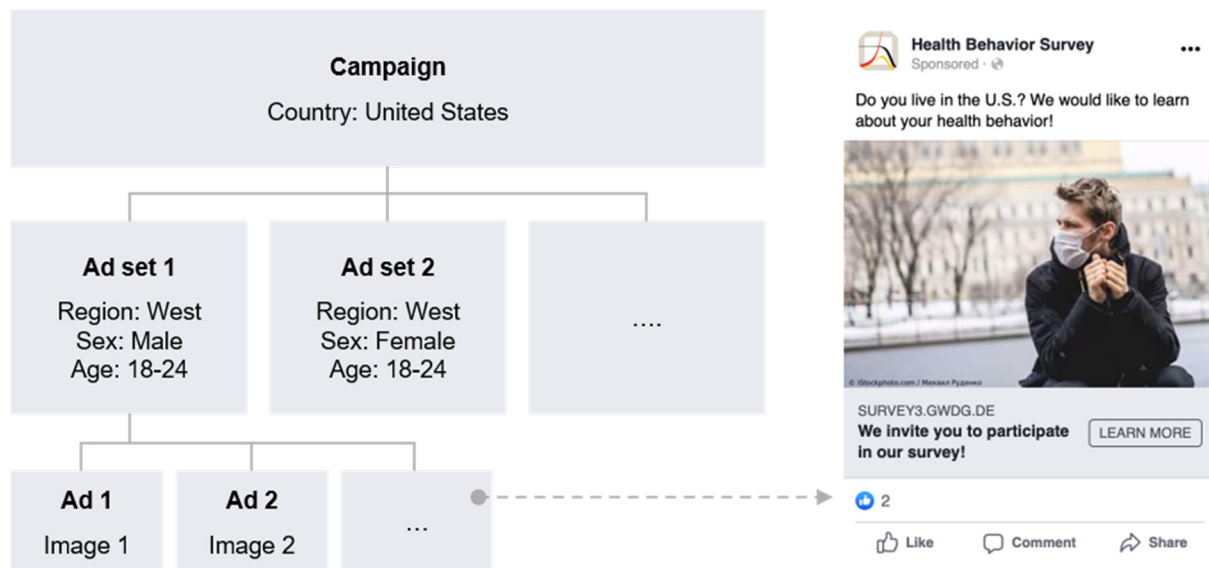


Figure 1. Illustration of Facebook advertising campaign used in the United States.

Source: Fig S1. in Perrotta et al. (2020)

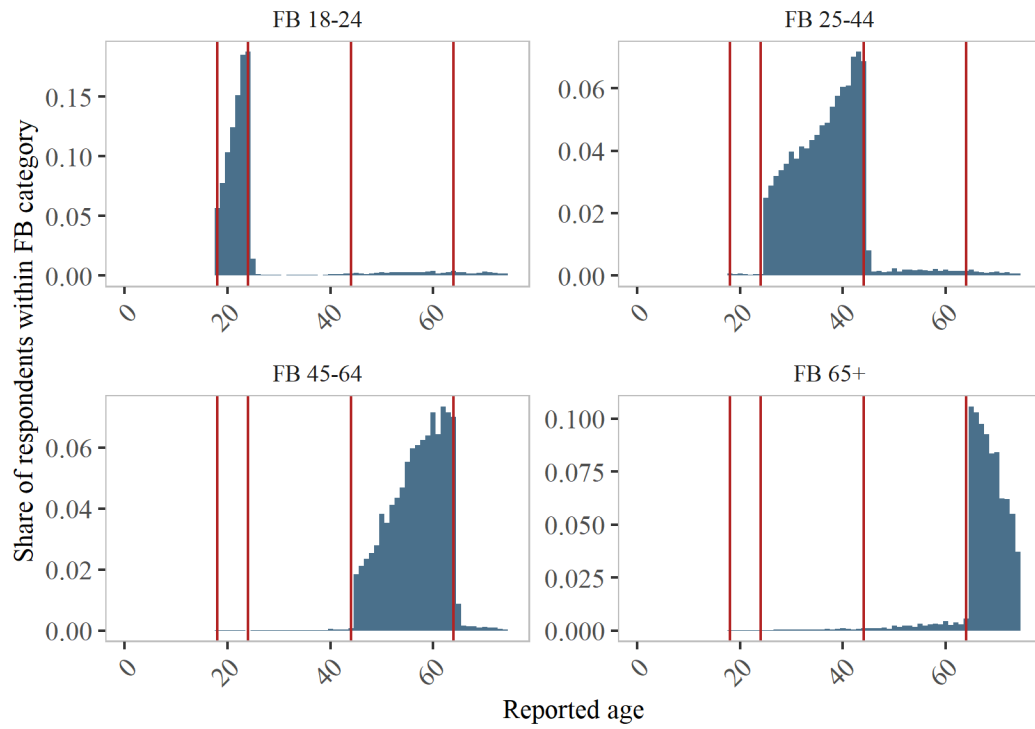


Figure 2. Share of respondents who reported a given age by Facebook's (FB) age classification. The red, vertical lines indicate age-group boundaries. Plot has been truncated at age 75.

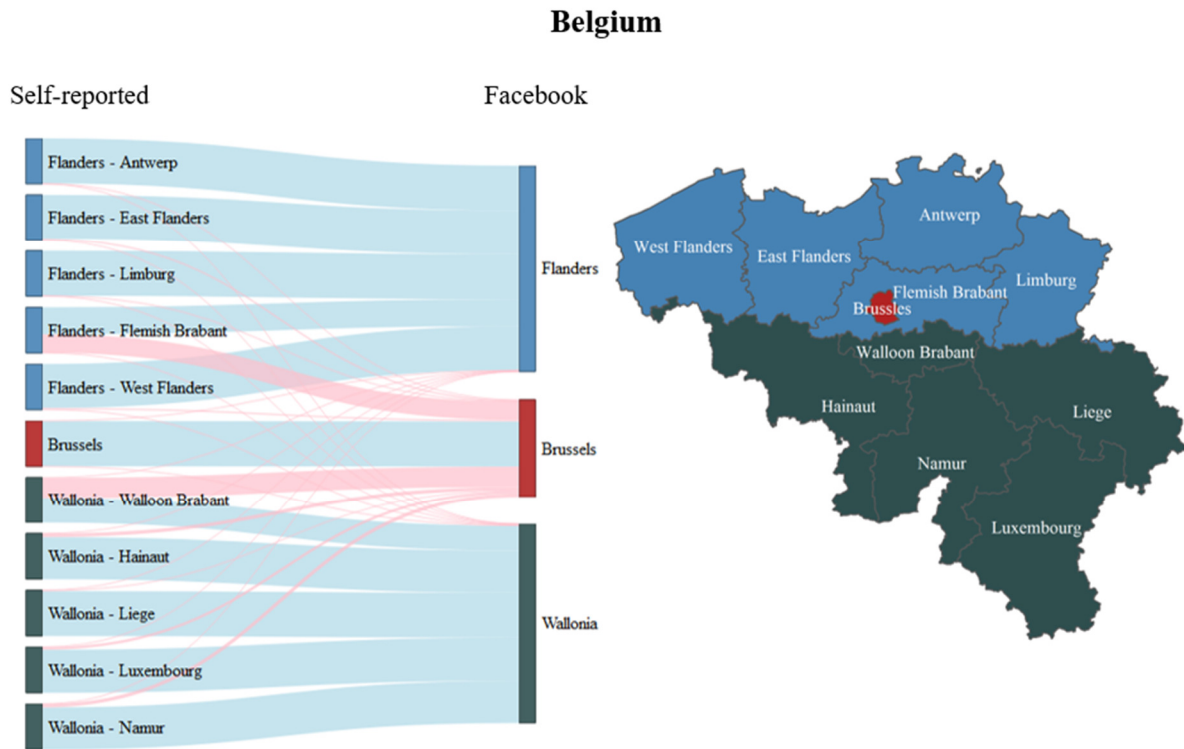


Figure 3. Respondents' reported region in comparison with Facebook's categorization in Belgium. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

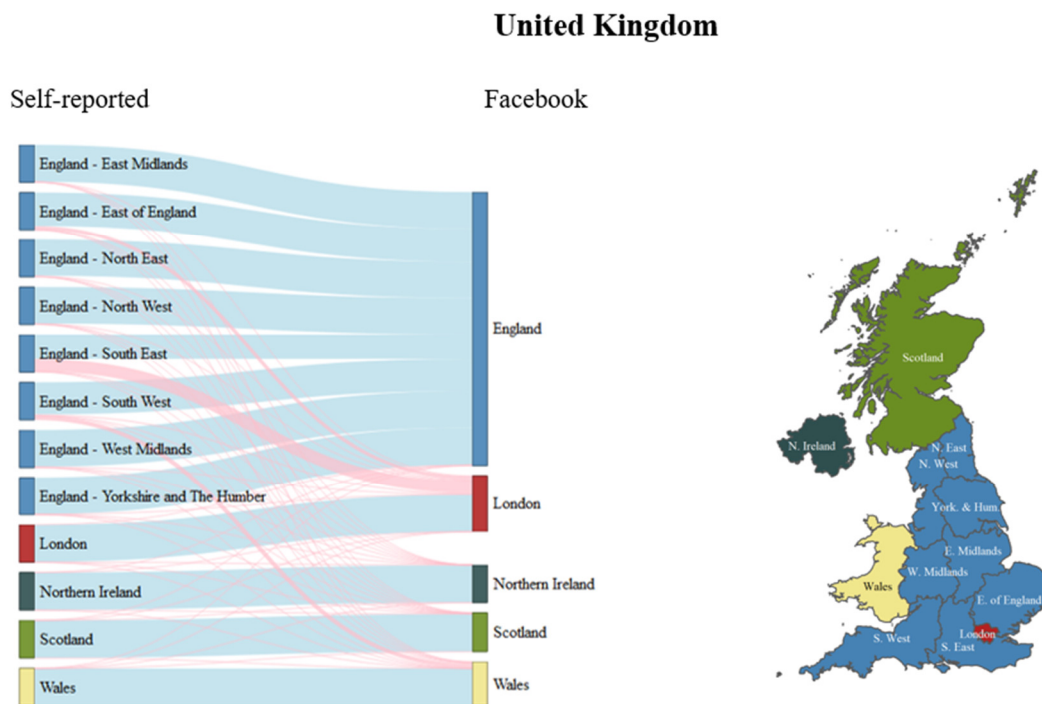


Figure 4. Respondents' reported region in comparison with Facebook's categorization in the United Kingdom. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

Supporting Materials

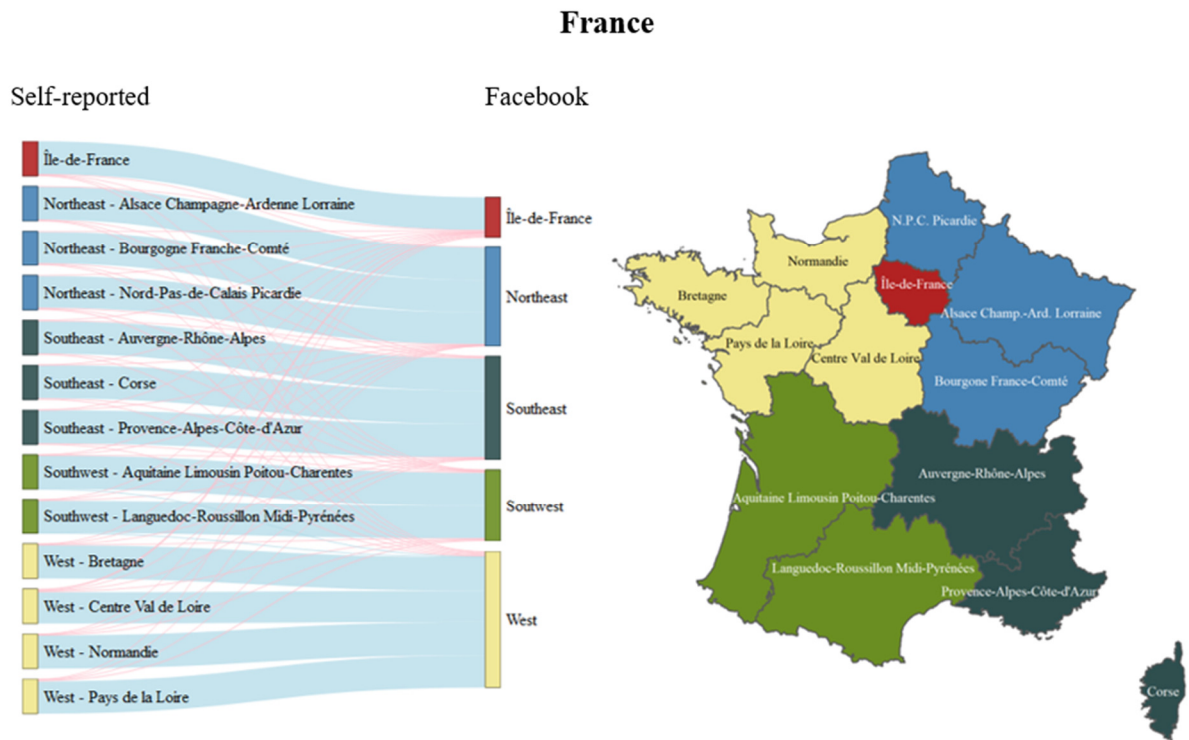


Figure A1. Respondents' reported region in comparison with Facebook's categorization in France. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

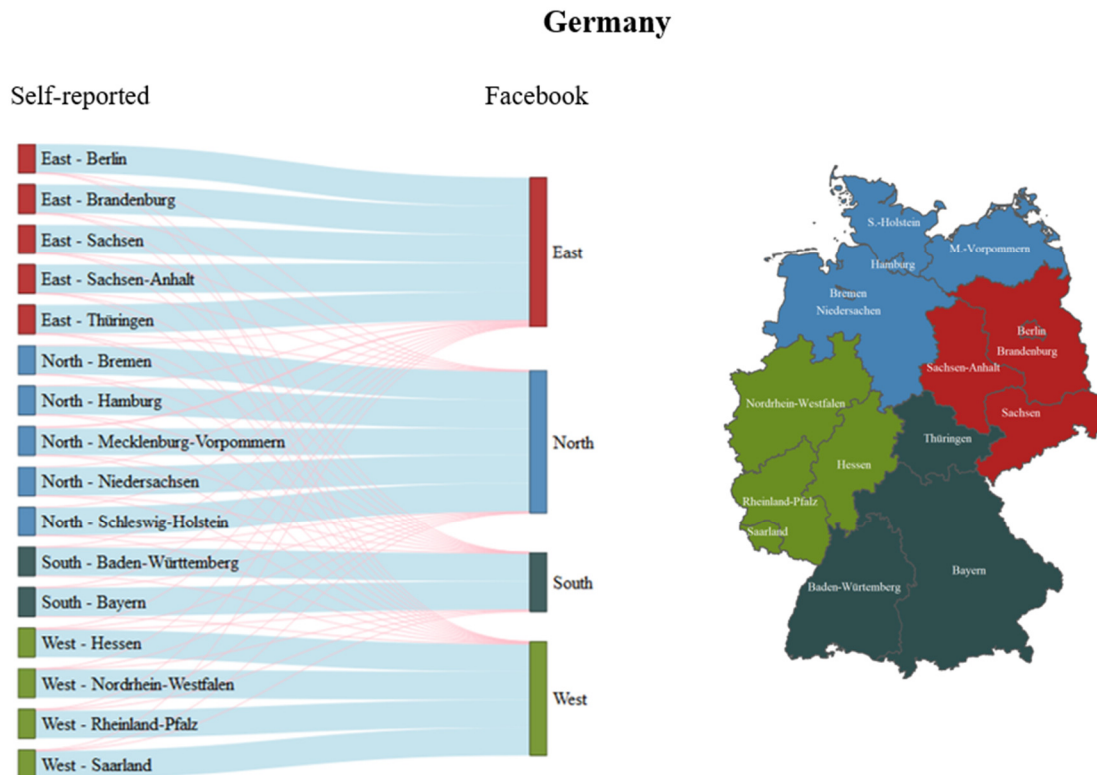


Figure A2. Respondents' reported region in comparison with Facebook's categorization in Germany. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

Italy

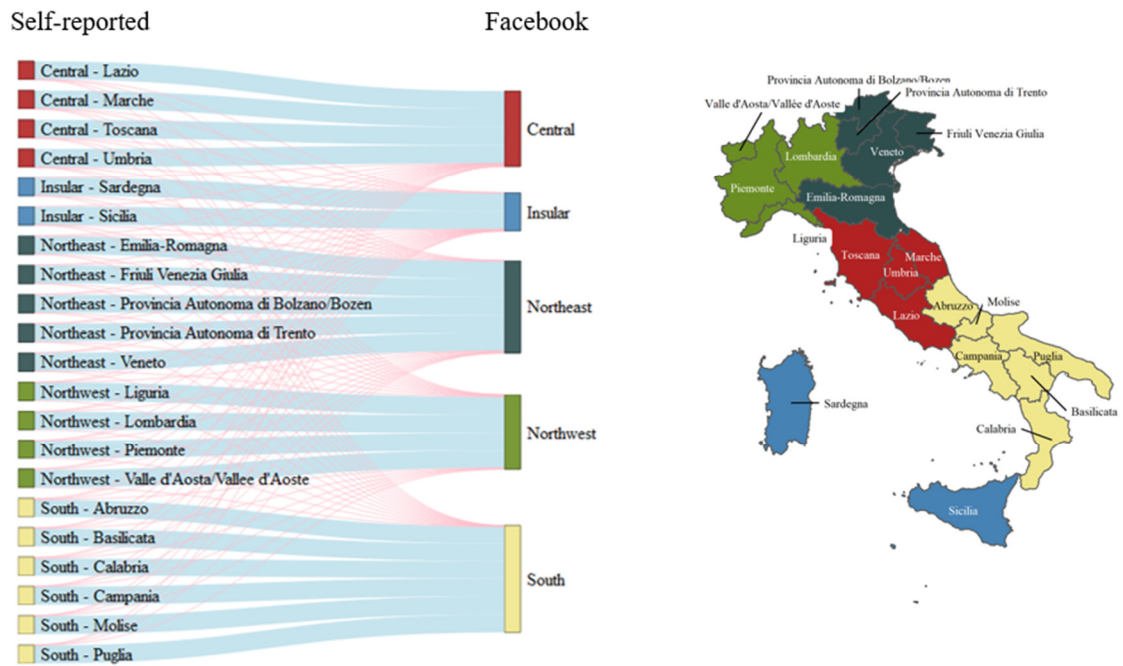


Figure A3. Respondents' reported region in comparison with Facebook's categorization in Italy. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

The Netherlands

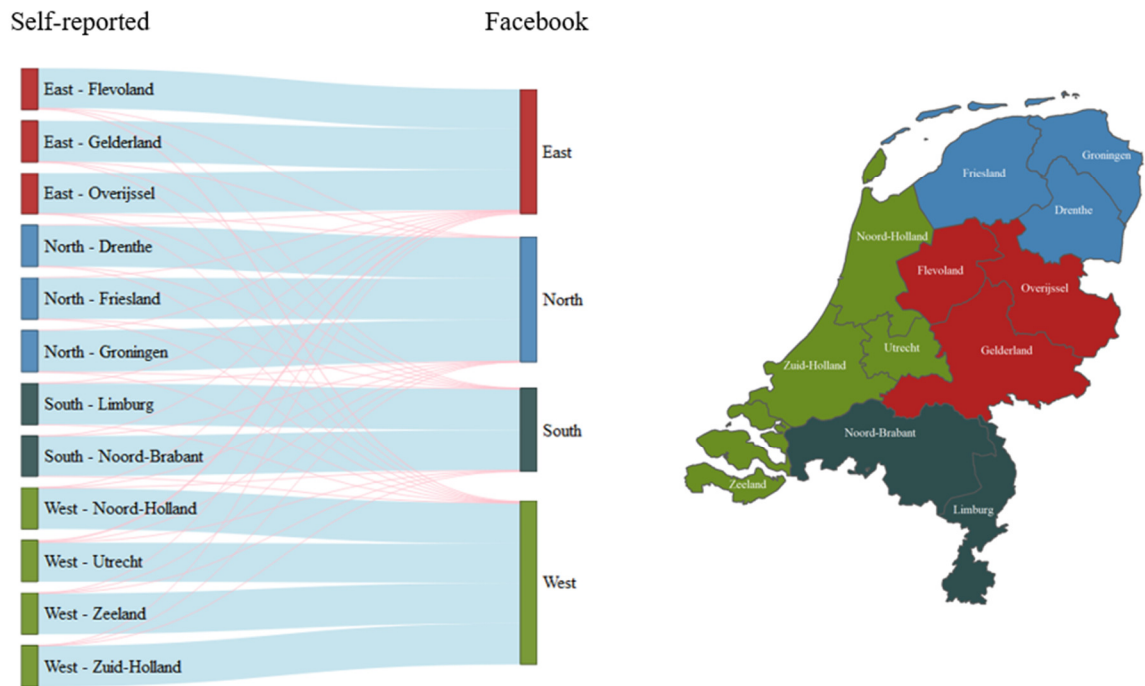


Figure A4. Respondents' reported region in comparison with Facebook's categorization in the Netherlands. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

Spain

Self-reported

Facebook

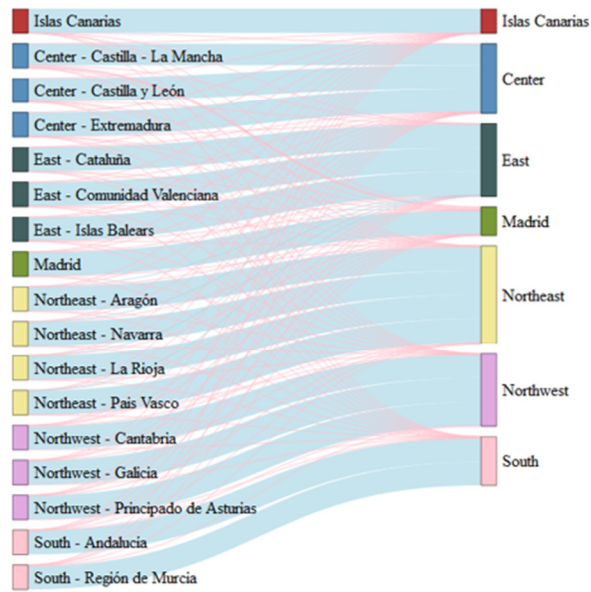


Figure A5. Respondents' reported region in comparison with Facebook's categorization in Spain. Blue lines indicate correct classifications, red lines indicate incorrect classifications.

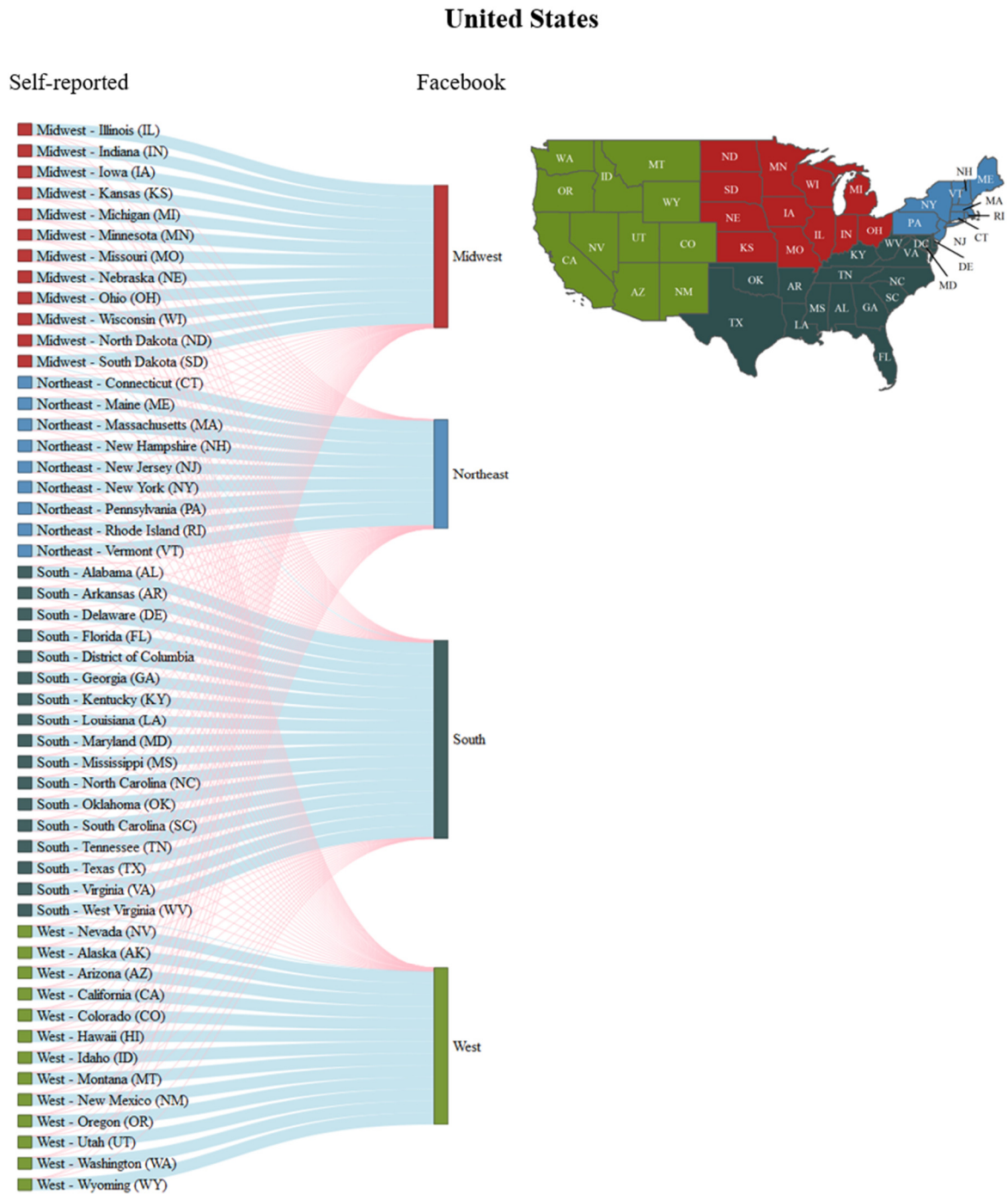


Figure A5. Respondents' reported region in comparison with Facebook's categorization in the United States. Blue lines indicate correct classifications, red lines indicate incorrect classifications.