

The Blind Spots of Measuring Online News Exposure: A Comparison of Self-Reported and Observational Data in Nine Countries

Sandra González-Bailón^a and Michael A. Xenos^b

^a University of Pennsylvania, sgonzalezbailon@asc.upenn.edu

^b University of Wisconsin-Madison, xenos@wisc.edu

Abstract: Measures of news exposure are common in research that tries to explain political knowledge, political engagement, opinion formation and, more generally, media effects. Much of that research employs self-reported measures obtained with surveys, known to suffer from accuracy problems. Observational measures, however, also suffer from limitations derived from data collection and instrumentation. Here we offer new comparative evidence on the nature of those problems. We show that commonly used self-report measures of digital news consumption are problematic for three reasons: they only pay attention to a small fraction of all available sources; they underestimate audience share; and they distort the relative position of news sites in visibility rankings. Measurement problems, however, also exist in observational studies, especially when mobile access is excluded from data collection. Our analyses quantify the magnitude of these problems, offering unprecedented comparative evidence of online news consumption that spans nine countries and a period of five years. We discuss the implications of our findings for future research on news exposure.

Keywords: news consumption; web; digital traces; exposure; measurement error.

Acknowledgements: The idea for this paper originated while the authors were visiting the Weizenbaum Institute in Berlin, Germany in the summer of 2019. We would like to thank Martin Emmer, Annett Heft, Curd Knüpfer and Pablo Porten-Cheé for their invitation to visit. This work has partly benefited from funding from the NSF (grant numbers: 1729412 and 2017655).

1. Introduction

How do people consume news? This question is at the heart of current discussions about the nature of the digital media landscape and the impact news exposure has on a variety of political processes. Digital media have multiplied the sources that are available to follow current affairs, but they have also allowed entertainment to overcrowd political content. Political accountability requires informed citizens, and measures of news exposure can help us analyze levels of political knowledge, political engagement, and shifts in public opinion. Understanding how people consume news in the digital environment is, therefore, essential to assess the state of our democracies and build theories about the political impact of online technologies.

Surveys have traditionally offered the main measurement instrument to assess news exposure. This approach assumes that respondents have an accurate recollection of the news sources consumed and/or the frequency of that consumption. In the past ten years, several scholars have explored a variety of biases that such measures introduce into estimates of news exposure, typically through comparisons to observational data. This work relates to three types of media exposure: exposure to TV news (Prior, 2009a, Konitzer et al., 2021); exposure to news on the web (Guess, 2015, Revilla et al., 2016); and exposure to news in social media (Guess et al., 2019). This scholarship suggests that commonly used survey measures introduce substantial measurement error, inflating estimates in the case of TV exposure and, to a lesser extent, in the case of web exposure. A recently published meta-analysis provides a thorough and systematic account of the modest correlations between self-reported and logged media use (Parry et al., 2021): fewer than 10% of the studies analyzed provide estimates falling within 5% of each other.

Alongside these discussions, multiple strategies for reducing bias and measurement error have been proposed, such as using open-ended prompts (Guess, 2015), or framing questions

around lists of specific news outlets and brands (Andersen et al., 2016, Dilliplane et al., 2013). Others have proposed combining questions about exposure with smartphone apps that prompt respondents to produce a novel kind of diary measurement (Ohme et al., 2016); or taking snapshots of screen activity to generate a rich stream of usage patterns (Reeves et al., 2019). However, given the inherent difficulties of working with open-ended survey responses, and the resources required to implement app-based measures, many researchers are still inclined to view list-based survey measurements as a reasonable compromise. Past research has suggested that this approach to media exposure offers a level of flexibility that is well-suited to the high-choice media environment and can provide predictive validity in the context of television news, which remains one of the most popular forms of news consumption (Andersen et al., 2016; Dilliplane et al., 2013). A number of major surveys follow this basic approach to analyze news consumption in the digital domain (e.g., Newman et al., 2019).

This paper contributes to these discussions by offering a specific set of comparisons between list-based recall data and observational data that is novel in three ways. First, the data we analyze integrates exposure to news sources through multiple devices (desktop, tablet, and mobile). Past research focuses on TV, or on specific social media, or on web browsing behavior tracked through plug-in extensions that only work in desktop computers. Excluding multi-platform estimates from observational studies, we argue, also leads to problems in the measurement of exposure and engagement with the news. Second, our data covers nine different countries: Australia, Brazil, Canada, France, Germany, Italy, Spain, the UK, and the US. Our analyses add a comparative perspective to the current literature, which is overwhelmingly focused on the US (with some exceptions, e.g., Scharkow, 2016, Jürgens et al., 2020, Araujo et al., 2017). We expand the empirical focus of past work to include countries that have been

disregarded so far (even though our focus is still predominantly confined to western democracies, leaving the global south largely unrepresented). And third, our data covers a period of five years (2014-2018), offering the longest observation window to date to compare survey and observational measurements of digital news exposure.

In what follows, we first review past strategies to measure news exposure, with a focus on digital news consumption. We review the questions used in survey-based research, and we then consider the observational measures that researchers have extracted from digital trace data. Section 3 introduces the data and the methods we use in our analyses. Section 4 presents the results, which we organize around three research questions: (1) What is the discrepancy in estimated audience reach? (2) What is the discrepancy in number of sources covered by the measurement? And (3) what is the discrepancy in the visibility rankings drawn from the data? Our comparisons suggest that – contrary to well-known patterns of *over*-estimation associated with television news exposure (Prior, 2009a, Prior, 2013b) – list-based recall measures of online news exposure show a robust pattern of *under*-estimation. This pattern is only visible if multi-platform measures (including mobile access) are used as observational metrics: desktop-only estimates (the predominant choice in past work) leads to a very different assessment of the difference in reach estimates. Focusing attention on just a small fraction of all available outlets, and the distortion in visibility rankings also limit, we argue, the characterization that surveys offer of the digital news environment.

In our discussion of these findings, we offer suggestions on how to move forward in news consumption research – addressing a fourth research question: how can we overcome some of the current limitations in measurement? The first suggestion is to pay more attention to activity in the long tail of news media options – a tail that surveys predominantly ignore but where the

niche behavior so characteristic of digital environments takes place. The second suggestion is to use alternative metrics to characterize news diets. We propose devising measurements that capture engagement, not just exposure, because they are more granular and informative than binaries (i.e., exposed/not-exposed). This is something prior research had already suggested in the context of survey data (Andersen et al., 2016) but observational data has higher resolution to capture temporal dynamics. We conclude with a summary of our results and implications for future scholarship on the nature of the digital news environment and, more broadly, on the effects of digital news exposure.

2. Measures of Digital News Exposure

Traditional approaches to measuring exposure with surveys involve asking respondents about their typical news use patterns, or whether they watched, read, or listened to news within a specified period. In the past ten years, researchers have analyzed why commonly used survey-based measures of exposure have measurement error. On the one hand, recall-based measures often place unrealistic demands on respondents' memory, and this translates into imperfect recall (Prior, 2009b). On the other hand, general questions about typical exposure to news content, or broad categories of sources, necessarily presume that the average person will reliably draw on a valid definition of what exactly constitutes "news", which is arguably a fuzzy term (Mutz and Young, 2011). As noted earlier, one approach to minimizing faulty recall involves explicitly listing the sources of exposure, for instance, specific TV programs (Dilliplane et al., 2013, Andersen et al., 2016). This strategy reduces the cognitive demands placed on respondents, and it promotes content validity by narrowing down a relevant domain of exposure. However, common

deployments of this measure fail to account for the actual amount of exposure (Prior, 2013a) and, when applied to TV consumption, they suffer from over-reporting (Prior, 2009a, Prior, 2013b).

Digital technologies have multiplied the sources that are available for consumption, in the process making the measurement of exposure even more difficult. However, digital technologies have also made it easier to track exposure as it happens, offering an important observational benchmark with which to compare self-reported survey data. The availability of digital traces has allowed researchers to provide some validation to survey measures. On the web, for instance, certain types of survey questions also lead to over-reporting (as with TV exposure) but this bias can be mitigated with the use of open-ended prompts (Guess, 2015). Researchers have also shown that surveys can be accurate in measuring social media activity related to the generation or dissemination of news content (Guess et al., 2019). In addition, research also suggests that surveys are unreliable when it comes to measuring exposure to false news stories. In a survey following the 2016 election in the US, researchers found that 15% of the respondents recalled seeing false news listed for them in the survey; and yet a placebo question that listed headlines made up by the researchers also gave a similar response: 14% responded seeing placebo stories that never existed in the first place (Allcott and Gentzkow, 2017). These findings put the question of the validity and accuracy of self-reported measures back at the heart of the news exposure debate.

Beyond validation, observational measures of exposure have allowed researchers to test hypotheses about ideological segregation (Gentzkow and Shapiro, 2011, Flaxman et al., 2016, Bakshy et al., 2015, Yang et al., 2020b); the impact of aggregators on news consumption (Athey et al., 2017); or the reach of untrustworthy information (Vosoughi et al., 2018, Grinberg et al., 2019, Guess et al., 2020). The sources of observational data that researchers have used in these

studies include web tracking (through panels of the online population or through the logs produced by search engines), and social media activity. Exposure across these studies is measured in slightly different ways: in some studies, it means accessing a news URL from desktop computers; in others, it means having a particular story or source appearing in the news feed; and yet in others it means posting and engaging with news sources in social media platforms. One difficulty in the current media landscape is that there are different ways of making the concept of “exposure” operational. Some options imply passive (incidental) exposure (what has been labeled as the ‘news-find-me’ phenomenon, Gil de Zúñiga et al., 2017); some others imply a more active type of deliberate exposure (as when an actor accesses a news source through intentional browsing). We argue that these two entry points to news co-exist but that each is more or less prominent depending on the mode of entry or the device used (e.g., desktop vs mobile).

Past research does not cast much light on this source of variation in observational data, but there is growing evidence suggesting that online intermediaries (i.e., search engines, social media) diversify news diets because of incidental exposure (Scharkow et al., 2020), and incidental exposure (i.e., the ‘news-find-me’ phenomenon) is more likely to happen on mobile via social media and notifications. There are other ways in which prior research is limited. The first limitation is that it is disproportionately focused on the US: there is a serious lack of comparative evidence in terms of validating survey measurements and analyzing different media contexts. There are some exceptions to this trend (e.g., Revilla et al., 2016, Jürgens et al., 2020, Scharkow, 2016, Araujo et al., 2017). This research expands the empirical focus beyond the US, but it also misses the cross-country comparative dimension and spans a much narrower time window (two months of data at most). The second limitation is that exposure is often treated as a

binary variable, i.e., a person reads or does not read a particular news site. This measurement disregards the actual level of engagement with the news, e.g., time spent consuming information from a particular source. Here, we offer evidence that addresses these two limitations to shed new light on the validity and usefulness of survey and observational measures.

3. Data and Methods

We use two sources of data to compare self-reported and observational estimates of online news exposure. We use data from the Digital News Reports (DNR), published by the Reuters Institute for the Study of Journalism, to obtain information on self-declared exposure to news sources online (Newman et al., 2019). These surveys are fielded by YouGov using an online questionnaire, with sample sizes of about 2,000 respondents in each country (see Table A2 in the Appendix). We use this survey because it is highly regarded in the academic community for its value as a research source; it is also unique in that it runs annually in several countries; and it asks questions about the frequency of online news exposure as well as about exposure to specific news outlets (which are our main unit of analysis). As noted earlier, this style of survey measure is particularly attractive to researchers who do not have access to observational data and want to guard against the limitations of fuzzier versions of the question, e.g., general questions about typical exposure that do not specify what constitutes “news”. In particular, the question asked by the DNR is “Which of the following outlets have you used to access news online in the last week (via websites, apps, social media, and other forms of internet access)?” Respondents in each country select the outlets from a list of about 30 of the most popular—where “popular” refers to the audience share of the outlets in each country (as provided by media measurement companies

like the one we use here as our observational source of data). In the end, the reports publish estimates for about 14 of these outlets (Fletcher and Nielsen, 2017: 485). We don't use data from the full set of 30 outlets because those in the lower tail of the distribution are nominated by very few respondents, so the estimates for those outlets are more unreliable; in addition, the outlets highlighted in the reports are also presented as the most relevant – hence our decision to focus on these. We use the percentages reported annually in each of the nine countries we consider.

Our observational data comes from the online measurement company Comscore. We use their Media Metrix Multi-Platform datasets to obtain the estimated percentage of the online population accessing websites in the category News/Information. These datasets rely on representative panels of the online population in each country, with sizes that vary between ~8,000 to 300,000 participants and that include activity through desktop and mobile devices, integrated in multiplatform estimates (see Table A1 in the Appendix for more details on the data). The list of web sites that are included in the category News/Information changes across countries, but the measurement methodology is the same: it relies on a combination of panel-based tracking (i.e., what panelists do while online) and census-based site analytics (i.e., traffic information generated at the site level). As far as we know, this is the only source of behavioral data tracking online activity that is directly comparable across the countries we consider here. We can also compare this data to the survey data because the estimates are weighted to be representative of the online population of each country. Likewise, in both cases news outlets are the basic unit of analysis, with audience reach (i.e., the fraction of the online population consuming the news outlet) serving as the main comparison metric. The Appendix offers more details about our two sources of data.

Our analyses rely on the outlets that were present in both the survey and the observational datasets. In line with our research questions, our goal is to identify discrepancies in the estimated reach of news sources and quantify the nature of that discrepancy. We follow prior work in making the assumption that observational data offers a more accurate approximation to the ground truth (Revilla et al., 2016, Prior, 2009a, Guess, 2015, Guess et al., 2019, Parry et al., 2021, Konitzer et al., 2021). As this research has suggested, self-reported data suffers from imperfect recall, inattentive responding, and social desirability bias, whereas behavioral trace data is immune to this type of subjective measurement error. Behavioral traces, on the other hand, suffer from other measurement issues: as we discuss below, desktop-only estimates also offer a distorted view of actual exposure because, as we argue, they do not accurately capture incidental encounters with news. One limitation of our approach is that we are unable to compare survey and observational data for online sources of local news. While the DNR has a generic category for ‘local/regional newspapers,’ the data do not allow us to parse out which newspapers are included in this group, so we cannot match them to our web browsing data. These local and regional outlets are still represented in the observational dataset, and they are part of some of the analyses below, but not those that provide a direct comparison with survey data. Nonetheless, relative to prior work in this area (which also focuses exclusively on national sources), our analyses still offer a more comprehensive validity test by expanding the number of countries considered as well as the observation window used to assess change.

4. Results

The most salient finding emerging from our analyses is a clear pattern of *underestimation* when it comes to list-based measures of online exposure to specific news sources. The observational estimates reveal moderate to large discrepancies for most news outlets included in the survey

estimates – directly opposite the well-known pattern observed for TV news (Prior, 2009a, Konitzer et al., 2021). What this finding suggests is that individuals are exposed to more online news from a particular outlet than we might infer based on list-based survey questions. We further identify two additional problems with the list-based approach. One is the bias that results from the selection of news sources: survey questions can only ask about exposure to a selected number of outlets (often limited based on considerations of survey length), which inevitably amount to a small proportion of all sources available online. In this part of our analyses, we consider the extent to which surveys miss activity in the long tail of media options that have a lower reach. The other problem is that a list-based approach can significantly distort the relative salience of news sources: in addition to underestimating reach, it can also lead researchers to assign substantially different ranking positions in visibility scales. The following sections discuss each of these issues in more detail (the Appendix casts additional light on discrepancies in overall trends over the 5-year period we consider, see figure A1).

4.1. Estimated audience

Here we consider our first research question: what is the discrepancy in estimated audience reach between the survey and observational estimates? Both data sources enable estimates of the proportion of individuals that are exposed to a particular news outlet. We refer to this proportion as the ‘audience reach’. Figure 1 plots the gaps in reach estimates for specific online news sources in the nine countries. The figure summarizes the differences in estimated audience size when we compare self-reported data to observational data from multi-platform access (including mobile). The horizontal axes measure the overall percentage reach for each source of data: the red dots locate the survey estimates and the blue dots locate the observational estimates. As the figure shows, surveys systematically underestimate the percentage of people who access news

sites in most countries and for most outlets. One interesting pattern is that the cases where surveys inflate reach (vis-a-vis observational estimates) tend to involve the smaller sites at the bottom of the observational ranking. Brazil and, for the most part, Australia are notable exceptions in that the data suggest the opposite trend: for most news outlets, the survey data offer higher estimates. This inconsistency with the other countries might be an artifact of how the survey was implemented in these countries. The DNR notes that in Brazil, for example, the sample is more representative of urban than national populations. Regardless, the evidence suggest that in most cases survey estimates deflate the actual proportion of the population that is exposed to news online. This finding is the opposite of what past work found in relation to TV news consumption (Prior, 2009a, Konitzer et al., 2021) and in relation to web exposure when non-open-ended questions were used (Guess, 2015). Crucially, these results would look very different if, for the observational data, we only used desktop-based estimates (see Figure A3 in the Appendix). Most prior research using observational data of online news exposure is based on desktop-only measurements, which means that they miss an important and increasing share of online activity since mobile access has become one of the main channels of exposure to online content (Yang et al., 2020b). Mobile access is also likely to make the role of intermediaries even more important in the diversification of news diets (Scharkow et al., 2020).

-- Figure 1 about here --

4.2. Selection of news sources

Here we consider our second research question: What is the discrepancy in number of sources covered by the measurement of exposure? One inherent challenge of survey instruments is the increasing number of available media options in our high-choice media environment.

Specifically, “the more personalized and fragmented media exposure becomes, the longer the lists of specific sources become as well” (Andersen et al., 2016). In contrast to the limited array of news sources in the traditional information economy, online news sources present a long tail of smaller sources (in terms reach) that extend to the right of the distribution headed by the most popular outlets. As a result, and keeping the fatigue of respondents in mind, most survey researchers end up with a reasonably good coverage of popular outlets at the top of the distribution, but they inevitably miss important outlets spread across the long tail. Indeed, this is the pattern that clearly emerges from our analyses when we examine aggregate pictures of source salience.

Figure 2 summarizes the reach distributions for news outlets included in the observational data (blue bars) and the survey data (red lines, inset plots) for the year 2018. Each blue bar represents a single news website, and its height is proportional to its audience reach; each red line in the inset plots also represents a news outlet, but its height is proportional to the audience reach according to the survey data. The pie charts summarize the fraction of all news sources present in the observational data that are also present in the survey estimates (red sections).

The first pattern that stands out from this figure is that, consistent with theoretical treatments of the networked information economy (e.g., Benkler, 2006), reach is highly concentrated in a few major outlets across all countries, although the levels of concentration are higher in some (e.g., Brazil) than others (e.g., Germany). This is consistent with previous work showing that news exposure is highly concentrated around a few sources (Gentzkow and Shapiro, 2011, Hindman, 2009, Flaxman et al., 2016, Yang et al., 2020a). These few outlets tend to be those usually considered in survey questions. Importantly, however, they are not

necessarily the most popular online or those attracting higher levels of engagement (more on this below).

The second aspect that stands out is that survey questions typically only give us information about a very small number of all available sources. In this case, because we focus on the top ~15 outlets, coverage amounts to 8 to 12% of all sources that the online population consumes. Again, this limitation stems from the practical concern of reducing respondents' fatigue and getting enough respondents identifying the sources; but the consequence of this is that surveys end up offering a very narrow window of a wider media environment that goes, for the most part, untapped. Even though the outlets at the tail of the online distribution have a much lower reach than the big outlets at the top, they create the niches that are so central in the digital media landscape. Breitbart, for instance, is a prominent example of this type of niche outlet – its relevance does not derive from its total audience reach, which is much lower than that of many other outlets, but from the engagement it generates in certain populations (again, more on this below).

Finally, Figure 2 also shows that, consistent with the underestimation patterns described in the previous section, survey estimates of audience reach are substantially lower than those obtained from the observational data –i.e., the range of the vertical axes in the inset plots is about half in most countries compared to the observational plots.

-- Figure 2 about here --

4.3. Salience of news sources

In this section we consider our third research question: What is the discrepancy in the visibility rankings drawn from the observational and survey data? Our analyses suggest that a third

problem arises from the way that survey data yields substantially different rankings of news sources in terms of relative visibility. This problem, when combined with the previous one, provides a clear picture of how a list-based, recall measure can lead researchers to form an inaccurate picture of the relative salience of various online news sources. Figure 3 helps visualize this issue by plotting associations between the rank position of news outlets in each of our nine countries, based on data aggregated from the list-based survey measures (vertical axes) and the observational data (horizontal axes). In any given plot, agreement between these two measures would show as an alignment of the estimates from both data sources along the diagonals. But as the panels of the figure clearly show, the estimates are far from aligning, especially in the cases of Germany, the US, Spain, Canada, Italy and France – where the correlation coefficients are all equal or below 0.5. Australia, the UK, and Brazil have the highest agreement in the relative salience of news outlets, with correlations that range between 0.7 and 0.85. These findings suggest that, for most countries, we are not just underestimating how many people are exposed to these news sources, but also how salient the sources are relative to each other. This discrepancy in rank position is similar if we use desktop-only estimates for the observational data (see Figure A4 in the Appendix).

-- Figure 3 about here --

5. Discussion

In recent years, researchers have begun to take the limitations of self-reported measures of news media exposure seriously, and a large and growing body of work suggests substantial measurement error in self-reported measures (e.g., Parry et al. 2021). Many of these errors arise from the difficulties that survey respondents face when asked to report how often they engage in activities that are of great importance to researchers but typically only of passing and vague

interest to them. Most researchers agree that there is value in employing alternative measurement strategies, such as observations drawn from digital trace data or more complicated app-based, diary measures. However, many are still reliant on traditional survey instruments, due to resource limitations or other factors of research design. In these cases, thoughtful researchers may seek to strike a compromise by opting for specific survey measurement strategies, following evidence suggesting that these may at least mitigate some of the most obvious problems of self-reported measures.

In this paper, we have explored what we believe is one of the most popular examples of such an approach, e.g., list-based recall questions. These items ask respondents to report whether they access news from specific online news sources, which are presented in a list usually compiled based on the reach ratings provided by media measurement companies (e.g., Comscore). Like many researchers who have probed the limitations of survey-based measures of media exposure, we do so by comparing survey data to data drawn from observational traces. Unlike previous studies, however, our comparisons draw on data from nine different countries over a period of five years, and use multi-platform estimates that consider online activity through desktop and mobile devices. Incorporating mobile access is, as we show, crucial for the findings we report: without it, the comparisons with survey data would look very different. In doing so, we reveal compelling evidence that list-based measures underestimate online news exposure – but so does observational data that does not consider multi-platform (including mobile) activity. We also show that the survey estimates paint a distorted picture of the online news environment by disregarding the long tail of media options and altering the relative salience of different online news outlets.

5.1. Future Research

So how do we alleviate some of the current blind spots in our measures of news exposure? One priority for future research should be to find ways to integrate different sources of observational data, e.g., traces left on the web but also on social media or digital TV. This integration is necessary to offer a more comprehensive account of news consumption, but there is still a scarcity of research in this domain (for an exception see Allen et al., 2019). Currently, this integration requires working with proprietary data that is prohibitively expensive for most researchers and that undermines the transparency of the research process (Lazer et al., 2020). Likewise, the data license agreement researchers need to sign to work with proprietary data precludes publishing research that explicitly compares data across sources or tries to reverse-engineer their methodology. This imposes serious limits on the robustness tests that can be made publicly available to the research community. It is still possible, of course, to publish different papers using different sources to see if they reach similar conclusions – to engage, in other words, in cumulative research. Future work should replicate the analyses reported here with other sources of observational data to test if our findings still hold – an exercise that can only enrich the discussion on how to best measure exposure to news in the digital age.

List-based, survey measures of news consumption usually rely on the ratings provided by the same media measurement companies providing observational metrics to decide which news outlets to include in the questionnaire list. Creating publicly funded data resources offering the same observational benchmarks would expand greatly our ability to do open science in this domain; it would also help address some of the privacy concerns associated with proprietary data. This challenge, in any case, goes beyond the specific needs of political communication

research, and it requires long-term strategic planning from several institutional levels (Lazer et al., 2020).

Our results also suggest some more specific guidelines for future research. One of the blind spots affecting survey measures affects outlets that may not represent large numbers in terms of reach but that can still trigger important effects. Coming up with better measures of engagement with the news (in addition to the binary exposed/not-exposed) is one priority for observational studies. Here, we consider our fourth research question: How can we overcome some of the current limitations in measurement? We argue that measures of time spent offer a more accurate representation of engagement with the news than binary operationalizations.

Figure 4 compares these two alternative measures of news consumption: the fraction of people accessing a particular news site (yes/no), and the time spent on that site (e.g., minutes spent, data aggregated for 2018). As the scatterplots show, surveys tend to focus on the outlets with a larger audience base, but not necessarily with the most engaged base. The correlation coefficients show that there is a significant disagreement between these two measures of exposure – one of which (time spent) is more granular than the other (count of exposures). Regional and local outlets, as well as niche outlets (e.g., Breitbart, RealClearPolitics), are in the category of sources with low reach but highly engaged audiences (in terms of time spent). Past work had already suggested that engagement metrics based on frequency of exposure (rather than mere exposure) are more appropriate to characterize media consumption (Andersen et al., 2016). Online tracking data is time-stamped, so it allows us to be more granular in measuring media engagement as a more direct measure of meaningful exposure.

-- Figure 4 about here --

Figure 5 pays closer attention to these two definitions of exposure by disaggregating the US data in terms of device used for news consumption. The idea here is to test if mobile access (which, as we argue, is more fleeting and conducive to passive or incidental forms of exposure) is associated with the consumption of certain news outlets, particularly with news outlets at the tail of the reach distribution. Panels A and B show that the average time spent by visitor is only weakly associated with the percentage reach of the news site, especially for desktop access – again, supporting the idea that smaller sites (in terms of audience base) can still attract high levels of engagement (in terms of time spent). Panel C shows that the percentage reach of news sites is largely similar regardless of the device used to access the sites; however, there is more discrepancy in the time spent: people using desktops spend more time on news sites than when access comes from mobile devices (panel D). These patterns suggest that regional and local outlets, as well as niche outlets (e.g., Breitbart, RealClearPolitics), are in the category of sources with low reach but highly engaged audiences (in terms of time spent on their sites). This type of tail activity is largely invisible to survey measures, but can be more easily tapped into with observational metrics.

-- Figure 5 about here --

5.2. Limitations

As all empirical research, our study has limitations that are important to note. First, the smallest temporal unit available within the Comscore data is one month, whereas the survey items we draw on from the DNRs ask respondents to indicate exposure “over the past week”. As a result, while we use a yearly average for most of our analyses, it is possible that the Comscore data include some instances of news exposure that are atypical, and thus less likely to be reflected in the estimates provided by survey respondents. This limitation may contribute to the patterns of

underreporting we discuss above. But given the substantial pattern of overall differences revealed in our analyses, we believe it is unlikely that this slight discrepancy in periodicity is responsible for the underreporting pattern we clearly identify; it also does not explain why, when we exclude mobile access, desktop-only estimates are often lower than survey estimates. A conservative interpretation of our findings considering this limitation would still suggest a pattern of underreporting in list-based survey recall measures for news exposure, especially for sites located in the long tail, beyond the reach of most lists used by survey researchers. In addition, providing a reference period (i.e., asking how often ‘in the past week’ someone got news rather than how often in general) has been reported to have no effects on survey responses (Pew Research Center, 2020).

Second, while our comparisons provide a compelling picture of measurement bias and overall distortion associated with list-based measures of exposure, they do not explain what underlies the general pattern of underestimation. Some demographic groups, for instance, might have a greater weight on the overall bias, which would imply that surveys are better (or worse) for certain sectors of the population. Even though the two sources of data we analyze are weighted to be representative of the online population of the respective countries, they are still based on different sets of individuals. This limits the research we can do about individual-level variance in the biases we identify (a limitation we share with recent research, e.g., Parry et al., 2021). In addition, the observational data is only available on the aggregate for news domains: our basic unit of analysis are registered domains, not individual users. The aggregated nature of our data helps protect the privacy of the panelists but the trade-off is that we lose individual-level information. This brings us to the third main limitation.

Since our observational measures work at the domain level, they do not offer insight on whether users accessing, say, the New York Times read the ‘politics’ section or if, instead, they just go to the ‘sports’ or other entertainment sections. This could explain why survey estimates are lower: if respondents (rightfully) do not consider their sports consumption as news consumption, they will not declare having used the New York Times – and these users will inflate the observational estimates. Having access to the specific URLs that people navigate while browsing a news media site would help determine if this is indeed what explains the discrepancies (not only between survey and observational estimates, but also between multi-platform and desktop-only estimates). It is certainly difficult to determine what goes on in the mind of survey respondents when asked about news exposure and, as we discussed above, there is an increasing body of empirical work showing that self-reported data suffers from imperfect recall, inattentive responding, and social desirability bias. Our findings suggest that the discrepancies are not only about underestimating the audience reach of news outlets: the discrepancies also affect their relative salience— in addition to the fact that list-based survey questions disregard most outlets in the long tail.

Our analyses also leave open the question of whether exposure makes any difference in political knowledge or behavior. There is increasing causal evidence suggesting that online activity has an impact on offline political action (for a recent review see Zhuravskaya et al., 2019), but there is yet not much large-scale evidence showing that exposure to news is equivalent to ‘meaningful’ processing of information or that it can lead to opinion change or behavior. Behavioral data is also limited in other respects, for which survey data might be more insightful. For instance, digital traces do not allow inferring reasons or motivations (e.g., what drives some people to consume certain news sources?); and they do not allow assessing how

much credibility people assign to certain news outlets. In this, surveys like the DNR are superior to any behavioral dataset –accessing a news domain is, after all, different from trusting the information published in that domain. These attitudinal data are crucial to understand some of the cross-country differences we identify here. Future research should pay more attention to the factors that explain the differences we document – and theorize about why the digital news landscape looks so different across media systems. This sort of theoretical work, however, requires settling on appropriate measurement first.

6. Conclusion

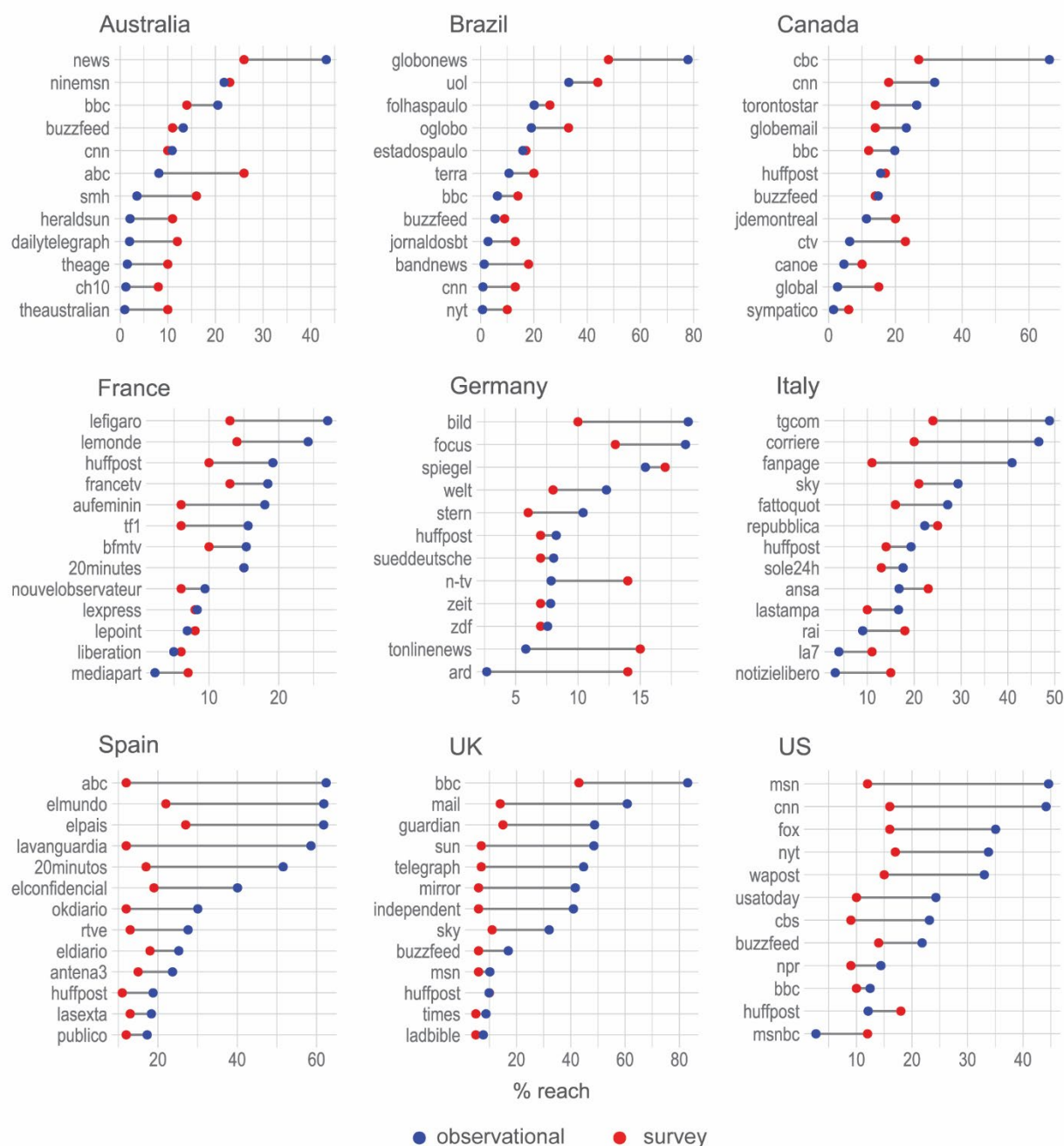
This paper presents evidence suggesting that list-based survey questions cannot successfully capture the patterns that characterize digital news consumption. The results discussed above add novel data to a line of research that had already identified bias in survey estimates when measuring exposure to TV news and to Web news in the US. Compared to that prior work, we consider a wider range of countries, a wider time window, and we pay more attention to activity taking place at the tail of the reach distribution (e.g., among the news outlets that are largely disregarded by survey questions). Crucially, we also incorporate in our estimates multi-platform access (including mobile) – which, as we show, is necessary to identify the patterns we report. We also propose using a second measure of exposure, in addition to audience reach, that considers the time spent in news sites – for which observational data provides more granularity than surveys. This second measure of exposure suggests that the least popular sites in terms of reach are not necessarily the sites with the least engaged audiences. In fact, many of those niche sites have audiences with engagement levels that are comparable to those of more mainstream news organizations.

References

- ALLCOTT, H. & GENTZKOW, M. 2017. Social Media and Fake News in the 2016 Election. *The Journal of Economic Perspectives*, 31, 211-235.
- ALLEN, J., HOWLAND, B., MOBIUS, M. M., ROTHSCCHILD, D. M. & WATTS, D. 2019. Evaluating the Fake News Problem at the Scale of the Information Ecosystem. *SSRN*.
- ANDERSEN, K., H. DE VREESE, C. & ALBÆK, E. 2016. Measuring Media Diet in a High-Choice Environment - Testing the List-Frequency Technique. *Communication Methods and Measures*, 10, 81-98.
- ARAUJO, T., WONNEBERGER, A., NEIJENS, P. & DE VREESE, C. 2017. How Much Time Do You Spend Online? Understanding and Improving the Accuracy of Self-Reported Measures of Internet Use. *Communication Methods and Measures*, 11, 173-190.
- ATHEY, S., MOBIUS, M. & PAL, J. 2017. The Impact of Aggregators on Internet News Consumption. *Stanford University Graduate School of Business Research Paper No. 17-8*.
- BAKSHY, E., MESSING, S. & ADAMIC, L. A. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348, 1130-1132.
- BENKLER, Y. 2006. *The Wealth of Networks. How Social Production Transforms Markets and Freedom*, New Haven, Yale University Press.
- DILLIPLANE, S., GOLDMAN, S. K. & MUTZ, D. C. 2013. Televised Exposure to Politics: New Measures for a Fragmented Media Environment. *American Journal of Political Science*, 57, 236-248.
- FLAXMAN, S., GOEL, S. & RAO, J. M. 2016. Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80, 298-320.
- GENTZKOW, M. & SHAPIRO, J. M. 2011. Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126, 1799-1839.
- GIL DE ZÚÑIGA, H., WEEKS, B. & ARDEVOL-ABREU, A. 2017. Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics. *Journal of Computer-Mediated Communication*, 22, 105-123.
- GRINBERG, N., JOSEPH, K., FRIEDLAND, L., SWIRE-THOMPSON, B. & LAZER, D. 2019. Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363, 374-378.
- GUESS, A. 2015. Measure for Measure: An Experimental Test of Online Political Media Exposure. *Political Analysis*, 23, 59-75.
- GUESS, A., MUNGER, K., NAGLER, J. & TUCKER, J. 2019. How Accurate Are Survey Responses on Social Media and Politics? *Political Communication*, 36, 241-258.
- GUESS, A., NYHAN, B. & REIFLER, J. 2020. Exposure to Untrustworthy Websites in the 2016 U.S. Election. *Nature Human Behaviour*, 4, 472-480.
- HINDMAN, M. S. 2009. *The Myth of Digital Democracy*, Princeton, NJ, Princeton University Press.
- JÜRGENS, P., STARK, B. & MAGIN, M. 2020. Two Half-Truths Make a Whole? On Bias in Self-Reports and Tracking Data. *Social Science Computer Review*, 38, 600-615.
- KONITZER, T., ALLEN, J., ECKMAN, S., HOWLAND, B., MOBIUS, M., ROTHSCCHILD, D. & WATTS, D. J. 2021. Comparing Estimates of News Consumption from Survey and Passively Collected Behavioral Data. *Public Opinion Quarterly*, 85, 347-370.
- LAZER, D., PENTLAND, A., WATTS, D. J., ARAL, S., ATHEY, S., CONTRACTOR, N., FREELON, D., GONZALEZ-BAILON, S., KING, G., MARGETTS, H., NELSON, A., SALGANIK, M. J., STROHMAIER, M., VESPIGNANI, A. & WAGNER, C. 2020. Computational social science: Obstacles and opportunities. *Science*, 369, 1060-1062.
- MUTZ, D. C. & YOUNG, L. 2011. Communication and Public Opinion: Plus Ça Change? *Public Opinion Quarterly*, 75, 1018-1044.

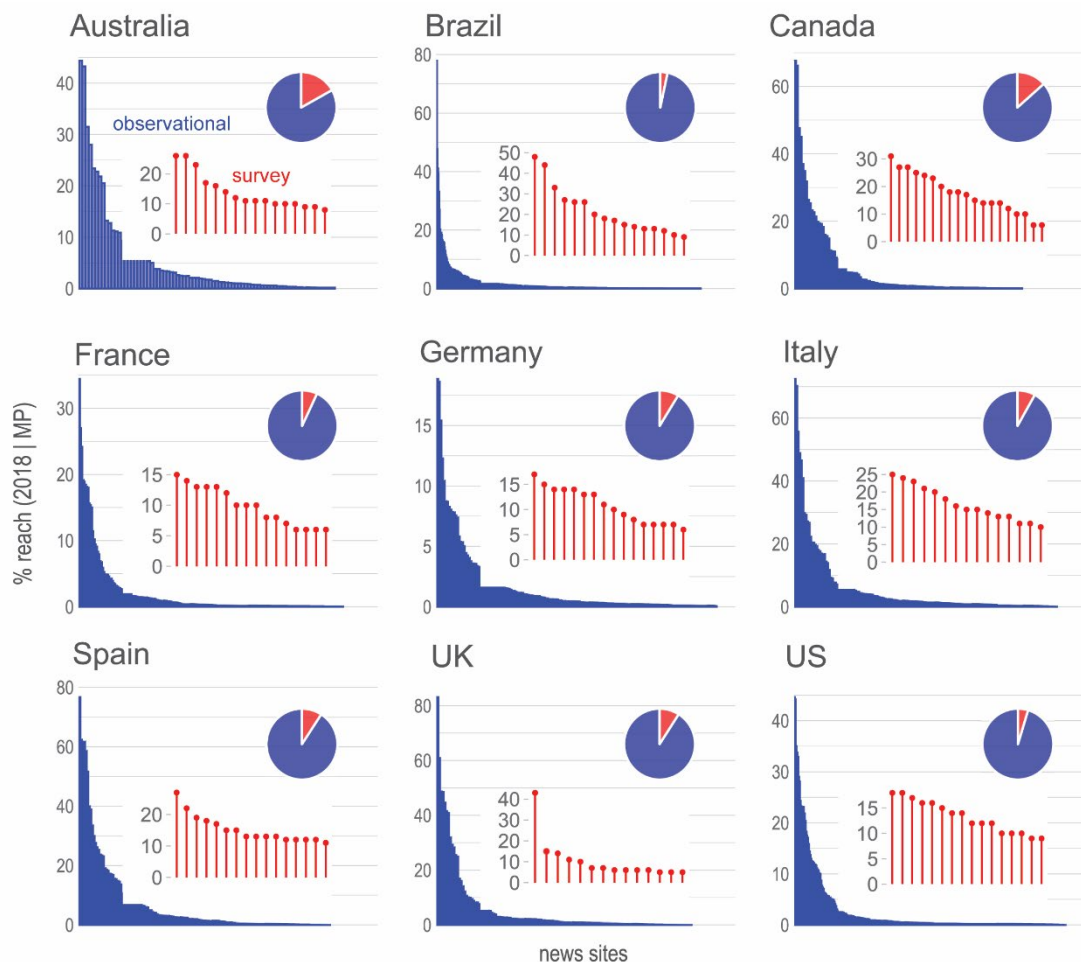
- NEWMAN, N., FLETCHER, R., KALOGEROPOULOS, A. & NIELSEN, R. K. 2019. Digital News Report. *Reuters Institute for the Study of Journalism*.
- OHME, J., ALBAEK, E. & H. DE VREESE, C. 2016. Exposure Research Going Mobile: A Smartphone-Based Measurement of Media Exposure to Political Information in a Convergent Media Environment. *Communication Methods and Measures*, 10, 135-148.
- PARRY, D. A., DAVIDSON, B. I., SEWALL, C. J. R., FISHER, J. T., MIECZKOWSKI, H. & QUINTANA, D. S. 2021. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*.
- PEW RESEARCH CENTER 2020. Measuring news consumption in a digital era. *Pew Research Center's Journalism Project*.
- PRIOR, M. 2009a. The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, 73, 130-143.
- PRIOR, M. 2009b. Improving Media Effects Research through Better Measurement of News Exposure. *The Journal of Politics*, 71, 893-908.
- PRIOR, M. 2013a. The Challenge of Measuring Media Exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication*, 30, 620-634.
- PRIOR, M. 2013b. Media and Political Polarization. *Annual Review of Political Science*, 16, 101-127.
- REEVES, B., RAM, N., ROBINSON, T. N., CUMMINGS, J. J., GILES, C. L., PAN, J., CHIATTI, A., CHO, M., ROEHRICK, K., YANG, X., GAGNEJA, A., BRINBERG, M., MUISE, D., LU, Y., LUO, M., FITZGERALD, A. & YEYKELIS, L. 2019. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Human-Computer Interaction*, 1-52.
- REVILLA, M., OCHOA, C. & LOEWE, G. 2016. Using Passive Data From a Meter to Complement Survey Data in Order to Study Online Behavior. *Social Science Computer Review*, 35, 521-536.
- SCHARKOW, M. 2016. The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data. *Communication Methods and Measures*, 10, 13-27.
- SCHARKOW, M., MANGOLD, F., STIER, S. & BREUER, J. 2020. How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences*, 201918279.
- VOSOUGHI, S., ROY, D. & ARAL, S. 2018. The spread of true and false news online. *Science*, 359, 1146-1151.
- YANG, T., MAJÓ-VÁZQUEZ, S., NIELSEN, R. K. & GONZÁLEZ-BAILÓN, S. 2020a. Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*, 117, 28678-28683.
- YANG, T., MAJÓ-VÁZQUEZ, S., NIELSEN, R. K. & GONZÁLEZ-BAILÓN, S. 2020b. Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*, 202006089.
- ZHURAVSKAYA, E., PETROVA, M. & ENIKOLOPOV, R. 2019. Political Effects of the Internet and Social Media. *Annual Review of Economics*, forthcoming.

Figure 1. Estimation of Audience Reach



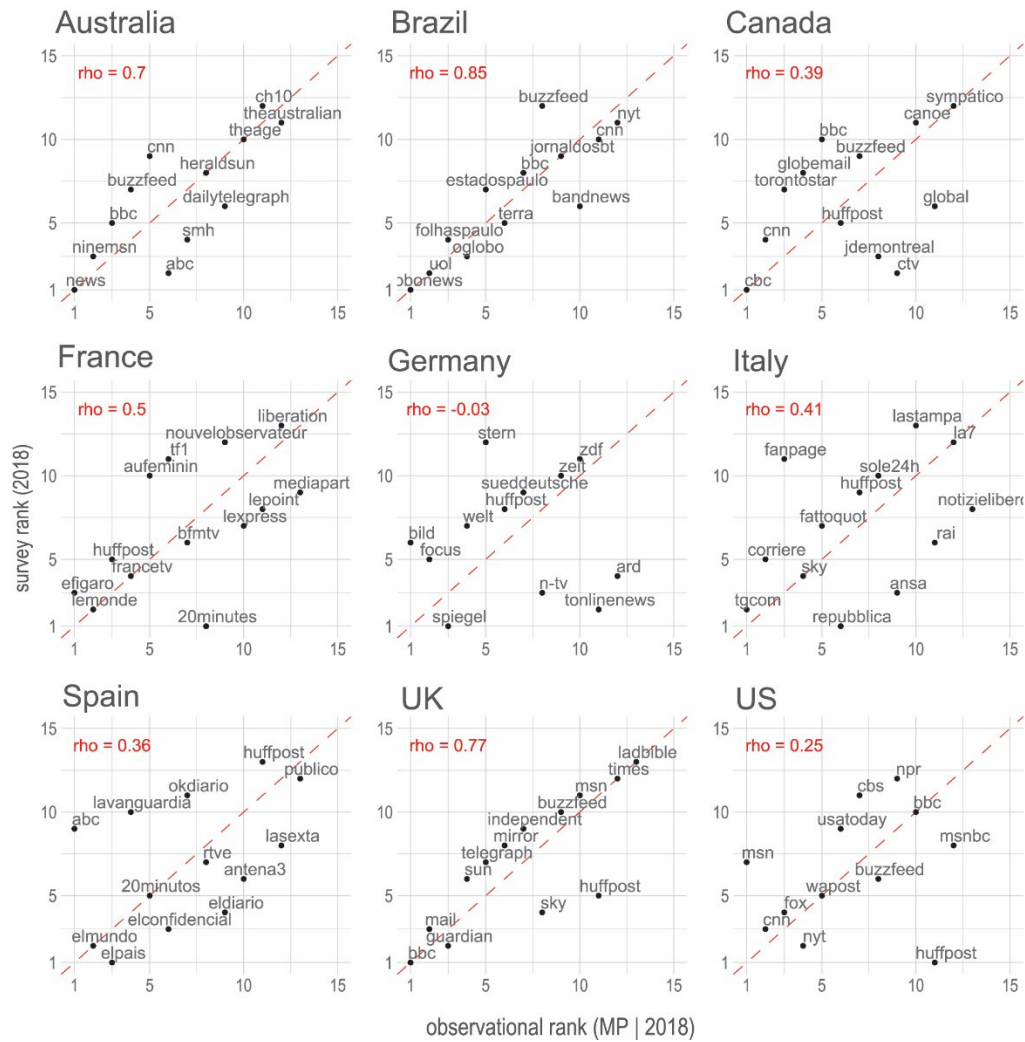
Note: red dots measure the 2018 percentage reach of specific news outlets (listed as vertical labels) according to the observational data (blue dots) and the survey data (red dots). There is a visible difference in the estimates for most news outlets in most countries. In particular, there is a clear pattern of underestimation in survey measures.

Figure 2. Selection of News Sources



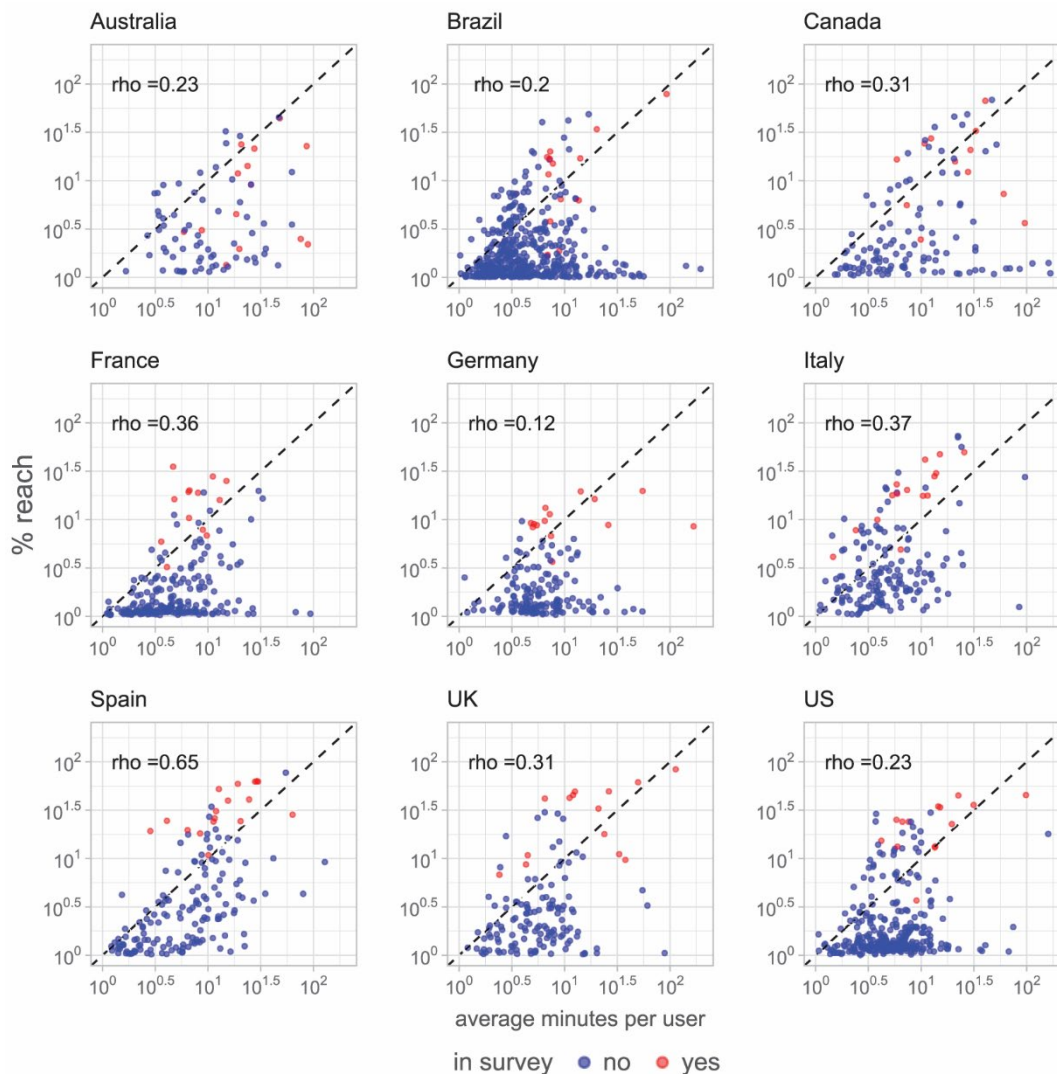
Note: vertical axes measure the percentage reach of news sources in 2018 in each country according to the observational data (blue bars) and the surveys (red bars, inset plots). The pie charts summarize the fraction of news outlets present in the observational data that are included in the survey estimates (red sectors). Three patterns stand out from this figure: audience reach is highly concentrated around a few outlets; surveys only give us information of 8-12% of all sources consumed online (longer lists are not feasible given respondents' fatigue and recall bias); and survey estimates are substantially lower than the observational estimates (e.g., the range of % reach is about half in most countries). Note that we use different y-axis ranges across panels to improve within-country comparisons (which would be more difficult if we fixed the limits to the same values across).

Figure 3. Relative Saliency of News Sources



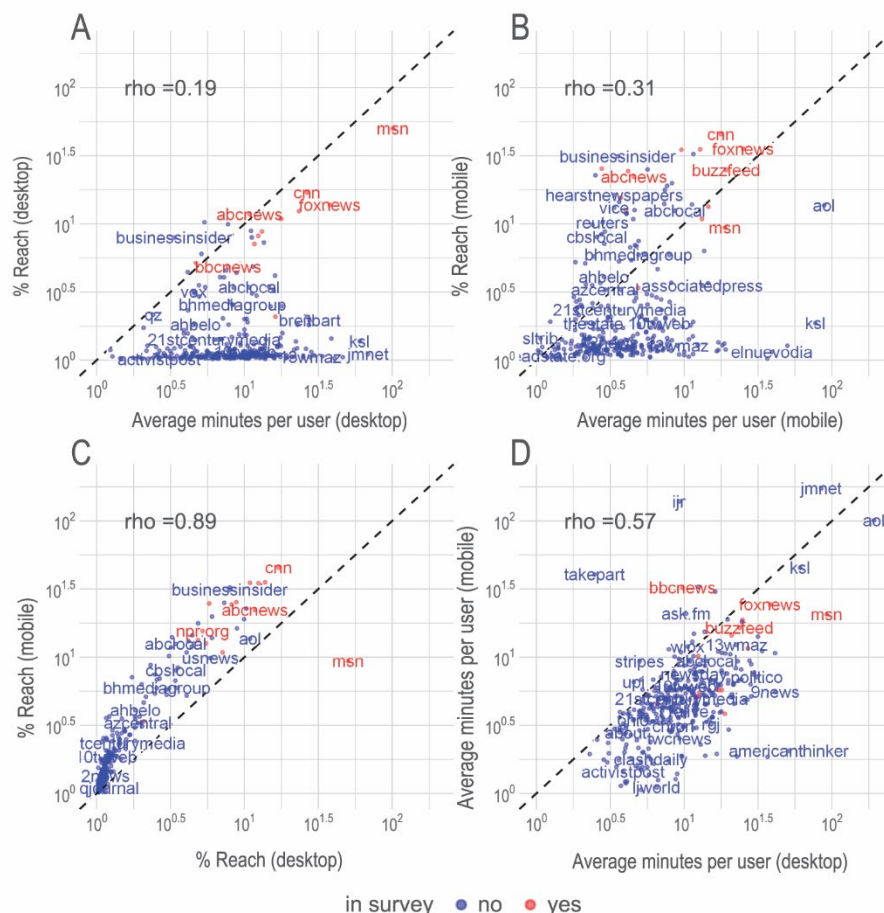
Note: the vertical axes tracks the 2018 rank position of news outlets in terms of visibility according to the surveys; the horizontal axes tracks the 2018 rank position according to the observational data. The farther the data points are from the diagonal line, the higher the disagreement in the two estimates. For Germany, the US, Spain, Canada, Italy, and France, the correlation coefficients are all equal or below 0.5. Australia, the UK, and Brazil have correlations that range between 0.7 and 0.8.

Figure 4. Relationship between the Reach of News Sites and Average Time Spent by Visitors



Note: the vertical axes measure the percentage reach of all news sites in the observational data; the horizontal axes measure the average number of minutes spent on the website, per visitor (2018 data). Red dots identify the news sources that are part of the DNR. As the plots show, surveys tend to focus on the outlets with a larger audience base, but not necessarily with the most engaged base. As the correlation coefficients show, there is a significant lack of agreement between these two measures of exposure: news websites that reach the largest number of people are not necessarily those in which audiences spend more time. Regional and local outlets, as well as niche outlets (e.g., Breitbart, RealClearPolitics), are in the category of sources with low reach but highly engaged audiences (in terms of time spent on their sites).

Figure 5. Breakdown of Reach and Time Spent by Device of Access



Note: observational data for the US, aggregated for 2018. These scatterplots shed light on the type of activity that takes place at the tail of the news reach distribution. Red data points and labels identify the news sites that appear in the DNR. Panels A and B show the same information displayed in Figure 5, US panel, but this time disaggregated by device used to access the news site (desktop computer in panel A, mobile device in panel B). These two panels show that the average time spent by visitor is only weakly associated with the percentage reach of the news site, especially for desktop access. Panel C confirms that the percentage reach of news sites is largely similar regardless of the device used to access those sites. Panel D shows more discrepancy in time spent: people using desktops spend more minutes on news sites, on average, than when they use mobile devices to access those same sites. Niche sites like Breitbart are comparable, in this regard, to much larger news outlets (in terms of audience reach), like CNN.

The Blind Spots of Measuring Online News Exposure: A Comparison of Self-Reported and Observational Data in Nine Countries

Appendix

A1. Data

Our observational analyses use data as provided by Comscore through their Media Metrix Multi-Platform Key Measures reports, which include access to web content through desktop (home and work) and mobile (Android, iPhone, and iPad devices). The estimates Comscore provides are based on a combination of panelists' passive tracking (at the individual level) and site traffic (at the device level). Panelists are recruited online to install Comscore's meters on their devices and fill a recruitment questionnaire eliciting basic demographic information. Table A1 summarizes the size of the panels in each of the countries we consider. The panels are designed to be nationally representative: they are weighted to Internet universe estimates using country-specific census data. Site traffic estimates are collected through tagging, a short code that is placed on a website and is used to track events (e.g., page views, mobile app call, etc). Traffic estimates report activity of content visited from all devices, including mobile phones and tablets. Comscore estimates are provide monthly. For most of our analyses, we average those estimates for the year 2018 (and yearly since 2014 for the temporal analyses below).

Our survey data comes from the Reuters Institute Digital News Reports, which rely on online questionnaires fielded yearly by YouGov. The estimates are representative of the online population in each country, and they are also weighted to targets (population with internet access) based on national census data. The question asked is "Which of the following outlets have you used to access news online in the last week (via websites, apps, social media, and other

forms of internet access)?” Respondents in each country select the outlets from a list of about 30 of the most popular outlets. In the end, the reports publish estimates for about 14 of these outlets, which offers the data we analyze here. Table A2 shows the sample sizes for the five years we consider.

Table A1. Panel Size for Observational Data

	Desktop	Mobile ³
Australia	8,000 ¹	NA
Brazil	60,000 ¹	4,800
Canada	30,000 ¹	7,000
France	20,000 ¹	3,500
Germany	24,000 ¹	3,500
Italy	18,000 ¹	5,000
Spain	25,000 ¹	10,000
UK	70,000 ¹	12,000
US	265,000 ²	35,500

Note: numbers provided by Comscore.

¹PC, ²PC and MAC; ³Android phone and tablet, iPhone and iPad.

Table A2. Sample Size for Survey Data

	2014	2015	2016	2017	2018
Australia	NA	2,164	2,021	2,004	2,026
Brazil	1,037	2,091	2,001	2,003	2,007
Canada	NA	NA	2,011	2,000	2,022
France	2,039	2,131	2,162	2,000	2,006
Germany	2,116	2,035	2,035	2,062	2,038
Italy	2,041	2,059	2,195	2,011	2,040
Spain	2,082	2,127	2,104	2,006	2,023
UK	2,271	2,313	2,024	2,112	2,117
US	2,384	2,588	2,197	2,269	2,401

Note: numbers provided in the Digital News Reports.

Both the survey and observational datasets are weighted to be representative of the online population in each country and they both estimate the audience reach of the same news outlets. These datasets do not allow us to explore bias at the individual level but at the domain level. Our question is not “what types of users are more likely to give biased answers?” but “which outlets attract more traffic?”

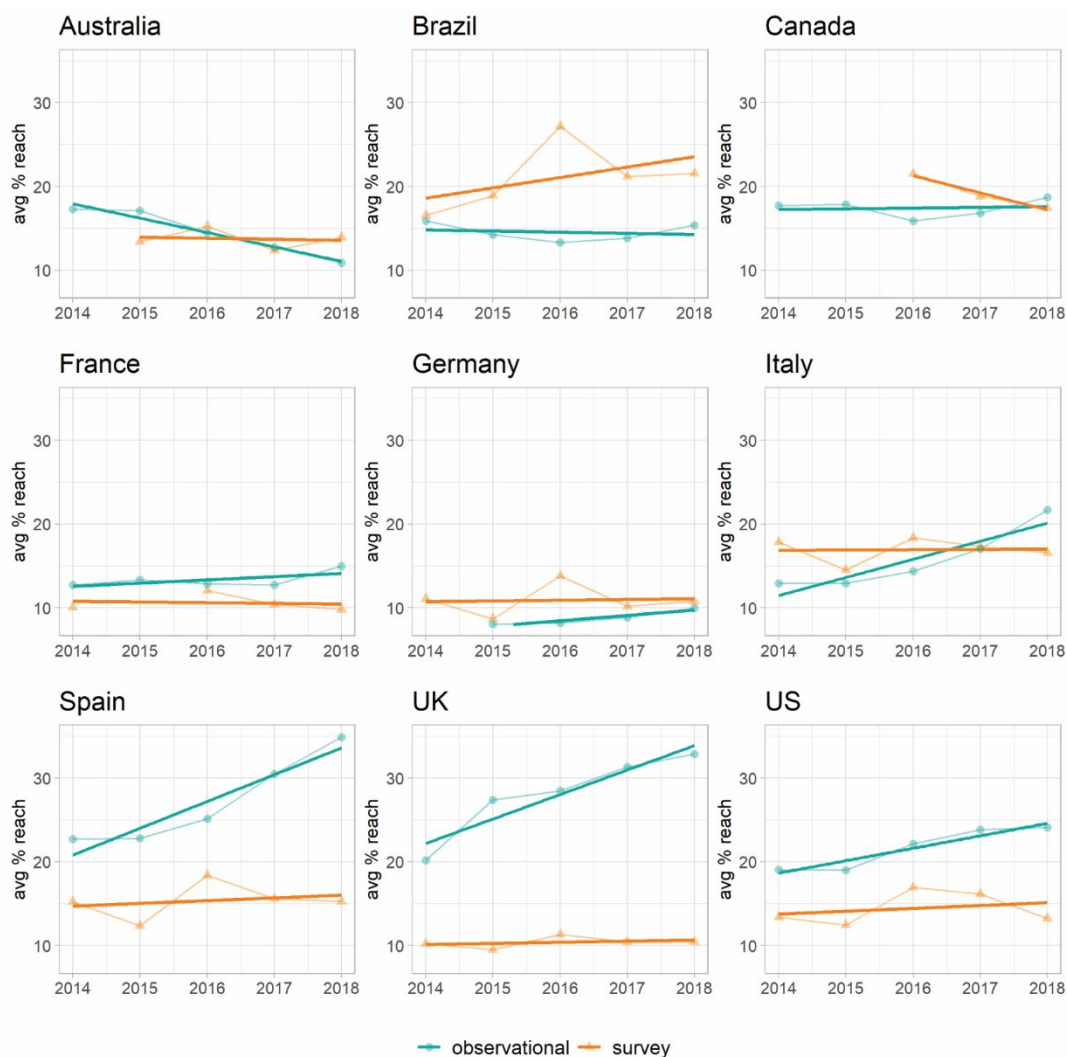
A2. Identification of trends

When examining temporal patterns in the reach of individual outlets, we also find disagreements between the list-based survey data and observational data. Figure A1 averages by year the reach of the news sources that appear in the surveys and their reach according to the tracking data. The vertical axes measure percentage reach, and the horizontal axes track time. As the panels show, in half of the countries the survey data persistently underestimates reach and mischaracterizes what, according to the observational data, are increasing trends. This means that, in most cases, there are some serious discrepancies not just in the size of the audience but in how the audience base evolves over time. For the subset of outlets that appear in the survey estimates, we see substantial disagreement between the two types of data in terms of whether they are gaining or losing reach. As we show below, the assessment of these temporal patterns is very different when we use desktop-only estimates (see Figure A5).

A3. Desktop vs Multi-Platform Access

The core of our argument is that surveys under-represent exposure to news. Yet observational data that only considers desktop access also under-represents access to news. This is not surprising given the fast rise of mobile access in recent years. As we show in Figure A2, the reach distribution for the news sites included in our observational data shifts to the right when mobile access is considered (note that the horizontal axes are log-transformed). However, most past research using observational data is based solely on data collected through desktop access – which explains why some of our findings might seem counter-intuitive when compared to past work. As we show in the figures below many of the trends we identify in the main text get reversed if we only consider desktop estimates. In the case of temporal patterns, the observational data shows decreasing trends across all countries (contrast figure A5 with figure A1) – which is not surprising given that most online news consumption has migrated to mobile devices. This is a particularly relevant finding since most prior research using observational data employs desktop-only estimates.

Figure A1. Longitudinal Changes



Note: each panel compares the average percentage reach of outlets according to the survey data and the observational data in each corresponding country for the period 2014-2018. Surveys persistently underestimate reach in four countries (France, Spain, UK, and US) and they fail to capture increasing trends in two other countries (Canada and Italy).

Figure A2. Reach Distribution of News Sites according to Desktop and Multi-Platform Estimates

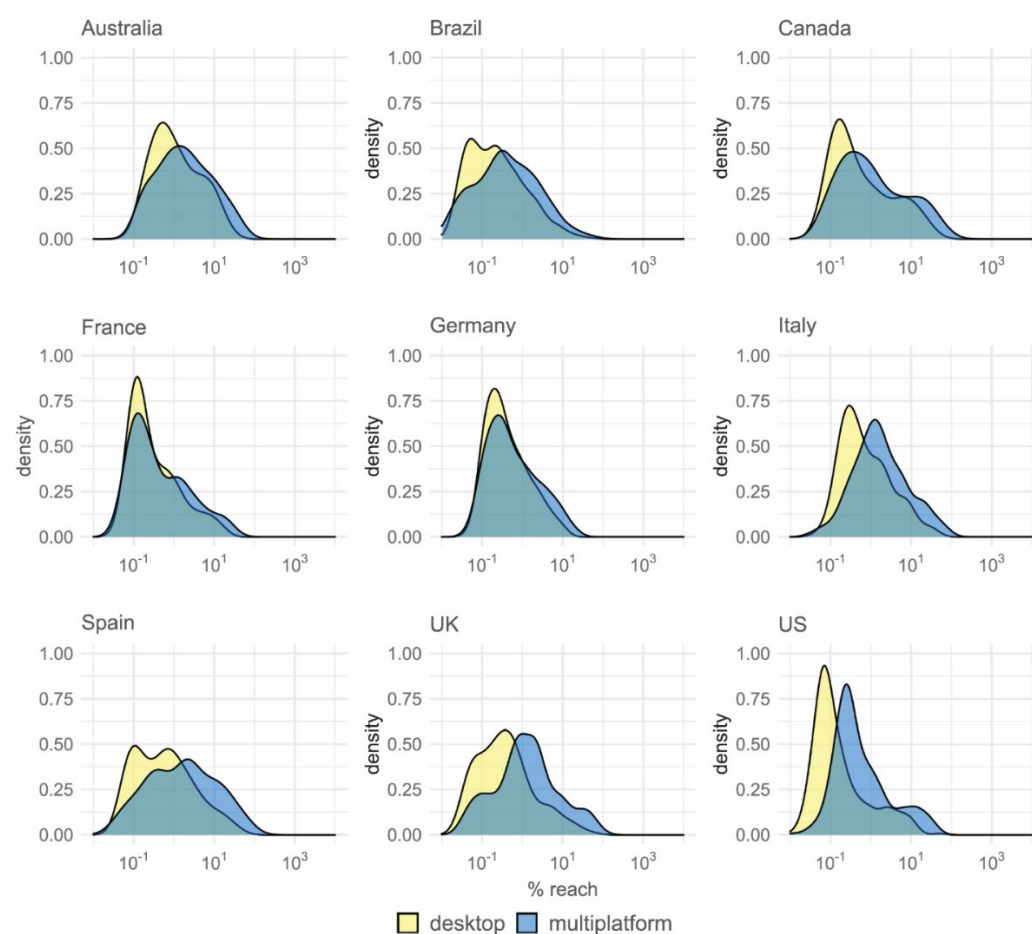


Figure A3. Estimation of Audience Reach (Including Desktop-Only Estimates)

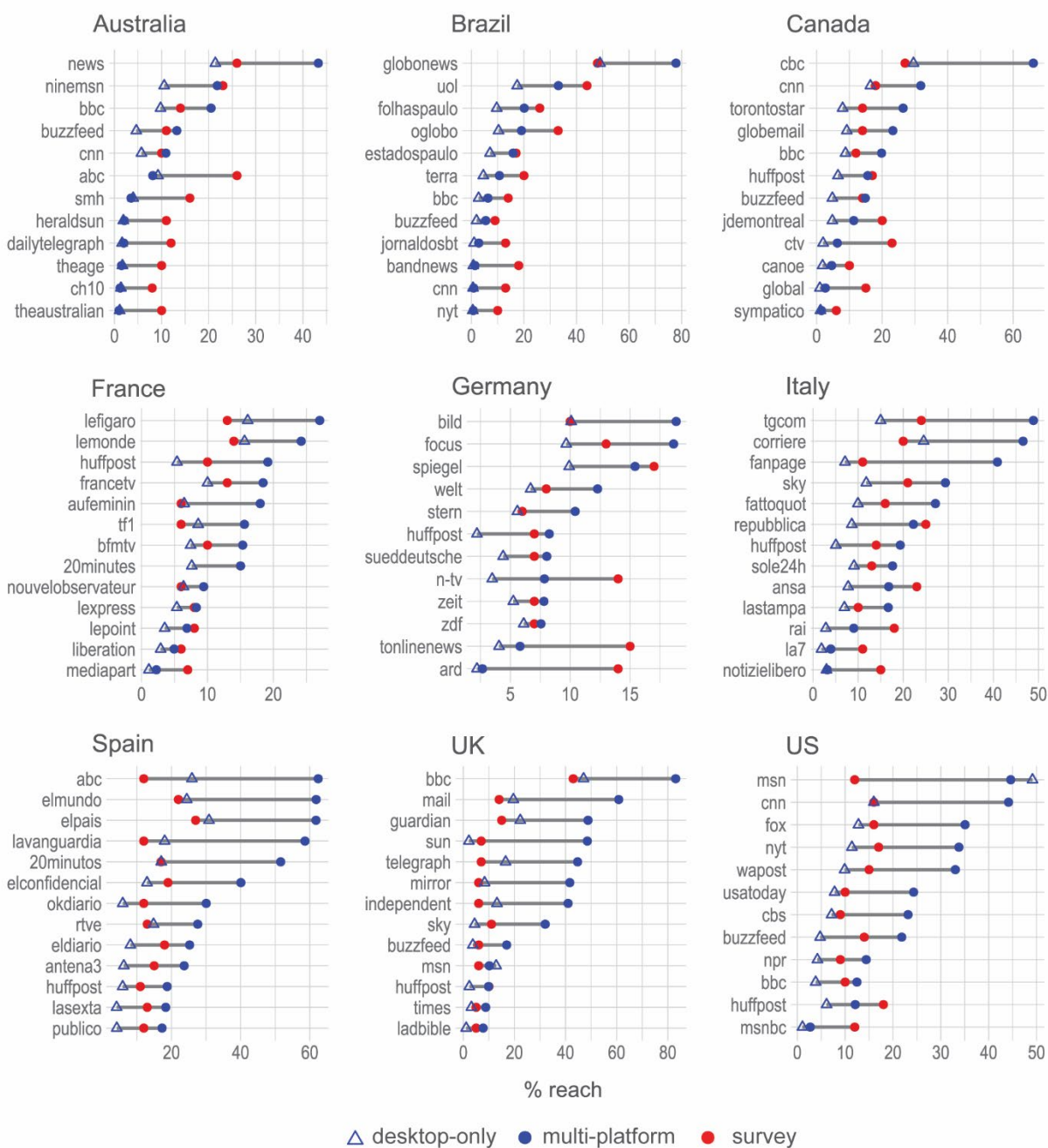


Figure A4. Relative Saliency of News Sources (for Desktop-Only Estimates)

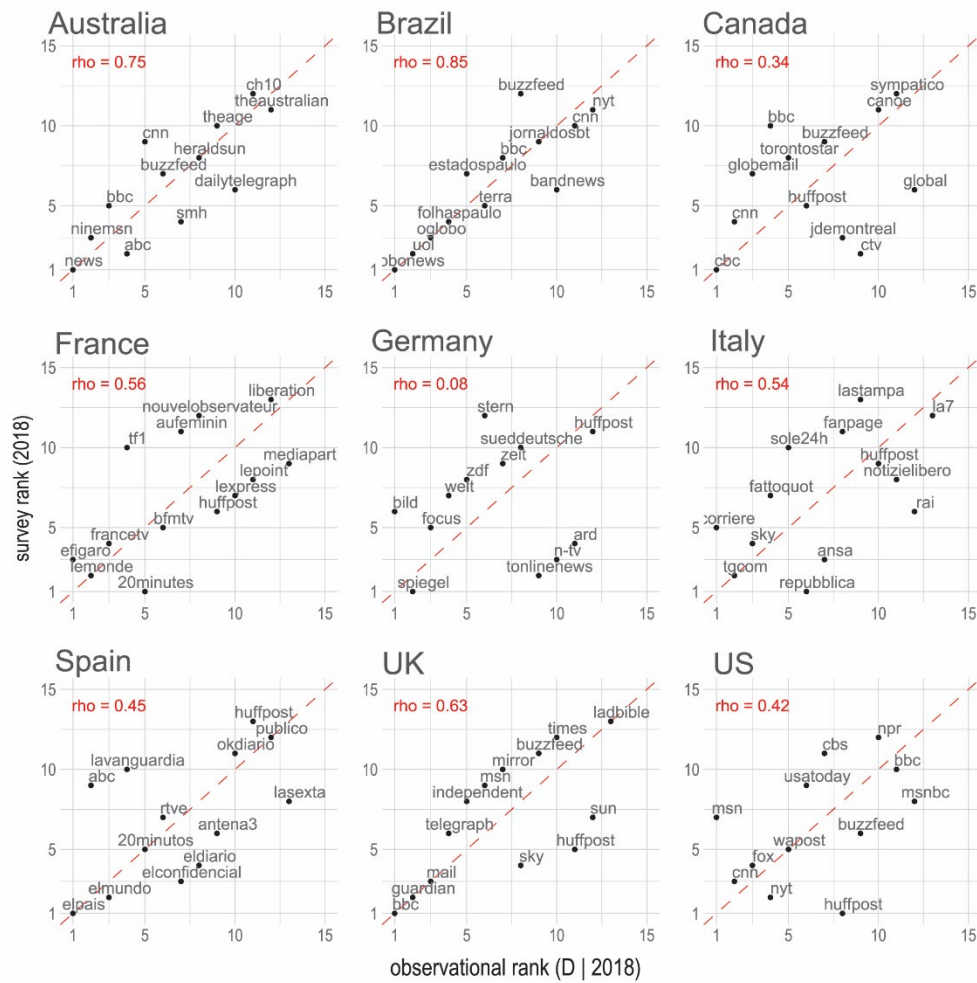


Figure A5. Longitudinal Changes (with Desktop-Only Estimates)

