

The use of poll aggregators in political journalism during Brazilian 2022 presidential campaign

O uso de agregadores de pesquisas no jornalismo político durante a campanha presidencial brasileira de 2022

Bruno Nogueira Galvão Pereira¹

ngpbruno@gmail.com

<https://orcid.org/0009-0000-3274-8572>

Arthur Guimarães de Oliveira²

arthurgdeoliveira.42@gmail.com

<https://orcid.org/0009-0004-4722-0905>

https://doi.org/10.14195/2183-6019_17_4

Abstract:

The present research aims to understand how Brazilian media outlets use electoral forecasting methods in journalistic articles within the scope of coverage of the 2022 presidential elections. The year 2022 was remarkable due to the launch of several poll aggregation tools by the press, which have been used as an alternative to traditional polls to reinforce the demanding electoral coverage. In this context, we collected data from journalistic articles of selected media outlets that cite their own poll aggregators from January 1, 2022, to July 31, 2022. The research maps the utilization of aggregators in newspapers and presents an analysis of the articles that use the poll aggregators. To do so, we classified the articles into three distinct categories, identified according to patterns found in the way they use the aggregator. Finally, it underlines the innovative character of the research on the use of poll aggregators by media outlets.

Keywords: Election forecasting; opinion polls; poll aggregator; political journalism; Brazilian elections.

Resumo

Este artigo visa compreender como a imprensa brasileira utilizou métodos de predição eleitoral em matérias jornalísticas durante a cobertura da eleição presidencial de 2022. O ano de 2022 foi marcado pelo lançamento de diversos agregadores de pesquisas de intenção de voto por parte da imprensa, que foram usados como alternativa às pesquisas tradicionais para reforçar a demanda pela cobertura eleitoral. Nesse contexto, coletamos dados apenas dos veículos que possuem e citam seu agregador em suas matérias jornalísticas entre 1 de janeiro de 2022 até 31 de julho de 2022. A pesquisa mapeia a utilização dos agregadores de pesquisa nos jornais e apresenta uma análise das matérias que utilizam os agregadores. Para isso, classificamos as matérias em três categorias distintas, identificadas de acordo com o padrão encontrado no uso do agregador. Por fim, reforçamos o

caráter inovador da pesquisa sobre o uso de agregadores na imprensa.

Palavras-chave: Predição eleitoral; pesquisas de opinião; agregador de pesquisas; jornalismo político; eleições brasileiras.

1 Pós-graduando em Ciência Política pela Fundação Escola de Sociologia e Política de São Paulo (FESPSP) e em Ciência de dados pela FIA Business School.

2 Graduado em Jornalismo pela Faculdade Casper Líbero. Repórter do JOTA.

Introduction

This research aims to analyze the use of election forecasting models as sources to produce journalistic articles by the Brazilian press during the pre-campaign period of the 2022 presidential elections. The main objective is to understand how Brazilian media outlets use methods that aim to anticipate electoral results to increase the volume of reports and differentiate themselves during electoral coverage. In recent years, Brazilian media outlets have begun to launch poll aggregators, statistical models that gather the results of polls on voting intentions. The principle is that while all polls should accurately reflect a “picture” of voter preferences at the moment, they are subject to a variety of errors in practice. Aggregation is a strategy to mitigate some of these imperfections. Random errors, such as those in sampling, and systematic errors, such as methodological biases, can cause poll results to differ from one another. When considering the production of different institutes, whose methodologies are also diverse, aggregators can reduce these errors

by adopting statistical techniques for normalizing the results. Choices for aggregating polls, however, vary. On one hand, it is possible to simply produce a statistical average of all available polls. Another alternative is to do the same but only with the results of polling companies whose quality is more recognized — or to do a weighted average of all polls, with more established companies having a greater weight. Considering the sample size is also a way to improve the estimates produced. Polls with larger samples tend to provide more accurate results than smaller ones. Weighting these results by giving higher weight to polls with larger samples can reduce errors in the aggregate (Pasek, 2015). Poll aggregators can be useful because they can capture changes in voting intentions as the election cycle unfolds. While traditional polls act as “stills” of that moment, aggregators can show the “footage” of the election. They, on the other hand, are dependent on polling companies and are restricted to that specific election, which makes it difficult to carry out an analysis capable of

comparing different elections (El-Dash, 2014).

With that in mind, this study aims to understand how statistical models are appropriated by Brazilian media outlets to guide the production of news articles during the pre-campaign period of the 2022 presidential elections. Our hypothesis is that poll aggregators can influence how news is produced. For this purpose, a literature review was carried out on the field of electoral predictive models, inferential statistics, public opinion studies, and presidential elections in Brazil. A collection of journalistic articles was reunited from media outlets whose aggregators are used at least once as a primary source of information to produce a news article on the elections. The decision was taken because we believe that media outlets that have their own forecasting models are, consequently, at a higher level of sophistication concerning statistical forecasting tools. The media outlets selected were *Poder360*, *Jota*, *Uol*, *O Povo*, *Estadão*, and *CNN Brasil*. The process was carried out through data scraping, with the browser extension

Web Scraper available for Google Chrome and Mozilla Firefox, in addition to the advanced search tool offered by Google. The database is available at: <<https://doi.org/10.5281/zenodo.7275014>>. The time interval for the scraping was from January 1, 2022, to July 31, 2022. With the data in hand, we mapped out the use of poll aggregators in publications of the selected media outlets and analyzed the application of such in journalistic articles. As far as we know, this is the first Brazilian scientific paper on electoral poll aggregators in the press and the first to raise their use in Brazilian media outlets.

Poll Aggregators

Poll aggregators represent a sophisticated method of capturing trends in the electoral scenario in real time, which is a major object of analysis in journalistic coverage of the electoral race (Louwerse & Dijk, 2022). Given its position of importance in journalism, naturally, some studies have investigated the quality of coverage of electoral polls, mainly studying whether it includes

explicit methodological information, for example, on statistical uncertainty (Andersen, 2000; Brettschneider, 2008; Gahner Larsen & Straubinger, 2012; Hardmeier, 1999; Miller & Hurd, 1982; Sonck & Loosveldt, 2008; Welch, 2002, as cited in Bhatti & Pedersen, 2016, p. 1). However, the study on the use of poll aggregation methods by the press is still scarce, proving to be a field that deserves further exploration.

Electoral poll aggregators are statistical models that unify the results of different polls on voting intentions in a single model, with different weights being assigned to certain polls, depending on the selection criteria. Polls by more established polling companies with a history of good results tend to weigh more in the final value of the aggregator than new polling companies, for example. In this regard, it's worth noting that, as a quality control measure, aggregators don't use data from all polling companies, but only from those with recognized integrity. Another possible criterion to be used is the survey methodology — in person or by telephone, stimulated or not, etc.

A moving average of the results in a given time window can also be applied, so that more recent polls have greater relevance. In general, the use of poll aggregators by the press comes at a time when different polls offer significant variance in their results, something that did not go unnoticed in their reporting (Marcelino, 2022). The advantage of using poll aggregators over an individual poll is that the aggregator can provide a more reliable result. "Combining estimates from multiple samples is in many ways the equivalent of collecting a larger sample. Larger samples, in turn, suffer from less sampling error than smaller ones" (Pasek, 2015, p. 597). In relation to other forecasting models, El-Dash (2014, p.1) highlights the dynamism of aggregation models as an advantage, since they "are capable of capturing all changes in voting intentions that occur during an electoral cycle". However, the dependence on being constantly fed with data from new polls to maintain their accuracy makes them incapable of, like structural forecasting models, "using historical and conjunctural information about the elections,

capturing trends that are repeated in different electoral cycles” (El-Dash, 2014, p. 1).

The use of poll aggregators by the media outlets is justified by themselves as a way of “contributing to a more accurate analysis of polls on electoral disputes” (Bramatti, 2022), through daily tracking of the electoral tendencies of candidates in polls and the treatment of biases in methodological aspects (*house effects*) of each polling company. Polls are subject to biases (Jackman, 2005), and in particular to biases specific to certain polling companies, known as “house effects”, which concern biases arising from the methodology employed by the institute conducting the survey. The way the interview is conducted (face-to-face, telephone, or internet), sampling and weighting procedures, day of the week or time of day a company interviews, age, ethnic, and gender composition of interviewers at a given company, the wording, and order of questions are all potential sources of bias in a poll. “The chief benefit of pooling poll results (after correcting for house effects) is that we are much better

positioned to ascertain movements in levels of voter support in response to campaign events” (Jackman, 2005, p. 501).

The aggregation of polls alone would not be efficient. This is because, with house effects, sampling variation is no longer the only source of error between different polls. The bias present in two polls can be such that it would be incorrect to consider them as estimators of the same population quantity and, therefore, their grouping would be invalid. The effect of a methodological bias is, in practice, to underestimate or overestimate a candidate’s performance. To contain them, the media outlet assigned a weighting system to each polling company, estimating how much a poll from a given institute is above or below the market average. In the case of electoral polls, there is a clear benchmark to assess the bias of each institute: the election result itself. By crossing data from previous elections, it is possible to evaluate the degree of bias in relation to competitors.

By pooling and smoothing the polls, we obtain a more precise

estimate of underlying vote intentions than can be formed from any single poll. By constraining the estimated trajectory of daily vote intentions to culminate with the known election result, it is possible to estimate bias parameters for each polling organization, and, in turn, to recover day-by-day estimates of vote intentions that are purged of the biases afflicting any one polling organization (Jackman, 2005, p. 515).

To deal with some of the main problems presented in polls (such as inaccuracy due to sampling errors, house effects, and blurred vision of the daily tracking of support for candidates and electoral trends due to the multiplicity of different polls pointing to different evolutions and of fluctuating periodicity), Jackman (2005, p. 508) proposes a model to aggregate polls in order to increase precision, estimate and adjust the methodological bias of any company and track trends and fluctuations in voter sentiment throughout the electoral campaign, described in the following Bayesian model:

$$y_i \sim N(\mu_i, \sigma_i^2),$$

Where, in an n number of polls i used in the aggregator, $i = 1$ until n , with the poll i made by the institute j_i on day t_j , y_i is the result of the poll i and has the probability distribution of N , σ_i^2 is the accuracy of the i -th poll, μ_i is the statistical average term of the poll, assuming that $\mu_i = \alpha_{t_i} + \delta_{j_i}$, where α_{t_i} is the voting intentions on day t_j and δ_{j_i} the effect of the polling company j_i that made the poll. “Incorporating the effect of the institute that carried out the research is one of the most important characteristics of this model, as the institutes are known to use different methodologies which clearly have an impact on the estimates”, reinforces El-Dash (2014, p.2) on the measure that deals with the house effects. This model acts as a Kalman Filter, a method “used in signal processing in engineering applications, for tracking a moving target with noisy (and possibly biased) observations.” (Jackman, 2005, p.509). The algorithm is usually applied in dynamic systems with uncertain measurements, as it manages to estimate the real values amid the noise. In the case of poll aggregators, it is useful for estimating values in the historical series when

there is a large interval without polls between one date and another.

It is worth making a disclaimer that the aggregation model proposed by Jackman (2005) was not designed to make forecasting, and it is discussed in the bibliography (Pasek, 2015) that aggregators are not intended to predict the electoral result, but rather to estimate the current context of the electoral race. The aggregator, after all, is not based on a “voting theory”, such as the “political economy” models (Lewis-Back, 2005), which seek to forecast the electoral result through explanatory variables — while aggregators merely normalize the numbers of the polls, which measure the moment, and not future estimates.

In short, it is argued that the use of poll aggregators collaborates for a more accurate “picture of the current political scenario” (as journalists often say), which can help to dispel the fog that surrounds this electoral poll market — with so many polling organizations publishing results weekly, increasing the frequency the closer to election day, the variation between the different results

can cause the opposite effect of what was expected: instead of helping the public to understand the scenario and choose their vote, the difference between the results can cause a noise that confuses the voter and, ultimately, generates distrust regarding the polls carried out. With the press so closely associated with polls (after all, it is the press that plays the role of reporting data to the public), a sense of public disbelief in polls could also undermine confidence in journalism, which is by itself not trusted by many people¹. The aggregators, therefore, can serve in two ways: to make the poll results safer from the statistical treatment of biases and, consequently, to bring more credibility to the polling companies and media outlets that publish it. At a time when the results of individual polls are increasingly questioned by ideological passions, aggregators emerge to reinforce the credibility of election forecasting methods before the general public and, in turn, can

¹ According to a survey conducted by the Reuters Institute and the University of Oxford, less than 50% of Brazilians have trust in journalism (Toledo, 2022).

serve as a source of authority for the media outlets that use it.

As Bhatti and Pedersen (2016) contend, the use of poll aggregators can significantly enhance the quality of electoral poll reporting. However, this improvement hinges on adequately addressing statistical uncertainty. Poorly grounded poll news may arise from journalists' lack of statistical understanding, as prior studies have highlighted journalists' numerical aversion and limited mathematical skills. However, another hypothesis suggests that such superficial reporting might stem from journalists and editors' desire for exciting "horse race" election narratives. "The issue might not only be the absence of methodological skills but also a genuine disregard for journalistic norms of reliability and scrutiny," (Bhatti and Pedersen, 2016, p. 9). Echoing this sentiment, Louwerse and Dijk (2022, p. 21) state that "our findings indicate that merely having a poll aggregator doesn't necessarily improve news quality; it's how data are presented by researchers (and poll aggregators) and the publication's organization that truly matters." An illustration is the

Dutch aggregator Peilingwijzer, which reports results based on uncertainty intervals, such as "party x is likely to get 11 to 15 seats" rather than giving a seat estimate as traditionally reported in Dutch media. Applying this logic to the Brazilian presidential election would involve reporting the candidate's vote intention range over time, rather than just the average vote intention from aggregated polls.

In terms of journalistic epistemology, we hypothesize that forecasting models influence the way news about elections is produced. Traditionally, newspapers tend to cover the moment of the electoral campaign through the perspective of the "horse race" (Welch, 2002) between the candidates, addressing the numbers of who goes up or down in polls. With the use of statistical methods such as the poll aggregator, there may be a tendency to produce content generated through its own sources — considering the poll aggregator as an internal production of the media outlet, even if it depends on data from external polls. This generates a potential for newsworthiness that deviates from the standard of reporting poll results,

since the publication of new poll results may cease to be "news", as aggregators track the fluctuation of candidates in real-time and indicate a future trend, something that polls alone cannot do unless they are made daily. The daily tracking carried out by the aggregators allows for newsworthiness criteria aimed at the technical analysis of the electoral trend. But even if a potential is pointed out, there are contrary discourses within journalists, which place the newsworthiness of individual polls as superior to the aggregated results.

The burden of seeking a weighted average of polls is missing the news. Datafolha was in the news last week precisely because of the result it brought alone. Had its numbers been diluted with others, which did not capture the same voter movement, it would not have been on the lips of the political world, which moved a lot in the face of the prospect of the first round. Nevertheless, there are merits in the joint analysis, as long as it is understood as a data study, not an effective poll. This, obviously,

Table 1. Path for scraping data from aggregators as of release date
Source: Authors

Media outlet	Launch of the aggregator	Path for scraping
Poder360	30/07/2018	https://www.poder360.com.br/tag/pesquisa-eleitoral/
Jota	2018	https://www.google.com/search?q=%22agregador%22+site:https://www.jota.info/&tbs=sbd:1,cd_r:1,cd_min:1/1/2022,cd_max:7/29/2022&tbm=news&sxsrf=ALiCzsagMqMDjd5tMAy7HOES2Xkp-gLSlg:1659203180114&ei=bG7IYvjHBv2I5OUP8MaI-Ak&start=0&sa=N&ved=2ahUKEwj4raTBlaH5AhV9BLkGHXAjAp84ChDy0wN6BAgBED0&biw=1536&bih=746&dpr=1.25
UOL	11/04/2022	https://www.uol.com.br/eleicoes/2022/pesquisas-eleitorais/
O Povo	13/05/2022	https://www.google.com/search?q=%22agregador%22+site:https://www.opovo.com.br/&tbm=news&sxsrf=ALiCzsa2nZsTkQnHnXyo_Hl-YXX9DrYX9w:1659203604860&source=Int&tbs=sbd:1&sa=X&ved=2ahUKEwjK4-iLl6H5AhVDirGHZGGD9YQpwV6BAgBECE&biw=1536&bih=746&dpr=1.25
Estadão	30/05/2022	https://tudo-sobre.estadao.com.br/pesquisa-eleitoral
CNN Brasil	14/06/2022	https://www.cnnbrasil.com.br/all-about/electoral-research/

will not happen, as can be seen in the headlines of Estadão and other media outlets with aggregators, such as the Jota website, where the tools gain the comfortable role of their own poll. Someone else's expensive budget is a mere detail (Mariane, 2022, para.11).

The ombudsman of the newspaper *Folha de S.Paulo*, José Henrique Mariane, not only states that the use of aggregators eliminates the newsworthiness of polls, but also stresses that the tool could not be understood as equivalent to a poll itself — without going into the merits that the numbers of the aggregators depend on the release of new polls to be updated. He also criticizes newspapers that use their aggregators as a source of news (Mariane, 2022). Therefore, investigating the impacts of forecasting models in the press is relevant because, in addition to representing a new investment for the media outlets and, therefore, their effects need to be studied, it can clarify their role in journalism in face of the questions raised about its potential for newsworthiness.

Methodology

When defining the scope, we chose to include in our work only media outlets that have aggregators. The selection criterion is to have at least one report on the electoral dispute where its aggregator served as the primary source of news. The decision was taken because we believe that media outlets that have their own forecasting models are, consequently, at a higher level of sophistication in relation to statistical forecasting tools. With the media outlets defined, we collected all the articles about electoral polls for the presidential race published by them from the launch date of each poll aggregator, as shown in the following table.

The collection of information was done through data scraping with the browser extension *Web Scraper* available for Google Chrome and Mozilla Firefox. As a criterion for selecting the news that would be collected, we chose to look for the articles that were categorized as news about electoral polls, since our intention is to analyze how the press uses forecasting models of the electoral campaign in the production of articles, such category

would fit better than if we had chosen to include all the news categorized in the field of “political coverage”; the news about polls better reflect the scenario of the electoral race and the attempt to predict the result than the news about the backstage of institutional politics and party movements, for example. With this choice, we accessed the website of each newspaper and made the scraper navigate by the *tag* of electoral polls. We understand, however, that this choice puts us in a fragile position since we depend on the content of the site being properly categorized by those responsible, running the risk of losing articles that fit our framework, but that have not been tagged as such in the *tag* we chose to browse. With that in mind, we did a double check using Google's advanced search resources, using the following input:

“agregador” site: <https://www.exemple.com.br/>

With this query, Google returns all results that contain the term in quotes within all indexed pages of the website specified. It's worth

Table 2. Number of articles collected from each vehicle from the launch of the aggregator until 07/31/2022
Source: Authors

<i>Newspaper</i>	Launch of the aggregator	Total articles	Articles that mentions “aggregator”
CNN BRASIL	06/14/2022	17	3
ESTADAO	05/30/2022	47	34
JOTA	08/22/2018	27	25
O POVO	05/13/2022	57	57
PODER 360	07/30/ 2018	210	120
UOL	04/11/2022	96	6
Total		454	245

noting that this search term returns a more restricted data set than the data scraping done by Web Scraper. In scraping within the websites, we collected all the articles about electoral polls, whether they had the term “aggregator” or not. The restriction made by Google, however, will not be an obstacle for most media outlets, since later, in the data processing stage, we will only keep the news that contains aggregators. However, this can generate distorted information for *O Povo* and *Jota*, for which we were unable to carry out the scraping through navigation by the tag, since it was not well structured and did not return articles within the framework or failed to include several that contained “aggregator”, but were not tagged as such. For these two, materials were collected only through Google. Therefore, we will not be able to compare the proportion of articles about electoral polls that include aggregators in relation to the total set for these media outlets.

The articles collection interval was selected from the launch date of each aggregator, as recorded in the previous table, until July 31, 2022. In other

words, in the case of *Uol*, articles were collected in the period between April 11 to the 31st of July. For *Poder 360* and *Jota*, which have aggregators since 2018, the starting point was defined as January 1, 2022. We defined this date to only include news about the 2022 presidential elections.

After the scraping was done, we created a *Google Sheets* with the raw data and another, in copy, where the clean data will be kept for the data treatment. This was done by excluding articles about elections for positions other than the presidential one, in addition to removing news that did not deal with polls of voting intentions, but which for some reason were included in the tag. News from the *Folha de S. Paulo* newspaper on the Uol website was also excluded. The database is freely available at: <<https://doi.org/10.5281/zenodo.7275014>>. After collecting and processing the data, we ended up with a dataset of 454 articles from the six selected media outlets. The numbers can be seen in Table 2. With the information in hand, we analyzed the publication data of the material and the text content, evaluating how

the article uses forecasting models. To do so, we classified the articles into three distinct categories, identified according to patterns found in the way they use the aggregator. We also approach patterns of journalistic production of different media outlets in relation to the particular use of the aggregators of each one, noting a variety of applications.

Results

Of the total of 454 articles collected, the newspaper Poder360 represents most of the production, equivalent to 47.8 percent of the total. The same result is seen in the total of articles that mentions “aggregator” throughout the text where, of the total of 245, the newspaper leads with 120 news — almost half of the sample of articles that mention the tool (48.9 percent). Uol has the second-highest volume of total articles, 96 (21 percent of the sample), but only 6 (2.4 percent of the sample) appear to mention the tool. O Povo adds up to 57 both in total and in the count of articles that mention aggregators, due to its collection having

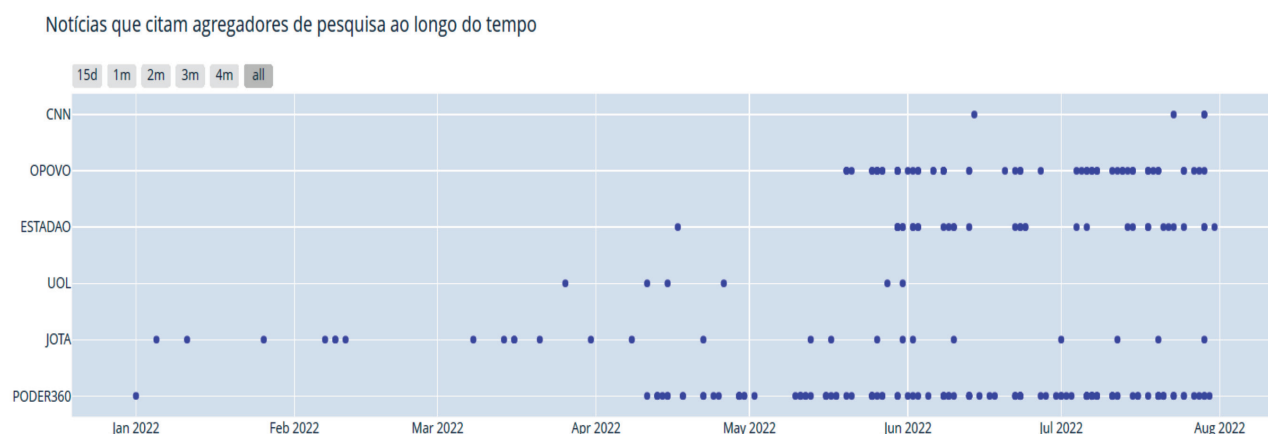


Figure 1. News that cites poll aggregators over time.

Source: Authors

been only for articles that mention the tool, as justified in the methodology section. The same problem occurred with Jota; however, even using filters to search only for news on aggregators, the Google search engine still returned 2 news that did not mention it, leading to a result of 25 out of 27 articles in total. Then, *Estadão* published a total of 47 articles on electoral polls (10.3 percent), of which 34 (13.8 percent) mention their own aggregation tool. Finally, *CNN Brasil* appears, with 17 articles (3.7 percent) and only 3 (1.2 percent) mentioning the aggregator.

Regarding the internal production of newspapers, CNN published 17 articles, of which 3 (17.6 percent) cited aggregators; *Estadão* published 47 news, 34 (72.3 percent) with aggregators; Poder360 had 210 news with 120 using aggregators (57.1 percent); Uol published 96 news and 6 (6.2 percent) mentioned aggregators. O Povo and Jota were not counted due to the alternative data collection process that excluded news without aggregators, making it impossible to compare internal production — even though they were kept in Table 2.

Regarding the frequency of use, we can observe that there is not a universal pattern of news publication between the different websites, as shown in Figure 1, below, which demonstrates the distribution over time of the total number of articles that mentions poll aggregators since their launch until July 31, 2022.

In Figure 1, we can observe different frequencies in the selected period. A first point to note is that, even among those who had an aggregator since 2018 (Jota and Poder360), the distribution of publications shows a considerable difference. Jota has a relatively stable frequency throughout the period, with publications since January. Poder360, on the other hand, has a single isolated piece of news in January and starts publishing more frequently only in May. From this, a first question can be raised: does the proximity of the election increase or reduce the production of news with aggregators? On the other hand, Uol starts with scarce production, has a break of one month between April and May and, until July 31, there was no more production. Most of the news focuses on April, when

the aggregator was incorporated into the website. The article isolated in March, before the launch of the aggregator on the UOL website, deals with the aggregator of the *Polling Data* website, by Neil El-Dash, who, in April, signed a partnership to publish it on the newspaper, thus being the same aggregator. O Povo and *Estadão*, which started in close periods, also have a similar distribution, with high frequency and aggregated volume of news since it was launched. The isolated article of *Estadão* in April, before the release, talks about a projection of the *Média Estadão Dados*, which aggregates the results of several polls in different Brazilian states. Finally, CNN has a very small number to make considerations about its frequency of publications, but it is noteworthy that the first isolated news in June refers to an article announcing the launch of the aggregator. The frequency ratio between the total number of articles on electoral polls and those using the aggregator can be seen in Figure 2, below:

In Figure 2, where the blue dots represent the articles that mention the aggregator and the red dots the ones

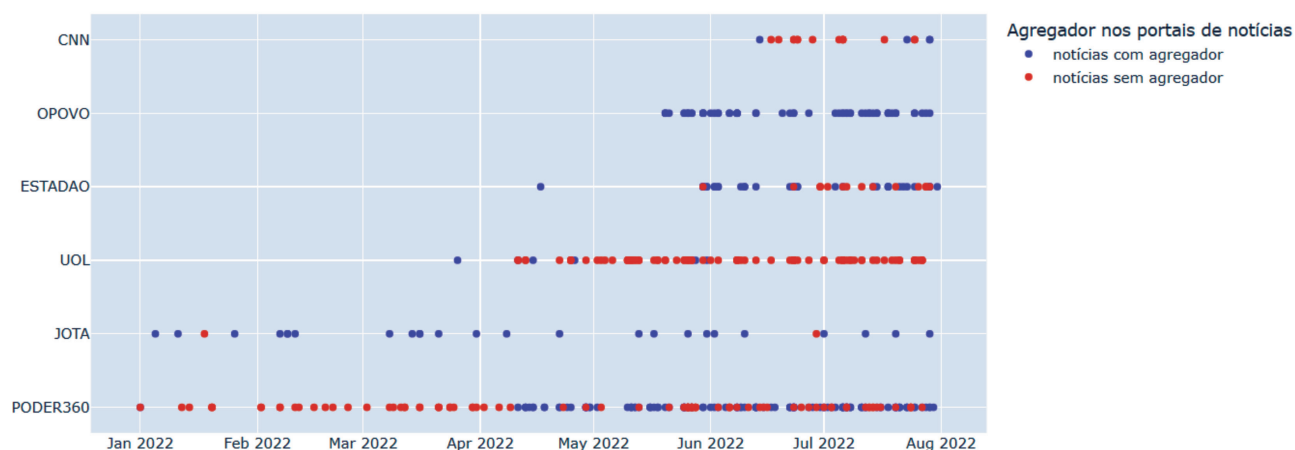


Figure 2. Total articles over time

Source: Authors

that do not, it is possible to see that the frequency of “regular news” is greater than those that use aggregators and, in some cases, in a discrepant manner. The case of UOL shows that there is a systematic production of news about electoral coverage, but the aggregator has not yet been absorbed in the flow. Others, such as Estadão, indicate the opposite scenario: the publishing frequency of the two categories of news described here are similar, with the aggregator even occupying spaces where there is a vacuum of “regular news”. Poder360 assumes a relatively regular pattern from April when it starts using the aggregator in its articles. Between January and April, the vehicle already had the tool but did not use it in any news. Unfortunately, few media outlets had aggregators since the beginning of the year or before that, making it impossible to analyze whether the proximity of election day or the beginning of the official political campaign period would have any correlation with an increase in the use of aggregators in the news — since the data available for this paper goes only until July 31, before the official campaign period,

it can be argued that “little time” was given for the most recent release aggregators to be integrated into the production of articles, indicating an internal operational issue of the media outlets, and not of the criterion newsworthiness itself.

For the qualitative analysis of the use of poll aggregators in media articles, data were categorized according to the way they use the statistical tool. Among the two categories already used previously, of news with or without aggregators mentioned throughout the text, we divided the first one into three subcategories, in order to contemplate the different ways that poll aggregators are applied in the articles. The results can be seen in the figure below:

News categorized under “*uses as primary source*” is, as the name implies, stories in which the newspaper’s aggregator is the news itself; the article is primarily about him. Of the total of 245 news that mention aggregators, 24 (9.7 percent) fit into this category. The second category represents those that use the “*aggregator as data*” to reinforce arguments. The criterion used to fit into this category

was that the article was not about the aggregator, but that it was mentioned to add information or context to the electoral race, showing the number that each candidate has registered in the aggregator, for example. This category has 22 articles (8.9 percent). Finally, the category that “*mentions in a promotional way*” refers to articles that (generally) include excerpts about the aggregator at the end of the text in a standardized way, presenting the tool to the reader, in a manner that is detached from the main content of the article. Links to articles that mention aggregators were also considered in this category, which in all add up to 199 news (81.2 percent). They were called “promotional use” precisely because they did not incorporate the aggregator in the story, but in an isolated way in the content, usually in a separate section of the article, after the end of the story, and it is not uncommon to find the same “sample text” presenting the tool in several different articles. What is being promoted here is the aggregator itself, which is in a separate part of the website, where the newspaper invites the reader to know more about

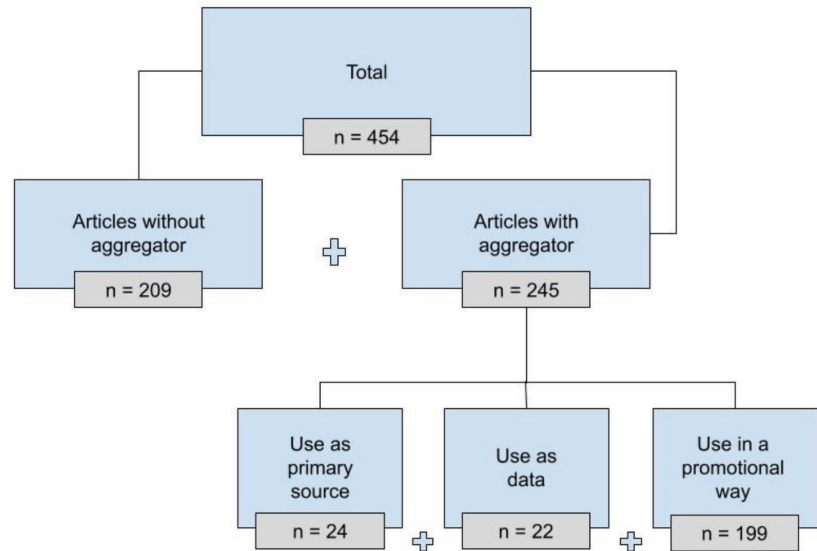


Figure 3.
Categorization of
articles according to
the use of aggregators
Source: Authors

News categorized under “uses as primary source” is, as the name implies, stories in which the newspaper’s aggregator is the news itself; the article is primarily about him

and explore the tool. It can be seen, then, that it is not a use of the tool integrated with journalism for the construction of the news, but a use of the journalistic medium as a showcase for presenting the tool — which, in the case of Jota and O Povo, is a Paid product for subscribers only. A division of articles published by category in the different media outlets, in absolute numbers, can be seen in the following figure:

Figure 4 illustrates the internal proportion of the production of each newspaper based on the three defined categories. CNN Brasil produced 3 news articles with aggregators, 1 (33.3 percent) as a primary source and 2 (66.6 percent) used as data; Estadão produced 47, 7 (14.8 percent) as a primary source, 4 (8.5 percent) as data and 23 (48.9 percent) as promotional; Jota produced 27, of which 6 (24 percent) as a primary source, 9 (36 percent) as data and 10 (40 percent) as promotional; O Povo had 57 articles, 7 (12.2 percent) as a primary source, 5 (8.7 percent) as data and 45 (76.9 percent) as promotional; Poder360 published 120 articles, all of them

promotional; Uol published 6 articles, 3 (50 percent) as a primary source, 2 (33.3 percent) as data and 1 (16.6 percent) as promotional.

Figure 5 demonstrates that Uol was the only website to produce more articles that use an aggregator as a primary source than other categories. The tendency observed in 4 of the 6 media outlets is to have more articles that mention the tool in a promotional way. Another trend pointed out is to use the aggregator more as a primary source than as data: 3 out of 5 media outlets do so. The disparity is not that great, however. The average use of aggregators as the main source is 28 percent, inflated by the expressive numbers of CNN and UOL. Note that the media outlet with the smallest sample in absolute terms in this case has the highest percentage in this category since they are the only two to be above the average. CNN also contributed to inflating the average of the news category they use as data, which reached 31 percent; Estadão, O Povo, and Poder360 were below. The average of news that uses it in a promotional way, 51 percent, was both inflated by Poder360, with 100

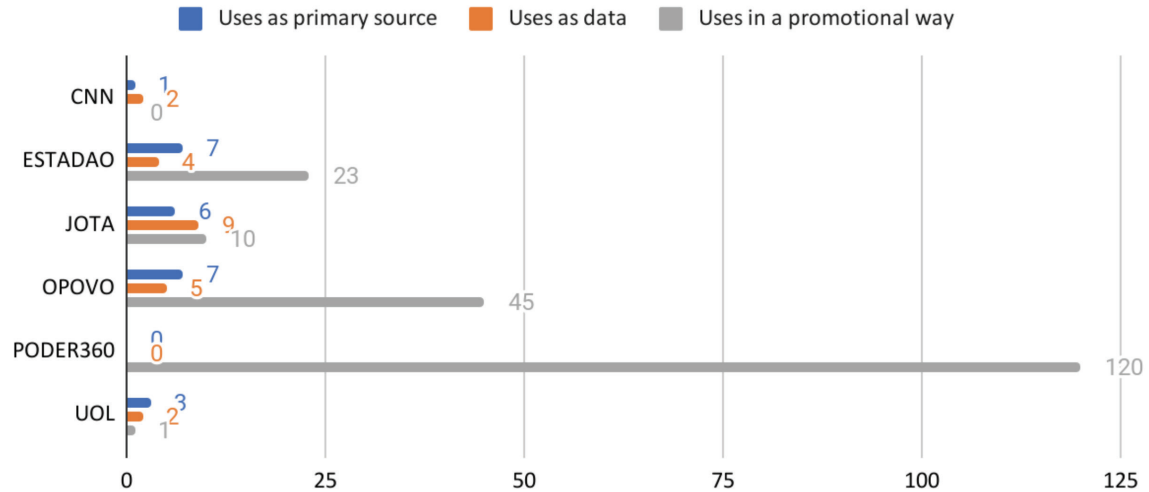


Figure 4. Aggregator usage in each newspaper by category in absolute numbers

Source: Authors

percent, and pulled down by CNN, with 0 percent.

In terms of reporting trends, we found that between news categorized as “poll aggregators as their primary source of information”, they extract and present this information differently. Some report in a manner similar to traditional poll result news, while others focus on different aspects, such as a candidate’s probability of being elected, rather than their estimated vote share. For instance, Jota’s six articles do not feature vote percentages in the headlines. Instead, they emphasize other elements, like the likelihood of a candidate winning in the first round, or the numerical gap between two candidates. Specific titles include phrases like “Chance of election ending in 1st round falls, JOTA’s poll aggregator points out,” and “Lula x Bolsonaro: Know how the voting intention of the two among women is.” Importantly, Jota factors in statistical uncertainty, demonstrated by their headlines that replace vote intention with victory probability. Contrastingly, Estadão headlines prominently feature their poll aggregator, “Média Estadão Dados,” and

include candidates’ vote percentages in most titles. However, only a subset of these articles considers statistical uncertainty when reporting the results. O Povo’s headlines always mention their aggregator. Still, the vote intention appears in just one title. Interestingly, this outlet provides unique insights by reporting on government approval and rejection rates, and trends for blank votes, null votes, and undecided voters. Notably, two of O Povo’s articles were originally from Estadão, disseminated via Agência Estado, a news agency that sells content for reproduction by other outlets. This allowed Estadão to propagate its aggregator even on a platform with its own aggregator. Uol sets itself apart by embedding the interactive graph of the aggregator directly into the articles. This unique approach allows readers to visualize the progression of candidates’ vote intention. Finally, CNN Brasil produced only one article that primarily uses a poll aggregator, focusing on the announcement of the tool’s launch without discussing specific candidates or vote intentions. These variances among outlets illustrate the diverse ways in which

poll aggregators are utilized in news reporting, reflecting distinct journalistic practices and perspectives.

Final Considerations

For this work, our research question was how statistical models are appropriated by Brazilian media outlets to guide the production of news articles of the 2022 presidential elections. Our hypothesis was that poll aggregators can influence the way news is produced. In this sense, we were able to observe marginal changes in newsworthiness that support the hypothesis. However, at least for the 2022 elections, it is possible to ascertain that the degree of influence is minimal. We can conclude that the use of forecasting models in Brazilian media outlets is still at an early stage and tracing a single pattern of use proves to be a difficult task. What can be observed are experiments among the different news websites. In this work, we try to categorize the articles that use aggregators in three application patterns — news that uses the aggregator as a primary source, news that uses it as data, and news

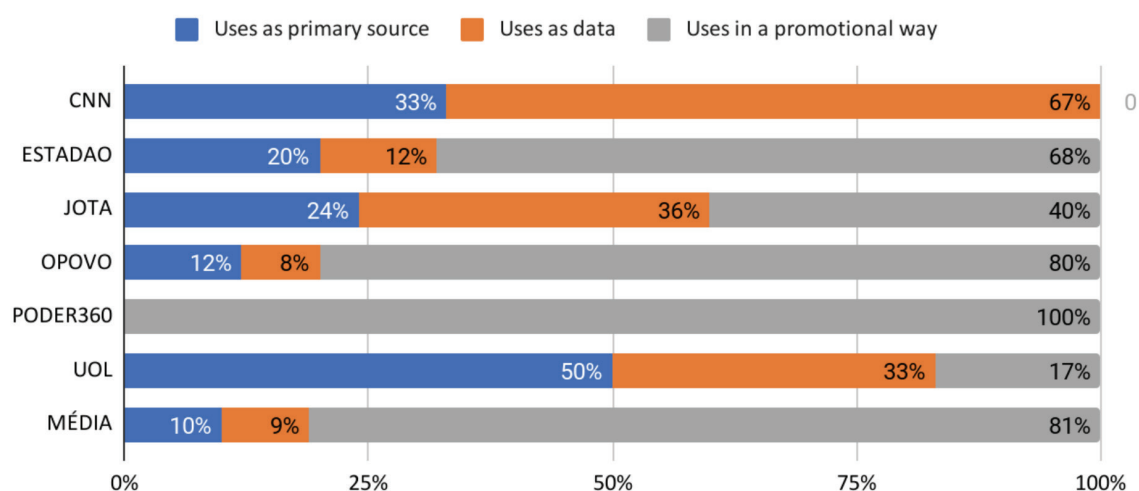


Figure 5. Aggregator usage in each newspaper by category proportionally

Source: Authors

that uses it in a promotional way —, but even these showed a plurality of ways of use within their respective categories. There are still a few articles in which the poll aggregator is the main news (10 percent of the sample), which may indicate that they still do not replace individual polls in terms of newsworthiness. Regardless of polls having a supposedly lower predictive capacity, they still retain greater symbolic capital in relation to aggregators in terms of newsworthiness, and, thus, newspapers tend to report more the dissemination of independent polls than the variation that an aggregator presented after the insertion of new data.

The high number of publications that mention the poll aggregator is not enough to state that the tool was fully inserted into the newsmaking, since the majority (81 percent) use it in a promotional way, that is, “external” to the news. In these cases, the aggregator does not appear integrated in the story, but as a complementary tool that the reader can check on another page, detached from the article. The result is not surprising, considering that only 2 newspapers

(Jota and Poder360) already had a poll aggregator before 2022. All the others were created this year, and it is worth mentioning that the data collection interval was a short period, considering the launch of the aggregator of each media outlet and, therefore, the data deal with a new product in which the newspapers still had no experience of working. This work was also limited to covering only the pre-election period when there is less news compared to the final stretch of the elections. If data from the electoral period had been considered, perhaps they would have contributed to a more complete picture of the usage of the tool, in view of the greater focus of the press on the elections and possible habituation to the use of the aggregator.

Although the use of aggregators is still in the beginning, it is possible to verify that they tend to gain more prominence, and it is worth checking if they will be used in future elections. CNN Brasil using its aggregator daily on television to make political analysis in elections is an example of this. The broadcaster even defined which candidates for state government

would be called for an interview (and the order to do so) according to their placements in the aggregator. Another example is Estadão, which was prolific among those who used the aggregator as a primary source and reported it in a similar way to an individual poll, in addition to having exported news about its aggregator in other portals due to the work of *Agência Estado*, which sells news to other media outlets. In addition to the news, this can be a way for the vehicle to expand its authority and influence, being able to put itself, in a way, on a par with competitors such as Poder 360 and Folha de S.Paulo, which have their own polls and export it to others media outlets.

Thus, aggregators have the potential to be used as a way for the press to gain credibility and notoriety. This possible usefulness is especially important at a time when both polls and the press that reports them are suffering from growing public distrust. According to Louwerse and Dijk (2022), polls serve as a valuable source of news, and electoral coverage focuses heavily on reporting the so-called “horse

race”. Bhatti and Pedersen (2016) argue that the search for news can result in a less qualified report in statistical terms. Added to this is the little knowledge that journalists usually have of mathematical concepts, and the search for interesting information to be reported can result in mistaken interpretations of the result. This is corroborated by the way in which information is passed on to the public. In the case of aggregators, 38 percent of the news does not even explain how the aggregator works. This can hinder the reader’s understanding of statistical uncertainty. In this sense, models developed to anticipate presidential election results should be more probabilistic than deterministic, offering a range of possibilities rather than just a number (Doskočilová, 2019). The same should apply to the press, which publishes the results. An example of work being carried out in this sense is Jota, which publishes news from the aggregator not based on the candidate’s voting intentions in a deterministic way, but reports the voting intention within an interval. In addition, a lot

of news deals with the probability of a candidate winning in the first round, rather than how many votes he will get in the first round. This way of reporting makes the uncertainty more evident.

That is, aggregators can be a good way to increase the credibility of the press by disclosing a number that is supposedly more accurate than individual polls, but this depends on two factors, which the aggregator alone is not capable of solving. The first is the way the result is reported since the quality of the news depends on the way the data are presented by the researchers and the way the publication is organized (Louwerse & Dijk, 2022). The second, in the Brazilian case, is to calibrate opinion polls so that they are more representative of the electorate that it failed to capture in the first round of the Brazilian presidential election. Even though the aggregator has techniques for normalizing the results, it is of no use if the polls on which they are based are detached from reality. When it comes to data, the old adage applies: garbage in, garbage out.

References

- Bhatti, Y., & Pedersen, R. T. (2016). News reporting of opinion polls: Journalism and statistical noise. *International Journal of Public Opinion Research*, 28(1), 129–141. <https://doi.org/10.1093/ijpor/edv008>
- Bramatti, D. (2022, May 30th). Estadão lança agregador de pesquisas; Lula tem vantagem de 18 pontos sobre Bolsonaro. *O Estado de São Paulo*. <https://www.estadao.com.br/politica/agregador-de-pesquisas-na-media-estadao-dados-lula-esta-16-pontos-a-frente-de-bolsonaro/>
- Doskočilová, K. (2019). *Forecasting election results in the Czech Republic* [Master thesis, Charles University]. Digitalni repozitar - University Karlovy. <https://dspace.cuni.cz/handle/20.500.11956/110193/>
- El-Dash, N. A. (2014). Prevendo o resultado da eleição presidencial brasileira de 2014. Trabalho apresentado no XXI *Simpósio Nacional de Probabilidade e Estatística*, Natal, Brasil. http://www.pollingdata.com.br/links/Apresentacao_Neale_Sinape2014.pdf
- Jackman, S. (2005). Pooling the polls over an election campaign.

- Australian Journal of Political Science*, 40(4), 499–517. <https://doi.org/10.1080/10361140500302472>
- Lewis-Beck, m. S. (2005). Election forecasting: Principles and practice. *The British Journal of Politics and International Relations*, 7(2), 145–164. <https://doi.org/10.1111/j.1467-856X.2005.00178.x>
- Louwerse, T., & Van Dijk, R. E. (2022). Reporting the polls: The quality of media reporting of vote intention polls in the Netherlands. *Acta Politica*, 57(3), 548–570. <https://doi.org/10.1057/s41269-021-00208-5>
- Marcelino, D. (2022, May 2nd). Por que as pesquisas eleitorais divergem tanto? Quem está certo? Quem sabe? *JOTA*. <https://www.jota.info/opiniao-e-analise/colunas/coluna-daniel-marcelino/por-que-as-pesquisas-eleitorais-divergem-tanto-quem-esta-certo-quem-sabe-02052022>
- Mariante, J. H. (2022, June 4th). Dois pontos: Uma praga. *Folha de São Paulo*. <https://www1.folha.uol.com.br/colunas/jose-henrique-mariante-ombudsman/2022/06/dois-pontos-uma-praga.shtml>
- Pasek, J. (2015). Predicting elections: Considering tools to pool the polls. *Public Opinion Quarterly*, 79(2), 594–619. <https://doi.org/10.1093/poq/nfu060>
- Toledo, J. R. (2022, 21 de setembro). Confiança na imprensa é menor do que no WhatsApp. *UOL*. <https://noticias.uol.com.br/colunas/jose-roberto-de-toledo/2022/09/21/confianca-na-imprensa-e-menor-do-que-no-whatsapp.htm>
- Welch, R. L. (2002). Polls, polls, and more polls: An evaluation of how public opinion polls are reported in newspapers. *Harvard International Journal of Press/Politics*, 7(1), 102–114. <https://doi.org/10.1177/1081180X0200700107>

