

Social Listening as a Public Health Surveillance Sensor: A Comparative Analysis of Vaccine Discourse in Brazil

Isabela Braga de Moura¹, Claudia Araujo¹

¹COPPEAD– Universidade Federal do Rio de Janeiro (UFRJ)
Rua Pascoal Lemme, 355 - Cidade Universitária (Ilha do Fundão) – Rio de Janeiro – RJ
– Brazil – CEP 21941-616

isabela.braga@coppead.ufrj.br, claraujo@coppead.ufrj.br

Abstract. *Social listening often assumes that digital signals about different vaccines are interchangeable. This study investigated this assumption by comparing COVID-19, influenza, and HPV vaccine mentions in Brazil (2025) using Meltwater. Analysis of mentions revealed structurally distinct signals: politicized and anger-driven on X for COVID-19; institutional and neutral in online news for influenza; clinical, female-authored, and responsive to awareness months for HPV. Findings demonstrate that social listening captures each vaccine's sociopolitical embedding rather than generic discourse, and that managers should monitor vaccines individually, prioritize engagement over volume, and use platform distribution as an early indicator of signal type.*

1. Introduction

Immunization is one of the most effective public health interventions [Gessner et al. 2017], yet its full societal value depends on a collective good that is inherently fragile: herd immunity. When individuals delay or refuse vaccines, they do not merely forgo a personal health product — they withdraw from a coordinated endeavor whose benefits are indivisible and relational. In low- and middle-income countries, this dynamic is particularly acute. Vaccine hesitancy, recognized by the World Health Organization (WHO) as a top-ten global health threat [World Health Organization 2019], interacts with weak health systems, historical distrust, and a rapidly evolving digital information landscape to create persistent pockets of under-immunization [World Health Organization (WHO) 2024]. This information landscape has been characterized as an infodemic — an overabundance of information, including misinformation and disinformation, that degrades the quality of resources available for health decision-making [Pan American Health Organization 2020; Zarocostas 2020]. Amid the COVID-19 pandemic, infodemic dynamics propelled social listening — the systematic monitoring and analysis of online public discourse — from a niche marketing practice to a core tool of public health surveillance. The WHO's Early AI-Supported Response with Social Listening (EARS) platform, for instance, now delivers real-time categorized data on COVID-19 narratives across dozens of languages [Purnat et al. 2021; White et al. 2023]. Social listening is increasingly regarded as a “sensor” for community perceptions, capable of detecting information voids and tracking sentiment toward health interventions.

A critical, often tacit assumption underlies this enthusiasm: implicit in the adoption of social listening for public health is the notion that digital signals about different vaccines are essentially interchangeable, applying a uniform analytical lens that

treats all vaccine-related content as manifestations of the same underlying phenomenon. Yet a growing body of research suggests otherwise. Vaccines are not homogeneous objects; they differ in their target populations, their historical narratives, their benefit-risk profiles, and, critically, their degree of politicization [Weinzierl et al. 2023]. COVID-19 vaccines became entangled with partisan identity and institutional distrust [Bulilan et al. 2024], while HPV vaccines navigate gendered, moral, and sexual health discourses [Santiago et al. 2024]. The question thus arises: When social listening is applied to different vaccines, does it capture a common “vaccine discourse” – or does it capture structurally different signals?

Answering this question is not merely an academic exercise. For managers of immunization programs, the utility of social listening as a surveillance tool hinges on understanding what exactly is being measured. If the signal for a politicized vaccine is predominantly political mobilization rather than health perception, then interpreting it as public sentiment toward vaccination could lead to misguided communication strategies. Conversely, if routine vaccines produce a cleaner health-related signal, then social listening can be a reliable complement to traditional epidemiological surveillance.

Brazil offers a uniquely valuable context for such an investigation. The country possesses one of the world’s most extensive public health systems (SUS), conducts large-scale annual campaigns for influenza and HPV, and experienced a highly polarized COVID-19 immunization debate [Gramacho and Turgeon 2021]. Moreover, survey data indicate that 86% of adult Brazilians use social media, placing the country among the most digitally engaged populations globally [Pew Research Center 2024]. This high connectivity generates a rich and diverse corpus of geo-locatable online discourse, making Brazil a particularly suitable setting for social listening research. By comparing the digital traces of COVID-19, influenza, and HPV vaccines over the same period (2025) using a single commercial social listening platform (Meltwater), we can isolate whether the signals differ and, if so, along which dimensions.

Accordingly, this study addresses the following research question: How does a commercial social listening platform configure different digital signals for vaccines with distinct degrees of politicization? We analyze three corpora — totaling approximately 494,000 social media and online news mentions — across metrics including platform distribution, automated sentiment, dominant emotion, top entities, author gender, temporal patterns, and engagement asymmetry. The analysis reveals that the three vaccines generate signals that are not merely different in volume but structurally distinct: each is anchored in different platform ecologies, emotional tones, and institutional or community logics.

2. Method

This study adopts a comparative social listening design, examining three vaccine-specific corpora collected from a single commercial platform over the same period. The following subsections detail the data source, search strategy and corpora construction, analytical metrics, and ethical considerations.

2.1. Data Source

Data was collected using Meltwater, a commercial social media intelligence platform providing access to publicly available online content across X, online news outlets (major

and smaller regional outlets), blogs, broadcast, podcasts, reviews, and forums. We collected all geo-located mentions originating from Brazil during the period January 1, 2025, to December 31, 2025. Location filtering relied on Meltwater’s native geolocation feature, which uses a machine-learning inference process based on publication metadata and profile information. Only content assigned to Brazil was retained.

The Meltwater platform provides automated sentiment analysis (positive, negative, neutral), emotion classification (e.g., anger, fear, joy, sadness), extraction of named entities (persons, organizations, locations), and keyword frequency distributions. All metrics used in this study were provided directly by the platform; no additional manual coding was performed on the full corpus.

2.2. Search Strategy and Corpora

A two-stage approach was used to build the three vaccine corpora. First, a broad vaccine-related query was designed to capture general immunization discourse in Brazil during the period January 1, 2025, to December 31, 2025. The query, entered in Portuguese, was: ((vacina* OR antivax* OR antivacin* OR imuniza* OR "anti-vax*" OR "anti-vac*") AND country:BR) NOT (bbb* OR "big brother brasil" OR (imun* NEAR/5 anjo)).

This query retrieved all publicly available mentions containing vaccine-related terms while excluding content associated with the Brazilian reality television show Big Brother Brasil, which had generated viral but irrelevant uses of the word imunidade (immunity). The initial corpus served as the universe from which vaccine-specific subsets were extracted.

Second, three vaccine-specific filters were applied to the initial corpus, producing the final analytic corpora (Table 1). Each filter was applied to the full vaccine-related dataset. Because the filters operated independently, a single social media post containing terms from more than one vaccine (e.g., a mention of both COVID-19 and influenza vaccination) was retained in each relevant corpus. The three corpora were analyzed separately. The resulting corpora comprised approximately 331,000 mentions for COVID-19, 123,000 for Influenza, and 40,100 for HPV. Detailed volume distributions by platform and by month are reported in the Results section.

Table 1. Vaccine-Specific Filter Terms

Corpus	Filter Terms (Portuguese)
COVID-19	COVID* OR coronavirus*
Influenza	gripe OR influenza*
HPV	HPV* OR Papilomavirus*

2.3. Analytical Metrics

For each corpus, a set of metrics was extracted or computed from the Meltwater platform. All metrics were provided directly by the platform; no additional manual coding was performed on the full corpora. Table 2 lists and defines each metric.

Table 2. Analytical Metrics Used in the Comparative Social Listening Analysis

Metric	Definition
Volume	Absolute number of mentions captured for each corpus over the study period, allowing temporal trend analysis.
Platform distribution	Percentage of mentions originating from X (formerly Twitter), online news outlets, and other web sources. Captures the dominant communication ecology for each vaccine.
Automated sentiment	Percentage of mentions classified by Meltwater's algorithm as positive, negative, or neutral. Used as an indicative proxy for the affective tone of discourse.
Dominant emotion	Most frequently detected emotion (anger, fear, sadness, joy) within each corpus, as classified by Meltwater's automated emotion detection.
Top entities	Most frequently mentioned persons, organizations, and locations. Used to identify whether discourse is anchored in political, institutional, or clinical referents.
Keywords by sentiment	Most frequent terms appearing in negative-classified and positive-classified posts. Provides a qualitative window into the content of discourse.
Temporal distribution	Trajectory of mention volume across the 12-month period, including peak detection and correlation with known campaign dates.
Author gender	Estimated gender distribution of unique authors, as inferred by Meltwater from profile metadata. Captures demographic skews in the audience for each vaccine.
Engagement concentration	Share of total engagement (shares, comments, reactions) captured by the top 5 % of editorial sources ranked by mention volume, supplemented by the ratio of engagement per mention for the highest-ratio source versus the highest-reach journalistic source. Measures the degree to which engagement is concentrated in a small number of sources, indicating amplification asymmetry.

All metrics were compared across the three corpora using descriptive statistics and visual inspection. No inferential statistical tests were applied, given the exploratory nature of the analysis and the non-independent structure of social media data. Two metrics required manual derivation. First, engagement concentration was computed by ranking content sources by mention volume, calculating the share of total engagement captured by the top 5 %, and comparing the engagement-per-mention ratio of the highest-ratio source with that of the highest-reach journalistic source. Second, top entities were classified into types (political, institutional, clinical) by the authors. Automated sentiment and emotion classification carry inherent limitations, particularly in Portuguese and in highly contextual or ironic political discourse. We therefore treat these metrics as indicative rather than as ground truth, and we focus our analysis on comparative patterns across corpora rather than on absolute values.

2.4. Ethical Considerations

This study was approved by the Research Ethics Committee of the Federal University of Rio de Janeiro (UFRJ) under protocol number 8.236.623. All data were collected in aggregated and anonymized form from publicly available sources; no individually identifiable user data were accessed or stored. The analysis focused exclusively on public discourse patterns, not on individual users. Limitations regarding the representativeness of social media data — particularly the overrepresentation of male users on X and the exclusion of private messaging platforms such as WhatsApp — are addressed in the Discussion section.

3. Results

The three vaccine corpora exhibited structurally distinct configurations across all analytical dimensions. Table 3 summarizes the main comparative metrics.

Table 3. Comparative Metrics for COVID-19, Influenza, and HPV Vaccine Corpora

Metric	COVID-19	Influenza	HPV
Total mentions	331K	123K	40K
Dominant platform (% of mentions)	X (77%)	Online News (62%)	Online News (64%)
Second platform	Online News (18%)	X (34%)	X (29%)
Negative sentiment (%)	57%	29%	15%
Neutral sentiment (%)	30%	63%	76%
Dominant emotion (%)	Anger (10%)	Anger (3%)	Anger (3%)
Top entity type	Political figures	Health institutions	Clinical/medical
Author gender majority (%)	Male (66%)	Male (58%)	Female (54%)
Peak activity period	Distributed across year	May 2025	March and October 2025
Top 5% sources — % of engagement	76%	65%	16.5%

3.1. Platform Distribution and the Political Ecology of Discourse

The three corpora displayed markedly different platform ecologies. COVID-19 discourse was overwhelmingly concentrated on X (77%), while Influenza and HPV discourse were hosted primarily by online news outlets (62% and 64%, respectively). X, as a platform optimized for short-form, high-velocity content and political commentary, amplifies polarizing and identity-based messaging. In contrast, online news portals in Brazil,

particularly reputable ones, tend to feature institutional voices, official announcements, and campaign coverage. The platform distribution alone thus anticipated the type of discourse each vaccine would exhibit before any content analysis was conducted.

3.2. Sentiment and the Gradient of Politicization

Automated sentiment analysis revealed a linear gradient in negative sentiment that paralleled the degree of politicization of each vaccine. COVID-19 exhibited the highest negativity (57%), followed by Influenza (29%) and HPV (15%). Conversely, neutral sentiment increased from COVID-19 (30%) through Influenza (63%) to HPV (76%). Emotion analysis reinforced this gradient: although anger was the dominant emotion across all three corpora, its intensity was markedly higher in the COVID-19 corpus (10%) than in the Influenza (3%) or HPV (3%) corpora. This suggests that politicization not only polarizes sentiment but also amplifies the emotional charge of discourse. The anger-laden language surrounding COVID-19 vaccination — driven by partisan conflict and institutional distrust — contrasted sharply with the predominantly neutral tone of HPV and Influenza conversations. These patterns suggest that automated sentiment and emotion metrics, despite their limitations, capture meaningful differences in the affective temperature of vaccine discourse.

3.3. Gender Inversion in the HPV Corpus

Author gender distribution revealed a notable pattern: both COVID-19 and Influenza corpora exhibited a male-majority author base, consistent with the known demographic skew of political and general-interest health discourse on X. The HPV corpus, however, was the only one with a female majority (54%). This inversion aligns with the target demographic of HPV vaccination campaigns—adolescent girls and young women—and with the clinical framing of cervical cancer prevention. It suggests that social listening can capture genuine demographic specificities when the health topic has a clear demographic anchor, and that platform and topic interact to shape who speaks and who is heard.

3.4. Temporal Dynamics and Institutional Anchoring

The temporal trajectories of the three corpora, displayed in Figure 1, underscore their distinct rhythms. Influenza mentions exhibited a sharp peak in May 2025, coinciding with the national influenza vaccination campaign.

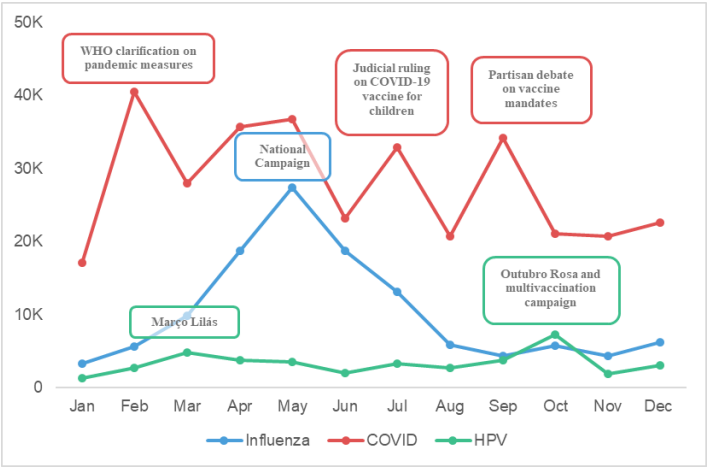


Figure 1. Monthly mention volume for COVID-19, influenza, and HPV vaccines in Brazil (2025)

HPV discourse, although lower in absolute volume, showed two pronounced peaks. The first occurred in March 2025 and was linked to *Março Lilás*, a national awareness campaign for cervical cancer prevention that explicitly promotes HPV vaccination. The second, and largest, occurred in October 2025, coinciding with the combination of *Outubro Rosa* (breast and cervical cancer awareness month) and a nationwide multivaccination campaign targeting children and adolescents up to 15 years, launched by the Ministry of Health on October 1 and culminating in a national mobilization day on October 18. These dual peaks demonstrate that HPV discourse responds sensitively to public health communication events, reinforcing the interpretation that this corpus is anchored in the institutional calendar.

COVID-19 mentions saw multiple spikes throughout the year, with no sharp campaign-driven peaks. Fluctuations were associated with political and judicial events rather than health communication initiatives. A surge in January 2025 aligned temporally with a statement by the WHO Director-General clarifying that the organization had not imposed lockdowns, masks, or vaccines during the pandemic. A spike in July coincided with a judicial ruling in southern Brazil that questioned the experimental nature of COVID-19 vaccines for children and rejected sanctions against parents who had not immunized their children. A further peak in September corresponded with partisan social media debates over the constitutionality of vaccine mandates and the division of responsibilities between federal and state authorities during the pandemic. These temporal associations suggest that where COVID-19 vaccination discourse rises in volume, it may do so primarily through political contestation rather than institutional health communication.

3.5. Engagement Asymmetry and Ideological Amplification

To quantify engagement asymmetry, editorial sources were ranked by mention volume and by engagement ratio (total engagements divided by total mentions). The three corpora exhibit structurally distinct profiles. In the COVID-19 corpus, the top 5% of sources concentrated 76% of all editorial engagement, with the leading high-ratio outlet achieving 523 engagements per mention against 30 for the highest-reach journalistic source — a 17:1 disparity. In the influenza corpus, concentration was lower (65%), yet the same high-ratio outlets dominated, suggesting a persistent amplification infrastructure independent of vaccine topic. The HPV corpus diverges sharply: the top 5% captured only 16.5% of total engagement, indicating genuinely diffuse engagement with no structural amplification.

4. Discussion

The comparative analysis of COVID-19, influenza, and HPV vaccine discourse in Brazil yields a central empirical finding: commercial social listening does not capture a homogeneous “vaccine discourse” but instead reveals structurally distinct signals that are consistent with the sociopolitical embedding of each immunization product. This finding resonates with and extends several strands of existing research.

First, the results empirically substantiate the critique advanced by Heyerdahl et al. [2023], who argued that purely online social listening risks misrepresenting local health

perceptions because it strips discourse of its social context. By demonstrating that the digital signal for COVID-19 is predominantly political — dominated by anger, concentrated on X, and temporally responsive to judicial and partisan events — our study provides quantitative evidence for the kind of contextual disembedding that the authors cautioned against. Conversely, the influenza and HPV corpora, with their institutional and clinical anchoring, illustrate that the same methodological tool can yield contextually meaningful signals when the health topic is not enveloped in partisan conflict.

Second, the findings problematize the implicit assumption of signal homogeneity embedded in unified social listening architectures such as the WHO EARS platform [White et al. 2023]. EARS applies a single public health taxonomy to all COVID-19 narratives, implying that vaccine-related content can be analyzed with the same set of categories regardless of the specific vaccine. Our data suggest that this assumption is untenable: the type of discourse captured is a function of the vaccine’s politicization, not merely of the topic “vaccination.” This aligns with Bulilan et al. [2024], who documented how COVID-19 vaccines became multilayered politicized constructions in the Global South, and with Santiago et al. [2024], who showed that HPV vaccine discourse in Brazil is traversed by gendered and moral framings distinct from those of other vaccines.

Third, the gender inversion observed in the HPV corpus — the only one with a female-majority author base — echoes the community-specific dynamics described by Rader et al. [2014] in their analysis of pharmaceutical bonding in social media. The authors found that online communities around specific health conditions and treatments often coalesce around shared identity and experience. The HPV corpus, concentrated on cervical cancer prevention and targeting adolescent girls and young women, appears to function as such a community, generating a discourse that is demographically distinct from the male-dominated political conversation around COVID-19. Moreover, the sharp contrast between the politicized COVID-19 discourse and the clinically framed HPV discourse resonates with Weinzierl et al. [2023], who demonstrated that these two vaccines elicit distinct misinformation themes and trust concerns on Twitter. Together, these findings reinforce the interpretation that the digital signal is shaped by the specific sociopolitical embedding of each vaccine.

4.1. Practical Implications for Health Managers

Three operational recommendations emerge from this analysis. First, public health managers should monitor vaccines individually rather than aggregating “vaccine discourse” — the aggregate signal is dominated by the most politically charged vaccine and obscures the dynamics of routine immunization campaigns. Second, engagement metrics should be prioritized over volume metrics for politically charged vaccines — where the top 5% of sources by volume concentrated 76% of COVID-19 engagement against only 16.5% for HPV — but are less decisive for vaccines with diffuse, institutionally anchored discourse. Third, the dominant platform should be used as a leading indicator of discourse type: high X concentration signals political discourse; high online news concentration signals institutional and informational discourse. Together, these recommendations enable infodemic managers to triage social listening data more efficiently, allocating analytical resources according to the type of signal each vaccine generates rather than applying a uniform lens to all immunization topics.

4.2. Toward a Brazilian Public Health Social Listening Taxonomy

The WHO EARS platform demonstrated the value of applying a purpose-built public health taxonomy to social listening data, enabling machine learning classification of narratives into actionable categories (White et al., 2023). Brazil has sufficient data volume and linguistic specificity to develop an equivalent system in Brazilian Portuguese. The structural differences identified in this study — particularly the distinction between political, institutional, and clinical discourse types — provide a foundation for the categorical architecture such a taxonomy would require. Building on these distinctions, future research should develop and validate this taxonomy in collaboration with public health institutions, following the example of Weinzierl et al. [2023], who derived separate misinformation and trust taxonomies for HPV and COVID-19 vaccines from Twitter discourse.

4.3. Limitations

Several limitations of this study should be acknowledged. The data were collected via a single commercial platform whose automated sentiment and emotion algorithms have known performance gaps in Portuguese, particularly in ironic and politically charged contexts. We addressed this by focusing on comparative patterns rather than absolute values, but the numerical estimates remain indicative. The study is also limited to public, geo-located online content; private messaging platforms (notably WhatsApp) and offline discourse, both central to vaccine decision-making in Brazil, remain inaccessible to commercial social listening. Finally, the observational design precludes causal claims about the relationship between online discourse and vaccination behavior.

5. Conclusion

This study demonstrated that a commercial social listening platform captures structurally distinct signals for COVID-19, influenza, and HPV vaccines in Brazil. Rather than reflecting a generic “vaccine discourse,” the digital traces of each vaccine correspond to its specific sociopolitical embedding: a politicized, anger-driven conversation on X for COVID-19; an institutional, campaign-anchored conversation in online news for influenza; and a clinical, female-majority conversation for HPV. These three configurations suggest that social listening is not a neutral sensor — it reflects the ecosystem in which each vaccine circulates, not merely public opinion about vaccination.

For infodemic management, the findings urge a shift from one-size-fits-all social listening to vaccine-specific, structurally aware surveillance. The practical recommendations offered — disaggregating vaccines, prioritizing engagement over volume, and using platform as a leading indicator — provide actionable steps in this direction. The development of a Brazilian Portuguese public health taxonomy, grounded in the institutional calendar and political landscape of the country, would further equip managers with the contextual intelligence required to interpret digital signals accurately.

References

- Bulilan, E. J. B., Uyheng, J. and Montiel, C. J. (8 oct 2024). Multilayered Politicized Constructions of Covid-19 Vaccines: A Global South View of Pfizer and Sinovac. *Journal of Social and Political Psychology*, v. 12, n. 2, p. 173–187.

- Gessner, B. D., Kaslow, D., Louis, J., et al. (1 nov 2017). Estimating the full public health value of vaccination. *Vaccine*, v. 35, n. 46, p. 6255–6263.
- Gramacho, W. G. and Turgeon, M. (6 may 2021). When politics collides with public health: COVID-19 vaccine country of origin and vaccination acceptance in Brazil. *Vaccine*, v. 39, n. 19, p. 2608.
- Heyerdahl, L. W., Le Marcis, F., Nguyen, T. T., et al. (2023). Parallel vaccine discourses in Guinea: ‘grounding’ social listening for a non-hegemonic global health. *Critical Public Health*, v. 33, n. 5, p. 579–593.
- Pan American Health Organization (2020). Understanding the infodemic and misinformation in the fight against COVID-19. . <https://iris.paho.org/handle/10665.2/52052>, [accessed on Dec 21].
- Pew Research Center (2024). 8 charts on technology use around the world | Pew Research Center. <https://www.pewresearch.org/short-reads/2024/02/05/8-charts-on-technology-use-around-the-world/>, [accessed on Apr 24].
- Purnat, T. D., Wilson, H., Nguyen, T. and Briand, S. (1 jul 2021). EARS - A WHO Platform for AI-Supported Real-Time Online Social Listening of COVID-19 Conversations. *Studies in health technology and informatics*, v. 281, p. 1009–1010.
- Rader, C. S., Subhan, Z., Lanier, C. D., et al. (27 may 2014). CyberRx: Emerging social media marketing strategy for pharmaceuticals. *International Journal of Pharmaceutical and Healthcare Marketing*, v. 8, n. 2, p. 193–225.
- Santiago, G. H. de P., Nardino, F., Machado, R. R. de S. and Lima, F. L. T. De (19 dec 2024). A vacina contra o papilomavírus humano em discursos de brasileiros na rede social X. *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, v. 18, n. 4, p. 879–897.
- Weinzierl, M. A., Hopfer, S. and Harabagiu, S. M. (1 apr 2023). Scaling up the discovery of hesitancy profiles by identifying the framing of beliefs towards vaccine confidence in Twitter discourse. *Journal of Behavioral Medicine*, v. 46, n. 1–2, p. 253–275.
- White, B. K., Gombert, A., Nguyen, T., et al. (2023). Using Machine Learning Technology (Early Artificial Intelligence–Supported Response With Social Listening Platform) to Enhance Digital Social Understanding for the COVID-19 Infodemic: Development and Implementation Study. *JMIR Infodemiology*, v. 3.
- World Health Organization (2019). Ten threats to global health in 2019. <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>, [accessed on Oct 17].
- World Health Organization (WHO) (jul 2024). Immunization coverage. <https://www.who.int/news-room/fact-sheets/detail/immunization-coverage>, [accessed on Feb 28].
- Zarocostas, J. (feb 2020). How to fight an infodemic. *Lancet (London, England)*, v. 395, p. 676.