Seah and Iwakuma *BMC Public Health* (2024) 24:2447
https://doi.org/10.1186/s12889-024-19813-y

BMC Public Health

RESEARCH

A quantitative content analysis of topical characteristics of the online COVID-19 infodemic in the United States and Japan

Matthew Seah[1*] and Miho Iwakuma[1]

Abstract

Background The COVID-19 pandemic has spurred the growth of a global infodemic. In order to combat the COVID-19 infodemic, it is necessary to understand what kinds of misinformation are spreading. Furthermore, various local factors influence how the infodemic manifests in different countries. Therefore, understanding how and why infodemics differ between countries is a matter of interest for public health. This study aims to elucidate and compare the types of COVID-19 misinformation produced from the infodemic in the US and Japan.

Methods COVID-19 fact-checking articles were obtained from the two largest publishers of fact-checking articles in each language. 1,743 US articles and 148 Japanese articles in their respective languages were gathered, with articles published between 23 January 2020 and 4 November 2022. Articles were analyzed using the free text mining software KH Coder. Exploration of frequently-occurring words and groups of related words was carried out. Based on agglomeration plots and prior research, eight categories of misinformation were created. Lastly, coding rules were created for these eight categories, and a chi-squared test was performed to compare the two datasets.

Results Overall, the most frequent words in both languages were related to health-related terms, but the Japan dataset had more words referring to foreign countries. Among the eight categories, differences with chi-squared $p \leq 0.01$ were found after Holm-Bonferroni p value adjustment for the proportions of misinformation regarding statistics (US 40.0% vs. JP 25.7%, ϕ 0.0792); origin of the virus and resultant discrimination (US 7.0% vs. JP 20.3%, ϕ 0.1311); and COVID-19 disease severity, treatment, or testing (US 32.6% vs. JP 45.9%, ϕ 0.0756).

Conclusions Local contextual factors were found that likely influenced the infodemic in both countries; representations of these factors include societal polarization in the US and the HPV vaccine scare in Japan. It is possible that Japan's relative resistance to misinformation affects the kinds of misinformation consumed, directing attention away from conspiracy theories and towards health-related issues. However, more studies need to be done to verify whether misinformation resistance affects misinformation consumption patterns this way.

Keywords Infodemic, COVID-19, Pandemic, Multi-country, Health information, Fact checking, Myth busting, Myth correction

*Correspondence:
Matthew Seah
seahmatthew@gmail.com
[1]Department of Medical Communication, Kyoto University, Sakyo-ku Yoshida-konoe-cho, Kyoto, Japan

Introduction

The COVID-19 pandemic has brought into the spotlight the growing *infodemic*: the "excessive amount of unfiltered information concerning a problem such that the solution is made more difficult" [1]. Between the mainstream media, statements made by politicians, social media platforms, instant messaging services, and changing guidelines released by official institutions, the typical person is constantly inundated with a barrage of information that presents both the *challenge* of discerning reliable information, as well as the *option* to take fringe or pseudoscientific theories as the truth. This represents a public health concern, as COVID-19 misinformation or "fake news" may spread anti-vaccine views or promote racial discrimination [2].

A multi-pronged approach is necessary to mitigate the impact of the infodemic, as no single intervention can achieve the breadth required to match the scale of the worldwide flow of information. Eysenbach proposes four pillars of infodemic management in his 2020 paper: infoveillance and infodemiology (surveillance of information supply and demand, as well as its quality); building eHealth literacy; improving the translation of knowledge between academia and larger outlets such as policymakers, mainstream media, and social media; and the peer-review process and *fact-checking* [3].

"Fact-checking" refers to the process of evaluating a statement for its factual accuracy or whether it has been framed in a misleading manner due to omission of context. Fact-checking has its origins in American TV segments devoted to checking the accuracy of statements made by American presidential candidates [4], though most current fact-checking content is produced by websites such as Snopes or FactCheck.org in the form of articles or videos.

Fact-checking alone cannot be the ultimate counter to misinformation – not only does it have limited effects on correcting perceptions of misinformation due to the strong biases and emotions involved when interacting with such information [4, 5], the *local politics of truth* [6], i.e. the historical and cultural contexts of the region, inform behavior and beliefs to a significant degree; for instance, close-contact burial practices in parts of west Africa stricken by ebola [7], or vaccine hesitancy in Japan following the HPV vaccine scare in 2013 [8]. Interventions targeting an infodemic need to take into account the nature and context of the region to be effective.

One of the few extant studies comparing the COVID-19 infodemics and national contexts across countries was published by Zeng et al. [9], in which they analyzed fact-checking article contents from the US, China, India, Germany, and France. Some key findings included the fact that non-health misinformation (e.g. regarding politics, or the origin of the virus) is nearly twice as common as health misinformation (e.g. COVID-19 being "just a cold"); Germany is relatively resilient to misinformation compared to the US or India owing to its low societal polarization and high trust in the news media; misinformation regarding the spread of COVID-19 or travel restrictions is common in China, likely due to China being the early epicenter of the pandemic as well as large-scale travel movements that occur around Chinese New Year; and wedge-driving misinformation along religious lines is common in India owing to the longstanding conflict between the nation's Muslim and Hindu populations.

Although there is already an abundance of cross-cultural research between the US and Japan, a comparative study of infodemics in these countries has yet to be done, and much has changed in the time since the publication of the Zeng paper – noteworthy developments including the progress made in global vaccination campaigns [10], and the emergence of the highly transmissible delta and omicron variants [11]. Furthermore, the national contexts of the US and Japan differ to a notable extent, in geographical, sociocultural, and historical terms, making it reasonable to expect differences in the types of misinformation that would gather more traction. Therefore, this research aims to provide an updated understanding of the COVID-19 infodemics in the US and Japan through a quantitative content analysis of the types of misinformation that appear in fact-checking articles.

Methodology

Data selection and gathering

In order to find the types of COVID-19 misinformation that gathered significant traction in the US and Japan, COVID-19 fact-checking articles were gathered from the top two largest fact-checking publishers: Politifact and FactCheck.org for the US, and Buzzfeed and InFact for Japan. All articles were written in their respective countries' languages (English for the US, Japanese for Japan). A summary of the data sources used is shown in Table 1 below. Articles included were published between 23 January 2020 and 4 November 2022.

Article URLs were scraped from the COVID-19 sections of each source in Python, using the Selenium library in Chrome 108.0.5359.124. Following this, a separate program was used to visit the listed URLs and scrape the article contents using the news-please library [16]. (Source codes can be accessed at https://github.com/seahmatthew/KyotoU-PublicHealth2023.)

Data analysis in KH coder

The open-source quantitative text analysis program KH Coder [17], developed by Koichi Higuchi at Ritsumeikan university, was used to analyze the article contents, with the US and Japan datasets in separate projects. As of January 2023, there are 5,761 published research articles

Table 1 Fact-checking article sources

Source	No. of articles	Remarks
Politifact (US) [12]	1402	Affiliated with the Poynter Institute, an American nonprofit school for journalists.
FactCheck.org (US) [13]	341	Affiliated with the Annenberg Public Policy Center of the University of Pennsylvania.
BuzzFeed (Japan) [14]	106	Fact-checking articles are gathered in a dedicated section, written by a handful of journalists.
InFact (Japan) [15]	42	Independent NPO consisting of members from the media industry, academia, and students.

which make use of KH Coder [18], many of which cover health-related research topics. Its strengths include functions for statistical analysis (e.g., term frequency) of large data files, as well as the KWIC Concordance function [19] which provides the capability to easily refer to the original data from any given result.

Word Frequency [19] was used to obtain an overview of the data as a preliminary step. Following this, Hierarchal Cluster Analysis [19] was used to explore groups of related words, and also to build the lists of terms to force pickup (such as "toilet paper" or "Moderna") which would not be picked up by default, and irrelevant terms to force ignore (such as "website" or "article"), which introduce noise due to appearing very frequently but not being indicative of any relevant themes. This took a process of trial and error especially when building the force ignore lists, as blocking certain seemingly irrelevant terms would sometimes turn out to hide an otherwise useable article.

After substantive force pickup/ignore lists had been built for each languages, the lists were compared to ensure that relevant keywords were ignored in both languages, although words that appear frequently as syntactic features in each language (such as "pants [on] fire" or "subject") were not duplicated in the same way.

Next, Hierarchal Cluster Analysis was re-run using the finalized force pickup/ignore lists to gather the terms to form the document coding files. For the U.S. dataset, the minimum Term Frequency (TF) was set to 90, Document Frequency (DF) to 1, and only nouns, proper nouns, and terms from the force pickup list were analyzed to minimize noise. For the Japan dataset, the minimum TF was set to 10, DF to 1, and only nouns, proper nouns, location names, and terms from the force pickup list were analyzed. For both datasets, the Ward method and Jaccard frequency were used, with the number of clusters shown being auto-chosen.

Based on the agglomeration plot turning points from the Hierarchal Cluster analyses, the prior Zeng paper [9], and familiarity with the data, it was decided to split the data into eight categories. From the categories and keywords found, coding files were built for the US and Japan datasets and applied to obtain the frequencies for each category. Articles could be assigned to multiple categories, and manual sorting was used to classify articles through a first pass after automatic sorting. Articles that failed to be classified in any category after both automatic and manual sorting were assigned to a separate Miscellaneous category.

After the code frequencies for each language had been obtained, chi-squared tests were carried out to test whether there were differences in the frequencies across countries. Holm-Bonferroni adjustment was used to adjust the *p* values.

Results

The agglomeration plots produced from the Hierarchal Cluster analyses are shown below in Fig. 1. The turning points show that somewhere in the range of seven categories would be ideal, but considering prior research and familiarity with the data, it was decided to generate eight categories.

The coding files created based on the categories and keywords found are shown in Table 2. A total of eight categories were created: government policy; resource shortages; statistics; measures to stem the spread of infection; masks and transmission; origin of the virus and resultant discrimination; COVID-19 disease severity, treatment, or testing; and vaccine efficacy, contents, or safety. Each category contains a set of keywords in its respective language that results in close association; for instance, "lockdown", "quarantine", and "border" associate highly with articles about measures taken to stem the spread of infection.

A summary of the top 50 words with the highest *tf* (term frequency) is shown in Table 3. Both the U.S. and Japan lists are topped by words pertaining to vaccination, masks, cases and testing, likely because these words are likely to appear across a broad range of categories. For instance, words pertaining to vaccination could appear in both articles about supposed deleterious health effects of vaccination, as well as articles about vaccination program plans or vaccine-related conspiracy theories.

A summary of the code frequencies, chi-squared test *p* values, and relevant excerpts from the data is provided below in Table 4. Articles that contained none of the eight predetermined codes are grouped in the "Miscellaneous" category. Chi-squared tests were carried out to compare

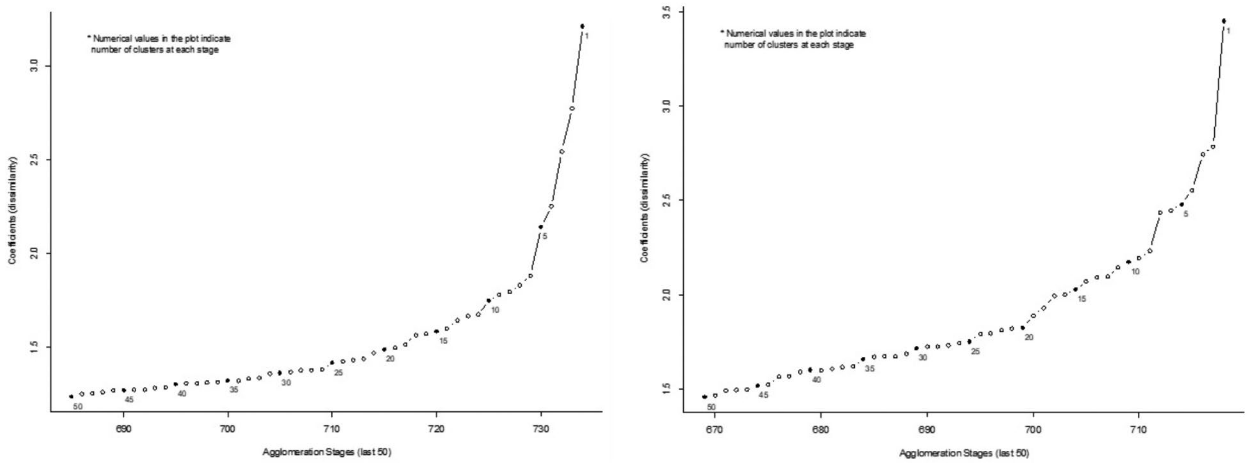

Fig. 1 Agglomeration plots produced by Hierarchal Cluster Analysis of the US (left) and Japan (right) datasets

Table 2 Keywords used in coding files

Category	English (US)	Japanese, translated (Japan)
Government policy	bill, legislation, lawmaker, proposal, Medicare, insurance, assistance, revenue, budget, spending, tax, income, unemployment, law, Act, stimulus, package, payment, rescue, relief, aid, mandate, executive order	policy, special measures law, law, legislation, criminal law, draft, government ordinance, budget, income, tax, support, unemployment, GoTo, Upper House, fine, imprisonment, regulation, handout, revision, measures, work hazard, penalty, Olympics
Resource shortages	demand, shortage, distribution, capacity, stockpile, production, toilet paper	toilet paper, production, supply, out of stock, material, shortage, collapse
Statistics	average, trend, surge, spike, wave, death toll, mortality, rise, count, fatality, figure, statistic	average, statistics, death toll, mortality, rise, fall, probability, spike, lethality
Measures to stem the spread of infection	quarantine, flight, ban, traveler, border, shutdown, stay-at-home order, lockdown	arrival at port, stay, isolation, entering the country, arrival and departure, lockdown, shutting borders, touch down, stay permit
Masks and transmission	cloth, covering, mask, face mask, droplet, particle	droplet, aerosol, mask, transmission route, conversation, infection source
Origin of the virus and resultant discrimination	gain-of-function, bat, Wuhan Institute of Virology, origin, discrimination, racism, manmade, tourist	artificial, bat, discrimination, bias, Wuhan laboratory, tourist
COVID-19 disease severity, treatment, or testing	ivermectin, hydroxychloroquine, chloroquine, remdesivir, malaria, therapy, cure, remedy, weed, swab, PCR, cold, flu, strain, severity, delta, omicron, antigen	green tea, PCR, ivermectin, treatment, negative, oxygen, cold, variant, hot water, delta, omicron, antibody
Vaccine efficacy, contents, or safety	VAERS, dose, shot, booster, mRNA, Pfizer/BioNTech, Pfizer-BioNTech, Pfizer, Moderna, side effect, adverse event, blood clot, allergic reaction, complication, heart, myocarditis, microchip, ingredient, breakthrough, vaccine, AstraZeneca, injection, anaphylaxis, immunity	vaccination, vaccine, arachnoid membrane, ovary, constituent, shot, DNA, gene, pregnant, adverse reaction, side effect, seizure, mRNA, Pfizer, Moderna, allergy, blood clot, heart, myocardium, microchip, prevention, anaphylaxis, causative relationship, infertility, breakthrough, AstraZeneca, immunity

the code frequencies across datasets, and *p* value correction was done using the Holm-Bonferroni method. Three categories stood out due to their relatively low *p* values and relatively high effect sizes: statistics, the origin of the virus and resultant discrimination, and COVID-19 severity, treatment, and testing.

Versions of Tables 2 and 3, and 4 with the original Japanese text are available in Supp_012024.docx.

The effect sizes ϕ for each category are shown below in Table 5. Only the category on the origin of the virus and resultant discrimination showed an effect size exceeding 0.1, a small effect. The two categories of statistics, and COVID-19 severity, treatment, and testing showed the next-highest effect sizes of >0.07. Hence, these three categories were chosen for further discussion.

Discussion

Similarities and differences between US and Japan categories

Selective reading of articles with high *tf* (term frequency) for the chosen categories produced a handful of similarities and differences. Within the statistics category (which was more common in the US dataset, 40.1% vs. 25.7%, ϕ 0.0792), misinformation from both countries tended to

Table 3 50 highest-frequency words for US and Japan (translated) datasets

US dataset (1,743 articles)				Japan dataset (148 articles)			
Word	**Freq.**	**Word**	**Freq.**	**Word**	**Freq.**	**Word**	**Freq.**
vaccine	8387	government	1366	infection	1047	symptoms	105
death	3897	testing	1333	vaccine	769	warning	104
case	3871	group	1164	vaccine (synonym)	527	America	101
virus	3768	hospital	1134	Japan	468	severe symptoms	99
state	3747	week	1106	test	350	approach	97
%	3528	month	1058	Tokyo	215	issue	97
U.S.	3250	use	1052	mask	214	risk	95
study	2807	COVID	1050	death	195	investigation	92
CDC	2630	way	1046	effect	179	nationality	90
pandemic	2529	school	1013	countermeasure	173	positive	89
datum	2312	official	1009	data	170	Pfizer	85
time	2280	Dr	1006	PCR	165	Taiwan	85
mask	2252	drug	990	healthcare	161	MHLW	84
country	2044	FDA	987	China	160	hospital	84
day	1912	flu	987	point out	148	Osaka	82
year	1869	trial	978	research	134	organization	81
disease	1853	response	974	patient	129	study	80
report	1819	outbreak	967	report	128	spread of infection	79
vaccination	1758	person	961	government	125	WHO	78
patient	1666	dose	935	overseas	122	effect	75
infection	1574	United States	932	situation	122	treatment	75
risk	1571	variant	916	immunity	119	influenza	74
child	1549	bill	905	prevention	119	world	74
China	1513	law	903	mayor	106	CDC	73
University	1470	company	891	news	106	cause	70

downplay the severity of the COVID-19 mortality rate, or otherwise make factually false statistical assertions. US misinformation tended to make more (invalid) comparisons to influenza, and there were false assertions that the US was performing statistically better in terms of mortality rate than other countries, while Japanese misinformation contained more assertions that vaccines increase mortality rate. Many of the US articles in this category were based on quotes from then-President Donald Trump.

Within the category regarding the origin of the virus and resultant discrimination (which was more common in the Japan dataset, 20.3% vs. 7.0%, ϕ 0.1311), misinformation from both countries asserted that COVID-19 was artificially made in the Wuhan Institute of Virology. However, US misinformation tended to focus on federal funding for the institute, and some articles tied the origin of the pandemic to Chinese meat-eating practices. Japanese misinformation focused more on Chinese people within Japan itself, such as warning of incoming tourist swarms or Chinese nationals taking up space in hospitals.

Within the category of COVID-19 severity, treatment, or testing (which was more common in the Japan dataset, 46.0% vs. 32.6%, ϕ 0.0756), both countries had misinformation about treatments for COVID-19, as well as about testing kits. While both countries mentioned ivermectin, hydroxychloroquine and marijuana as COVID-19 treatments were exclusive to the US dataset, while green tea and hot water were exclusive to the Japan dataset. More US articles tended to downplay the severity of infection by likening it to the flu. There were pieces of misinformation in the US that stemmed from misinterpretation of test kits, while there were Japanese assertions that COVID-19 test kits are faulty or ineffective.

Overall, non-health misinformation appeared more frequently than health misinformation, echoing findings from other studies analyzing fact-checking articles [9] or social media posts [20].

In addition, while the category frequencies for masks and transmission did not appear to differ, the contents of articles in these categories showed differences: articles from the US dataset tended to be regarding misinformation on the effectiveness of masks as a means for preventing transmission, while articles from the Japan dataset tended to be on ancillary topics, such as the country of manufacture of masks, or mask shortages. Mask-wearing as a means for preventing disease transmission while sick is an established aspect of Japanese culture [21].

Table 4 Code frequencies and examples (Japanese text translated)

Category	Term freq. %	χ^2 adjusted *p* value	Example statement
Government policy	720 41.3 (US)	0.712	"…this stimulus 'deal' (which) provides MORE funding to foreign governments and to American arts centers, than to the American people"
	52 35.1 (JP)		"Regarding imprisonment for refusing to be hospitalized, PM Suga explained that 'the National Governors' Association requested for the penalty'"
Resource shortages	322 18.5 (US)	0.776	"Although grocery store dairy shelves remain sparse, dairy farmers are being forced to dump thousands of pounds of milk down the drain"
	21 14.2 (JP)		"masks and toilet paper are made using the same materials"
Statistics	698 40.1 (US)	0.005	"Current survival rate for COVID19 in the US is 98.54%. Let's share this story. Positive vs. Panic"
	38 25.7 (JP)		"the death rate was almost the same in both vaccinated and unvaccinated groups"
Measures to stem the spread of infection	365 20.9 (US)	1.000	"The CDC can detain anyone with a fever 'indefinitely'. Vaccination is a way people could get out of detention"
	29 19.6 (JP)		"the Japanese government let in over 3000 Chinese nationals under a special scheme in April alone"
Masks and transmission	343 19.7 (US)	0.313	"N95 and surgical masks both provide 95% protection, while sponge and cloth masks offer none"
	39 26.4 (JP)		"wearing a mask during pregnancy decreases blood oxygen levels, causing the umbilical cord to shorten"
Origin of the virus and resultant discrimination	122 7.0 (US)	< 0.001	"COVID-19 started 'because we eat animals'"
	30 20.3 (JP)		"so there were 2 Japanese people on a plane that arrived from Wuhan on the 29th, apparently someone went home after refusing to be tested…"
COVID-19 disease severity, treatment, or testing	569 32.6 (US)	0.007	"NIH COVID Treatment Guidelines Approve Ivermectin"
	68 46.0 (JP)		"The PCR tests will also give positive results for other types of flu, so they don't work for detecting COVID-19"
Vaccine efficacy, contents, or safety	897 51.5 (US)	0.895	"The second booster has eight strains of HIV"
	77 52.0 (JP)		"vaccinated people spread the disease to everyone around them"
Miscellaneous	101 5.8 (US)	0.660	"…mobile COVID-19 testing station bears a logo that depicts an ancient deity of death"
	5 3.4 (JP)		"73 Japanese police officers infected while handling dead bodies"

National contextual factors that affect misinformation consumption

As outlined above, there are some differences in the contents of the COVID-19 misinformation circulating in the US and Japan. A few of the numerous contextual factors that may have influenced these differences will be described further below.

Importantly, it should not be assumed that a cause-and-effect relationship is at play, as a myriad of factors influence consumer (and macro-level) information-seeking habits. For instance, on the micro level, there are consumer culture factors that influence patterns of consumption, such as social influences or social class [22]; on the macro level, society-level factors such as the quality of official communications can affect attitudes towards health measures [23]. Some evidence also exists to suggest that in certain countries, the demand for certain kinds of misinformation fluctuates based on the epidemic curve [9]. While a comprehensive list of every potential influencing factor would be beyond the scope of this research, it can be seen that local context can indeed influence information-seeking habits. Understanding the concerns and mindsets of those grappling with the infodemic should be a priority in determining what countermeasures to take (e.g., targeted messaging, rapid response, etc.).

On the topic of the high prevalence of political figures involved in US misinformation, a survey conducted by

Table 5 Category χ^2 and effect sizes

Category	Country	(+)	(-)	χ^2	ϕ
Government policy	US	720	1,023	2.15	0.0337
	JP	52	96		
Resource shortages	US	322	1,421	1.69	0.0299
	JP	21	127		
Statistics	US	698	1,045	11.85	0.0792
	JP	38	110		
Measures to stem the spread of infection	US	365	1,378	0.15	0.0089
	JP	29	119		
Masks and transmission	US	343	1,400	3.77	0.0446
	JP	39	109		
Origin of the virus and resultant discrimination	US	122	1,621	32.50	0.1311
	JP	30	118		
COVID-19 disease severity, treatment, or testing	US	569	1,174	10.80	0.0756
	JP	68	80		
Vaccine efficacy, contents, or safety	US	897	846	0.02	0.0030
	JP	77	71		
Miscellaneous	US	101	1,642	1.50	0.0282
	JP	5	143		

the Reuters Institute for the Study of Journalism in 2020 [24] found that American information-seeking habits surrounding COVID-19 are strongly tied to political affiliation. Left-leaning respondents were likely to trust the news media and unlikely to trust the government; the opposite was true for right-leaning participants. Trump was himself a major direct source of COVID-19 misinformation [25], and many of the erroneous claims he made are reflected in the data, especially in the Statistics and Origin categories. The significant sway a person's political beliefs hold over their information-seeking behavior in the US is likely to be associated with the country's highly polarized political climate. This finding of the high frequency of misinformation from politicians in the US is echoed in the Zeng paper [9], and the same paper found that this connection between societal polarization and political misinformation was also clear in India.

In the Japanese dataset, articles pertaining to the origin of COVID-19 from China were much more frequent and pointed in general; as opposed to US articles which mostly addressed conspiracy theories of American funding for the Wuhan Institute of Virology or the animal origins of the virus, articles in this category in the Japan dataset tended to focus directly on Chinese nationals, either as disproportionate occupants of Japanese medical institutions, or as spreaders of COVID-19 inbound from China. Japan's relative geographical proximity to China and popularity as a Chinese tourist destination, as well as existing anti-Chinese sentiment that has been worsening progressively since the 1980s [26], may explain to some extent the personal nature of Japanese misinformation in this category.

At first glance, it may seem surprising that both the US and Japan have similar proportions of articles discussing vaccine efficacy, contents, or safety, especially given the heavy role US political figures played in leading supporters to act contrary to evidence-based findings [27]. In an article published in the Japanese journal *Chiryo* in 2021, the founders of HPV vaccine awareness group MinPapi describe how vaccine hesitancy in Japan may have been exacerbated by the human papillomavirus (HPV) vaccine side effect scare in 2013 [28]; years later, addressing vaccine hesitancy through their new website CoviNavi continues to be a challenge.

Additionally, a 2021 survey conducted in Japan showed that Japanese respondents were uncertain in general about what sources of COVID-19 information they could trust [20]. 24.7% of respondents believed there was no information source they could trust, and only 26.0% of respondents felt they could trust health experts. This stands in stark contrast to the results from the aforementioned Reuters study, where over 80% of American respondents on both sides of the political spectrum felt they could trust health experts. This difference in response to the infodemic – picking sides, as opposed to being assailed by uncertainty – may actually help to explain why vaccine misinformation is relatively common in both countries; one possible interpretation is that a limited segment of the American audience consumes vaccine misinformation in greater per capita amounts, while a more general segment of the Japanese audience consumes vaccine misinformation in lower per capita amounts.

Disinformation resilience and its effects on misinformation consumption

In a 2020 paper, Humprecht et al. outline a framework for cross-national comparisons of *disinformation* (henceforth "misinformation") *resilience*: the degree to which online misinformation is likely to receive exposure and be spread [29]. Political factors limiting misinformation resilience include societal polarization, and frequency of populist communication; media-related factors include low trust in news media, weak public news services, and audience fragmentation; economic factors include a large advertisement market size, and high social media usage. Using this framework in a comparison of the US with 16 other mainly European countries, the authors found that the US scored the lowest in misinformation resilience, owing to its fragmented media landscape, large ad market, low trust in news, highly polarized society, and frequent populist communication.

In comparison to the US, Japan scores notably lower in terms of populist communication [30]; NHK, the public broadcasting network, attains comparable viewership to other networks [31] as opposed to American public broadcasters with one- to two-thirds the viewership of major American TV networks [32, 33]; major TV news networks in Japan attain roughly two times the viewer share of US TV network providers, with Yahoo! News dominating the online news market with over 50% weekly usage [34]. While a formal comparison has yet to be done in the literature, these factors suggest that Japan may be more resilient to misinformation than the US. It is possible that this affected the sizes of the datasets that could be obtained, leading to the US dataset being more than ten times as large than the Japan dataset.

While it stands to reason that increased misinformation resilience would lead to lower spread and consumption of misinformation, its effect on the types of misinformation consumed is less clear. In the Zeng study [9], Germany stood out as one of the studied countries with high misinformation resilience; compared to the other countries which tended to contain high proportions of articles on political conspiracy theories, lockdown measures, or transmission methods, misinformation from Germany was centered on COVID-19 treatment and vaccines, similarly to the Japan dataset used in this report. If we consider the nature of rumors and misinformation as an answer-seeking response to a perceived external threat [35], one possible interpretation of this pattern is that increased misinformation resilience in the midst of the pandemic contributes to *lower distraction with non-key issues* – the *key issue* in this context being the health impact of COVID-19 and how it can be avoided or treated. The "Miscellaneous" category is mostly comprised of articles on these *non-key issues*, including those bordering on absurdity or conspiracy; while this category was not notably differently sized between the US and Japan datasets, the Japan data had a noticeably lower proportion of misinformation along the lines of the "deity of death" US article.

Strengths and limitations of this study

In comparison to prior studies which used fact-checking articles as data, this study uses a larger sample size for the US dataset and offers a Japanese dataset for the first time. In particular, using KH Coder allowed for multiple categories to be assigned to a single article, which reflects the data more accurately than other studies [9] that are limited to a single category for each article. Additionally, quantitative content analysis using KH Coder allowed for counting the term frequencies in the large datasets, as well as for referring back to the original data when needed using the KWIK Concordance function.

However, as to the limitations of the study, the span of misinformation covered in this report is limited to that selected by the editorial teams in a "gatekeeping" process [36] for the four online news sources used; in particular, fact-checking in Japan is a relatively new endeavor, with the InFact team and website notably smaller than established fact-checking organizations from the US. This has negative implications for the generalizability of the Japan data, and a larger future dataset would likely give richer results. In addition, since the categorization processes were carried out automatically, there may be a handful of data points that have not been categorized correctly. More studies should be done to further verify the relationship between the misinformation resistance of a country and the types of misinformation that spread within it. Future studies of this nature will have larger and more varied datasets to work with, whether they are about COVID-19 or any other infodemic. Finally, the effect sizes found for the sections discussed here are all of small magnitude, meaning that it should not be inferred that certain segments of misinformation should receive disproportionate amounts of focus in countries that seem vulnerable to that kind of misinformation.

Practical implications

In combination with aggregated data from other countries, data on the types of misinformation which are comparatively common in the country provides policymakers a reference point when allocating resources to tackling misinformation, through means such as rapid-response messaging [37]. Of course, this data should be weighed against the actual likely impact of said misinformation spreading in the populace; any given piece vaccine misinformation is likely to do more harm overall than a wild claim of a vaccination center bearing a logo of a "deity of death".

This research also opens up new avenues for further research – for instance, research to verify whether modifying our taking a culturally-relevant approach to tackling misinformation results in better correction outcomes. One possible example would be altering the tone of messaging to be firmer and more succinct in an environment like Japan, where misinformation likely spreads out of uncertainty instead of certainty in misinformation, while a more indirect approach may be more effective in places like the United States where misinformed beliefs are grounded in certainty.

Conclusion

Using quantitative content analysis, this study shows the similarities and differences in the COVID-19 infodemics in US and Japan since the start of the pandemic. Differences were found in the proportion of articles mentioning statistics, the origin of the virus and resultant discrimination, and COVID-19 severity, treatment and testing, though the effect sizes were seen to be small.

Several facets of national context appear to support the trends seen in the data, such as the history of the HPV vaccine in Japan leading to increased distrust of COVID-19 vaccines. In addition, application of a misinformation resilience framework appears to show that in countries with higher resilience, distracting *non-key issues* such as conspiracy theories attract less attention compared to *key issues*, which refer to COVID-19 health impacts and other health information in the context of the pandemic. Understanding the types of misinformation in circulation gives policymakers and educators direction in developing strategies to counter this misinformation.

Lastly, it should be reiterated that fact-checking, even when done through appropriate channels in a culturally relevant manner, cannot be relied upon as the sole measure with which to combat an infodemic. Not only does fact-checking have heavily limited effects on correcting misinformed beliefs [4, 5], a deluge of fact-checking information may even backfire by contributing to information overload and avoidance in the intended audience [38], or by simply acting as a dissemination channel for the misinformation that would not have been spread otherwise [36]. Fact-checking has a place as one of the pillars of infodemic management – there is a need to uphold journalistic integrity, and to provide a reliable source for a more invested, informed reader subset. The other pillars of infoveillance and infodemiology, the gradual process of building eHealth literacy in the populace, and providing clear, timely translations of scientific findings to actionable messages need to be upheld in tandem as a long-term strategy for decreasing the impact of misinformation [3].

Abbreviations

COVID-19	Coronavirus disease 2019
HPV	Human papillomavirus

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-19813-y.

Supplementary Material 1: This material contains the chi-squared test assumptions, Holm-Bonferroni adjusted *p* values used in the results, and untranslated Japanese text for Tables 2 and 3, and 4

Acknowledgements
Not applicable.

Author contributions
This paper was written in whole by MS. The manuscript was reviewed by Associate Professor MI. Both the authors read and approved the final manuscript.

Funding
This research was supported by operating funds allocated by Kyoto University to the Department of Medical Communication.

Data availability
The dataset supporting the conclusions of this article is available in the GitHub repository, https://doi.org/10.5281/zenodo.8282744 at https://github.com/seahmatthew/KyotoU-PublicHealth2023 [39].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 7 September 2023 / Accepted: 16 August 2024
Published online: 09 September 2024

References

1. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods To Analyze Search, communication and publication behavior on the internet. J Med Internet Res. 2009;11(1):e11.
2. Organization WH, Managing. the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation 2020 [updated 23 September 2020. https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation
3. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res. 2020;22(6):e21820.
4. Young DG, Jamieson KH, Poulsen S, Goldring A. Fact-checking effectiveness as a function of format and tone: evaluating FactCheck.org and FlackCheck.org. Journalism Mass Communication Q. 2018;95(1):49–75.
5. Chou W-YS, Gaysynsky A, Vanderpool RC. The COVID-19 Misinfodemic: moving beyond fact-checking. Health Educ Behav. 2021;48(1):9–13.
6. Foucalt M, Power/Knowledge. Selected interviews and other writings 1972–1977. Great Britain: Harvester, Limited;; 1980.
7. Chandler C, Fairhead J, Kelly A, Leach M, Martineau F, Mokuwa E et al. Ebola: limitations of correcting misinformation. Lancet. 2014;385(9975).
8. Yagi A, Ueda Y, Nakagawa S, Ikeda S, Tanaka Y, Sekine M et al. Potential for cervical cancer incidence and death resulting from Japan's current policy of prolonged suspension of its governmental recommendation of the HPV vaccine. Sci Rep. 2020;10(1).

9. Zeng J, Chan C-H. A cross-national diagnosis of infodemics: comparing the topical and temporal features of misinformation around COVID-19 in China, India, the US, Germany and France. Online Inf Rev. 2021;45(4):709–28.
10. Verbeke R, Lentacker I, De Smedt SC, Dewitte H. The dawn of mRNA vaccines: the COVID-19 case. J Controlled Release. 2021;333:511–20.
11. Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. Lancet. 2021;398(10317):2126–8.
12. Politifact. Latest Fact-Checks In Coronavirus. https://www.politifact.com/factchecks/list/?page=1&category=coronavirus
13. FactCheck.org. Issues: coronavirus. https://www.factcheck.org/issue/coronavirus/
14. Japan B. Fact Check. https://www.buzzfeed.com/jp/badge/factcheckjp
15. InFact. COVID-19 FactCheck. https://infact.press/tag/covid-19-factchecks/
16. Hamborg F, Meuschke N, Breitinger C, Gipp B, editors. news-please: A Generic News Crawler and Extractor. 15th International Symposium of Information Science; 2017.
17. Higuchi K. A two-step Approach to quantitative content analysis: KH coder tutorial using Anne of Green Gables (Part I). Ritsumeikan Social Sci Rev. 2016;52(3):77–91.
18. Higuchi K. Research done using KH Coder 2023. https://khcoder.net/bib.html
19. Higuchi K. KH Coder 3 Reference Manual. 2017.
20. Ishibashi M, Sekiya N. Exploratory study on the spread of rumors and psychological factors related to COVID-19. Japanese J Risk Anal. 2021;31(2):123–32.
21. Burgess A, Horii M. Risk, ritual and health responsibilisation: Japan's 'safety blanket' of surgical face mask-wearing. Sociol Health Illn. 2012;34(8):1184–98.
22. Hoyer WD, MacInnis DJ, Pieters R. Consumer Behavior. 6 ed. Cengage Learning; 2012.
23. Nara Y. The role of Risk Communication in COVID-19 Public Health Management. Infect Control: Japanese J Infect Control. 2021;30(9):958–61.
24. Nielsen RK, Fletcher R, Newman N, Brennen JS, Howard PN. Navigating the 'Infodemic': How People in Six Countries Access and Rate News and Information about Coronavirus. 2020.
25. Editors NEJoM. Dying in a Leadership Vacuum. N Engl J Med. 2020;383(15):1479–80.
26. Sekiyama T. Japan's policy toward China under strong anti-chinese sentiment: a case of terminating yen loans to China. East Asia. 2012;29(3):295–311.
27. Adams A, Hackstadt A, editors. Distrust in Institutions: Reference and Library Instruction During an Infodemic. DGO2021: The 22nd Annual International Conference on Digital Government Research; 2021 2021-06-09: ACM.
28. Kinoshita T. How should healthcare professionals reach the apathetic populace: on my experiences running MinPapi/Cov-Navi. Chiryo. 2021;103(12):1527–31.
29. Humprecht E, Esser F, Van Aelst P. Resilience to online disinformation: a Framework for cross-national comparative research. Int J Press/Politics. 2020;25(3):493–516.
30. Hawkins KA, Aguilar R, Castanho Silva B, Jenne EK, Kocijan B, Rovira Kaltwasser C. Global populism database. 2 ed. Harvard Dataverse; 2019.
31. VideoResearch. Weekly Top 10 Most-Watched TV Shows: VideoResearch; 2022. https://www.videor.co.jp/tvrating/
32. Center PR. Public Broadcasting Fact Sheet 2021. https://www.pewresearch.org/journalism/fact-sheet/public-broadcasting/
33. Center PR. Cable News Fact Sheet 2021. https://www.pewresearch.org/journalism/fact-sheet/cable-news/
34. Newman N, Fletcher R, Robertson CT, Eddy K, Nielsen RK. Reuters Institute Digital News Report 2022. Reuters Institute; 2022.
35. Difonzo N, Bordia P, Rumor. Gossip Urban Legends Diogenes. 2007;54(1):19–35.
36. Lwin MO, Lee SY, Panchapakesan C, Tandoc E. Mainstream News Media's role in Public Health communication during crises: Assessment of Coverage and correction of COVID-19 misinformation. Health Commun. 2021;38(1).
37. Scales D, Gorman J, Jamieson KH. The Covid-19 infodemic — applying the epidemiologic model to Counter Misinformation. N Engl J Med. 2021;385(8):678–81.
38. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara S. -e. from information seeking to information avoidance: understanding the health information behavior during a global health crisis. Inf Process Manage. 2021;58(2).
39. Seah M, Iwakuma M. KyotoU-PublicHealth2023-MatthewSeah. 2023.

Publisher's note