Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

12 Studying Social Media and Journalism with Publicly Available Cross-national Secondary Data

Michael Chan

Introduction

Adapting to evolving audience behaviours brought by new communication technologies has long challenged journalism and its practitioners. The global diffusion of social media, significantly, has fundamentally altered the production, distribution, and consumption of news in societies around the world (Waisbord, 2019). These developments, in turn, have led to thriving interdisciplinary research with perspectives drawn from such fields as communication, sociology, and political science, which examined the intersection of social media, journalism, and the audience (Steensen & Westlund, 2021). This chapter focuses on the utility of publicly available cross-national secondary survey data intended to be shared with the public and how researchers can utilize such data to build on the literature examining the dynamics of social media and journalism.

Secondary data analysis is the analysis of existing data already collected by others (Donnellan & Lucas, 2013). It is a well-established research approach in the social sciences, going as far back as Durkheim's study on suicide based on public data (Pickering & Walford, 2000). The benefits of using secondary data for researchers are obvious. It saves them time and financial resources if data has already been collected that can be used to address their research questions. There are two ways to obtain secondary data. The most direct is requesting it directly from the author(s) of a published study. This is rare for utility reasons (e.g., the data is specific to that study, and it is already published). The common way is to obtain secondary datasets intended to be shared with other researchers. The scale and scope of such data are often much larger than what a single researcher can typically collect (Donnellan & Lucas, 2013), and the quality of the data is derived from rigorous sampling and data collection techniques that provide representative country samples. Moreover, as the data are intended to be shared, they are usually already formatted and "cleaned" and accompanied by extensive documentation such as codebooks and questionnaires that help the researcher understand the theoretical and organizational foundations of the datasets and how they were operationalized and measured.

DOI: 10.4324/9781003568766-16

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

Many of the publicly available datasets are also cross-national as they comprise data from multiple countries (See Appendix). Researchers can, of course, use only data from a single country relevant to their own research. In this chapter, the focus is more specifically on secondary data analysis across *multiple* countries. As noted by comparative communication researchers, obtaining research findings from multiple countries can attest to the generalizability of theories previously derived from single countries. Researchers can also examine the extent to which the objects of analysis (i.e., independent and dependent variables) are shaped by country-level conditions. Analyses of this kind draw important theoretical attention to macro-level structures that can explain the similarities and differences in the dynamics of media and communication phenomena among societies (Esser & Hanitzsch, 2012). For example, Chan and Yi's (2024) multi-country analysis of the relationship between people's social media use for political information and their perceived quality of democracy in their country showed that the relationship was conditioned on the degree of affective polarization at the country level. At lower levels, it was positive, which is a normatively desirable outcome. But, at higher levels of polarization, the relationship was negative, thus pointing towards the potential for democratic backsliding in the long term. The divergent relationship resembled what Holbert and Park (2020) call "cleaved moderation," which makes a more substantive theoretical contribution than other types of moderation.

Despite the availability and utility of such data to generate new knowledge, there are few go-to learning or reference resources available to scholars considering using cross-national secondary data for their research. In this chapter, I draw from my own published work and those of others to provide an overview of some of these data sources and how they have been applied in social media and journalism scholarship, such as the Reuters Institute Digital News Report (DNR), World Values Survey (WVS), and Varieties of Democracy (V-Dem), among others. I elaborate on the opportunities and challenges researchers face when using these data sources (i.e., theoretical, conceptual, operational) and implications for research design and analysis, such as multilevel modelling for analysing clustered data. I also highlight the technical aspects of handling large datasets and running statistical analyses using them, which can be computationally intensive.

Overall, this is not a "how to" chapter on deriving research questions from cross-national secondary data. Rather, it presents a series of concrete considerations researchers should consider when selecting and using cross-national secondary data for social media and journalism studies.

Cross-National Secondary Data: Issues and Considerations

Data Fit and Data Discovery Approaches to Knowledge Creation

Primary and secondary research have the same starting point, which is a set of research questions and tentative hypotheses that a researcher wants to address

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

(Donnellan & Lucas, 2013). For secondary research, the next step involves searching for datasets that provide the relevant variables for statistical analysis. Keyword searches are a straightforward way to search for cross-national secondary data, though several datasets already have well-established reputations. For example, the WVS, which covers 80,000+ respondents across 70+ countries, is very well known among those engaged in public opinion, political science, and communication, and it has been used to explore the relationship between media use and a range of political and social attitudes, such as support for democracy (Goidel et al., 2017) and anti-science perceptions (Mede, 2022). This can be considered a deductive approach, seeking available data to fit pre-existing research questions.

Alternatively, researchers can use secondary data inductively to explore and uncover new concepts and insights. For example, Lee et al. (2019) developed the concept of "consumptive news feed curation" based on Reuters DNR question items that tapped into how news consumers personalized their social media news feeds. His analysis across six East Asian countries showed that curation was positively related to news engagement in all countries. Using the Worlds of Journalism Study (WJS), which surveyed 27,000+ journalists in 65 countries, Hamada (2021) used two question items to develop a "professional autonomy" scale, which was found to be positively correlated with Freedom House's press freedom score. Toff and Kalogeropoulos (2020) used a DNR question that was asked to over 67,000 respondents across 35 countries on the extent to which they were "actively trying to avoid the news these days." They theorized that news avoidance should not be conceptualized simply as a low level of news consumption. Instead, it can also be a purposefully strategic behaviour with different antecedents and consequences from general news consumption. These are just a few examples demonstrating the potential for researchers to use cross-national secondary data to generate new concepts and insights and answer pertinent research questions related to the antecedents and consequences of social media and journalism, whether through a deductive or inductive approach or a combination of the two.

Availability of Question Items for Social Media and Journalism Research

There are two main types of cross-national secondary data available to researchers related to social media and journalism, and the social sciences more generally, and the majority of them are based on survey data from individual respondents. The first type is well-established, large-scale regional or global public opinion surveys focusing on citizens' attitudes and beliefs on society and politics. These surveys also contain a subset of news and media-related questions that can be used to study the aspects of social media and journalism. Examples include the *WVS*, *Asian Barometer Survey*, *Afrobarometer*, *Eurobarometer*, and *Latinobarómetro*. Questions on social media generally focus on whether respondents use any of the platforms and for what purposes.

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

Notably, *Latinobarómetro* also asks about attitudes towards social media, such as whether it makes politics "better" or "worse." Some surveys also contain relevant questions on the normative role of the press, such as whether "The news media should constantly investigate and report on government mistakes and corruption" (Afrobarometer); "Freedom of expression should be guaranteed even if it means circulation of fake news" (Latinobarómetro); while others ask about media bias: "TV news favors the governing party" (World Values Survey). The second type is surveys explicitly focused on news and journalism, such as the *Reuters Institute Digital News Report* and the *WJS*. The DNR features a deep and diverse set of questions on individual news consumption and social media news engagement, supplemented by topical questions that change every year. For example, in the 2023 report, a subset of questions asked whether and where respondents have been exposed to discourses criticizing journalists. In the 2024 report, respondents were asked about their comfort levels with AI-generated news. The WJS differs from the other surveys because the respondents are working journalists rather than general citizens. So, the questions are more targeted towards assessing the role of journalists and their working conditions and perceptions of the environment for journalism.

Both types of surveys can provide useful cross-national data for exploring social media and journalism-related phenomena. However, researchers will often find that their proposed variables can only be measured with a single question item. Therefore, it is not uncommon for discerning reviewers to question the reliability and validity of key variables derived from secondary data. For example, in Chan and Yi (2024), the core independent variable for the study "social media use for information" was a binary Yes/No measure as the original answer choices from the WVS were "Have done," "Might do," and "Would never do." A reviewer noted this core measure was a "very significant weakness" and that there was a "huge difference between people who have used social media for political information once and those who regularly consume all their information this way." Our rationale in the response letter, which convinced the reviewer, was that 70% or more of respondents did not use social media for political information at all in 19 of the 27 countries. Even if a continuous measure of social media questions were available, it would only apply to small subsets of social media users in those countries. Therefore, a binary measure was still appropriate for studying the general population, which was the focus of the study. This example exemplifies that while the limitations of secondary measures are often (and should be) stated in the limitations section of the discussion, researchers can also build a strong argument for the use of their measures in the methodology section by contextualizing the variable from a cross-national perspective. After all, the use and non-use of communication technologies and exposure to or not to certain channels for news can vary widely across countries.

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

The Challenge of Country Selection

An important consideration for researchers utilizing cross-national datasets is deciding which countries to include in the analysis. An intuitive approach might be to include all countries based on data availability. However, more discerning reviewers might demand that country selection be justified on theoretical grounds. Conversely, researchers who use only a subset of countries in their study might be queried by reviewers familiar with the dataset why they only selected those countries.

Using a well-established media-related typology or index is a common approach that combines a theoretical framework and justification for subset country selection. Perhaps the most popular is the comparative media systems typology by Hallin and Mancini (2004), which categorized 18 Western democracies along different structural dimensions, resulting in the classification of the media systems as following the polarized pluralist, democratic corporatist, or liberal models. Thus, in their study on the extent to which countries in Europe (and the United States) are resilient to misinformation, Humprecht et al. (2020) took a subset of the same 18 countries from the DNR dataset as their study sample. Kalogeropoulos et al. (2017) used the same media systems typology but selected only six countries from the DNR dataset, though they ensured that each media system was represented by two countries. Also using DNR data, Fletcher et al. (2020) conducted a 12-country study on polarized news behaviour among audiences. While nine countries were selected based on the media systems typology, the authors also added three more countries not originally in the typology (Australia, Poland, and Czechia) based on their arguments that the countries shared common characteristics with others in the typology.

Outside of Europe, other media typologies can be used to explain and justify sample selection. For example, using DNR data, Chan et al. (2019) conducted a comparative study of social media news engagement among news consumers in six Asian markets (i.e., Taiwan, Japan, Korea, Hong Kong, Malaysia, and Singapore). As well as being located in Asia, another important criterion for country selection was that the countries represented different levels of press freedom, such that two of them were "Free," "Partly free," and "Not free," respectively, based on the Freedom of the Press report.

Apart from media typologies and indices, other studies justified country selection based on their theoretical research questions. Hamada et al. (2024) compared journalism practices between 24 Western-majority and Muslim-majority countries; so, it was logical for them to select a subset of Western and Muslim countries from the WJS dataset for their analyses. Nisbet et al. (2012) combined data from the Afrobarometer and Asian Barometer Surveys because they wanted a sample with greater heterogeneity of political systems ranging from authoritarian states to developing democracies that were particularly suited for the authors' research focus on citizens' demands for democracy.

Thus, data availability can be a consideration for secondary research when deciding which countries to include in a study, but it should not be the most

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

important one because reviewers will always scrutinize the logic behind country selection for cross-national studies. Instead, it should be driven by theoretical considerations as new knowledge derived from cross-national comparative research is more meaningful when "the objects of analysis are compared on the basis of a common theoretical framework" (Esser, 2013, p. 115).

Selecting an Analysis Strategy

After country selection and confirmation of which measures to use in the analysis, the next important decision for researchers is the analytical strategy. Regression analysis is familiar to all researchers who use quantitative methods as it is the common statistical approach to modelling how a battery of independent variables predicts a dependent variable. Most researchers are also well-versed in the variants of regression statistics, such as linear, logistic, and Poisson regression, respectively, for continuous, binary, and count outcomes. For analysing cross-national data, however, there are additional considerations as there are different ways in which the comparative dimension can be incorporated into the regression model, as shown in Table 12.1. Which approach a researcher takes can depend on several factors, such as the research focus and the number of countries in the analysis.

Using Approach 1, Chan et al. (2021) used 2018 DNR data to test separate regression models for each of the seven countries to predict social media news engagement. Among the findings were that people who followed or subscribed

Table 12.1 Analytical Approaches for Cross-National Comparative Research

Approach	*Country Samples*	*Level of Analysis*	*Regression Model Specification*	*Example Studies*
1	Separated	Individual	Single-level regression model for each country	Chan et al. (2021) Fisher et al. (2019)
2	Combined	Individual	Single-level regression model with country as a factor or moderator	Kalogeropoulos et al. (2017) Fletcher and Nielsen (2018)
3	Combined	Multilevel	Multilevel regression with individual and country-level predictors	Humprecht et al. (2020) Toff and Kalogeropoulos (2020)
4	Combined	Multilevel	Multilevel regression with individual and country-level predictors plus cross-level interactions	Chan et al. (2022)

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

to news organizations, and journalists on social media were more likely to engage with social media use across all seven countries. Statistically, the country effects were absorbed by the intercept of each model, and the regression estimates were specific to each country. Therefore, this can be considered a descriptive comparison approach because it summarizes the overall findings across different countries, but it does not directly test cross-national differences statistically.

In contrast to the above is the country-as-factor approach (Approach #2), in which country differences are explicitly measured. Kalogeropoulos et al. (2017) combined data from six countries in their study on social media news engagement using the 2016 DNR data. Then, they dummy-coded each country as a categorical variable using the k-1 method and included them in the regression models with the United Kingdom as the baseline category. Among the findings was that respondents in the United States, Italy, and Spain, on aggregate, were significantly more likely to comment on news through social media compared to those in the United Kingdom. This approach thus considers the variance of the dependent variables explained by country. The four-country study by Fletcher and Nielsen (2018) extended this by adding interaction terms to examine how countries moderated the relationship between incidental exposure and the number of online news sources accessed by respondents. Their findings showed that the relationship was stronger in the United Kingdom than in Australia and Italy. Because there can only ever be one reference group, studies that use this approach must change the reference category and rerun the regression model multiple times to accommodate all pairwise comparisons among the countries.

More ambitious comparative studies are not only interested in differences among countries, but they also want to examine how specific media and communication-related processes or outcomes at the individual level are shaped and contextualized by specific structural conditions at the country level (Esser, 2013). Research of this kind typically uses a multilevel approach, also known as hierarchical or mixed-effects modelling, because it accounts for the clustering of individuals under higher-order country-level characteristics. Using such an approach also provides opportunities to combine secondary datasets containing data for the same countries. The "resilience to online disinformation framework" tested by Humprecht et al. (2020) using Approach #3 is an illustrative example. They proposed that certain structural characteristics of countries make them more "resilient" to disinformation, such as higher media trust and funding of public service media, along with lower levels of populism and polarization, among others. To create the 18-country dataset, they combined data from multiple secondary sources, such as V-Dem for some variables and the Global Populism Database (GPS) among others from the same 18 countries. To examine how these variables predicted users' perceived exposure to disinformation, they combined the country-level data with individual-level data from the 18 countries contained in the DNR. The final dataset thus comprised individuals in the 18 countries clustered within the structural variables

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

of the same countries. Put simply, each dataset "contributes" its own set of variables to the combined dataset and is merged by "country." Among the findings of Humprecht et al. (2020) was that lower media trust at the country level was related to greater exposure to disinformation at the individual level. This contributed to the theoretical argument that countries with higher media trust are more resilient to disinformation than countries with low media trust, as individuals in those countries are less exposed to misinformation.

Toff and Kalogeropoulos (2020) also combined secondary data from multiple sources, including individual-level data of 36 countries from the Reuters DNR and country-level data for the same countries from the press freedom index (Reporters Sans Frontières) and political freedom index (The Economist) to generate a battery of individual-level and country-level predictors on people's tendency to avoid the news. The findings showed that while individual-level variables such as age, gender, and ideology predicted news avoidance, country-level variables such as lower press freedom and political freedom also accounted for a substantive portion of the variance in explaining news avoidance.

Researchers wanting a more direct analysis of how contextual factors influence individual-level processes would include one or several cross-level interactions in their models (Approach #4). This follows the same logic as moderation analysis for conventional regression, which examines how the relationship between two variables varies at different levels of the moderator. For multilevel analysis, the emphasis is typically on how individual-level relationships differ at different levels of the country-level factor. For example, Chan et al. (2022) combined individual-level WVS data with country-level V-Dem data to examine how the relationship between individuals' online news use and their political participation is contingent on levels of online surveillance and censorship across 44 countries. Their findings supported the "informational theory of repression" proposed by the previous literature, which was demonstrated by a three-way cross-level interaction where the relationship between online news and political participation was stronger under the conditions of high surveillance and low censorship.

The above examples show how different regression approaches can address a variety of cross-national comparisons derived from secondary data. Practically, specific approaches are more suited than others based on the number of countries. For example, the country-as-factor approach would not be practical for a 20-country sample as this would entail 19 interaction terms, which makes interpretation of the findings more challenging.

Multilevel Modelling, Secondary Data, and Reviewer Feedback

As noted earlier, reviewers will most probably ask for justifications behind the country selection for cross-national comparative studies if they are not in the manuscript. There may also be feedback that is specific to multilevel models and analysis. For example, there have been criticisms by some scholars on the

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

use of cross-national secondary data that was derived from qualitative judgements, such as V-Dem, the Polity Project, and the Global Party Survey, among others, because such judgements were susceptible to personal biases (Little & Meng, 2023). For example, V-Dem relies on "Country Experts" to quantify whether a country is "polarized" or not, rather than through representative population surveys. Others counterargued that aggregating responses from representative surveys of respondents also leads to bias, as such responses were also judgement-based (Knutsen et al., 2023). These disagreements and debates can sometimes spill over into the review process, as a reviewer from one side of the argument may question the choice of variables and their validity. Such was the case of Chan and Yi (2024), where the two core country-level moderators (affective and issue polarization) were derived from V-Dem data. This necessitated a long explanation in the response letter, which ended up in the online supplementary file. In the first step, we summarized the ongoing scholarly debate on the use of expert coders for country-level indices. Then, we performed tests of construct validity by correlating the study measures with other measures, such as the Comparative Study of Electoral Systems (CSES) survey, which is based on aggregated individual-level data. This extra work spanning five pages eventually satisfied the reviewer.

In other fields like political science and sociology, where multilevel modelling for comparative studies is more prominent, there have also been technical discussions on what is the minimum number of countries required in a comparative study in order to obtain reliable estimates (Stegmueller, 2013). For example, Bryan and Jenkins (2016) recommended 25 countries for linear regression and 30 countries for logistic regression. This could be another potential issue raised during the review process. Therefore, researchers should have a clear understanding of the methodologies behind the secondary data sources and anticipate challenges to their validity from peer reviewers.

Learning Multilevel Modelling

The examples of multilevel analysis above exemplified the potential of combining multiple publicly available cross-national datasets from different levels of analysis to propose new theoretical frameworks and test existing theories (see also Chapter 10 of this handbook). Despite its promise, the learning curve of the technique can be steep, even for researchers who are already well-versed in conventional linear regression. More than three decades ago, communication scholars already noted that "all individual mass communication behavior is performed in a social, cultural, institutional, and organizational context" (Pan & McLeod, 1991, p. 148). Yet, despite intermittent articles that highlighted the benefits of multilevel analysis for communication scholarship (i.e., Slater et al., 2006), it has not gained much traction in the field. Thus, multilevel regression is unlikely to be taught in graduate-level research programs in media and communication. This means that self-learning is the most likely path to learning the technique.

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

There are several introductory texts on multilevel regression with varying levels of technical detail. A good introductory text in its third edition is *Multilevel Analysis: Techniques and Applications* (Hox et al., 2017). *Multilevel Modelling* (Garson, 2020) is perhaps the most user-friendly and accessible text as it provides a general overview of the logic of multilevel modelling that is accompanied by step-by-step visual examples showing readers how to conduct the tests using the most common statistical software, including Stata, SPSS, SAS, R, and HLM. Most researchers in the social sciences are familiar with at least one of these software programs, so the text allows researchers to learn multilevel regression by doing. Moreover, as the same tasks are repeated across the different software programs, readers can compare the relative strengths of each software. For example, HLM, unsurprisingly, has more advanced features because the software was explicitly designed for hierarchical linear modelling. Hayes (2006) has also written a short and accessible primer on multilevel modelling for SPSS users.

Another important consideration with regard to software is that secondary data files can be much larger than a typical dataset that a researcher handles. For example, the World Value Survey csv file is over 170 MB, and the full V-Dem SPSS dataset is over 950 MB. Therefore, even relatively straightforward data-wrangling tasks such as filtering out a subset of the sample or recoding a variable can take a very long time to complete in programs like SPSS. For such users, one way around this is to import the SPSS data file into R and then use R to perform all necessary data cleaning and transformation tasks before generating a new SPSS .sav file and conducting the analysis. An alternative is to use R to conduct the analyses using the well-established *lme4* or *glmm* packages. While it may take some time to learn the R syntax, the relative advantage of using R is not only its speed in completing data wrangling tasks and various functions for running analyses for large multilevel regression models but also its flexible ways to visualize the findings that are not possible in other programs, such as cross-level interactions. The WVS and V-Dem survey, among others, also provide their data files in R format. There are various ways to learn multilevel regression for R, including online tutorials (e.g., Vacca, 2022) and short primers in journals (e.g., Brown, 2021). Another great resource is from published cross-national comparative studies that have "open materials" badges where the authors share their R scripts from a repository (e.g., Chan & Yi, 2024). By downloading the scripts and data, researchers can directly replicate the published findings and, in the process, learn the logic and workings behind the R code.

Key Takeaways and Lessons Learned

Cross-national comparative analysis is extremely valuable for theory building and testing as it allows researchers to contextualize their findings according to macro-level structures and influences. This helps to explain the patterns of findings within countries. As Esser (2013) noted: "It provides a contrast to

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

naïve universalism, countering the tendency to presume that communication findings from one's own country also apply to other countries" (p. 113). Most social scientists, however, lack the resources to independently collect representative cross-national sample data. Such large-scale efforts are typically undertaken by well-funded research institutes and teams. Therefore, utilizing cross-national secondary data allows researchers to explore compelling research questions and ideas with high-quality data from multiple countries, through either deductive or inductive approaches. Such data can potentially be used to address various aspects of social media journalism from the perspective of news producers, news consumers, and the work of journalists (see Hendrickx & Opgenhaffen, 2024). For example, the 2024 DNR included a question asking respondents how easy or difficult it was to discern trustworthy and untrustworthy news and information from different social media platforms such as Facebook, Instagram, and LinkedIn. One can reasonably assume that variables at the individual level such as education and media literacy improve discernment. There can also be potential explanations at the macro level. For example, structural factors such as a polarized news information environment or the online reach of public service media could also have a role. Research questions like this can be answered by combining individual and country-level data from multiple secondary data sources together with a theoretically informed framework and a strong rationale for country selection and comparison.

As discussed in this chapter, conducting cross-national comparative analyses using publicly available data presents several unique challenges. The most pressing issue is methodology. Although comparative communication is a well-established subfield dating back to the 1990s, the literature on conducting such analyses is sparse and scattered. Consequently, communication and journalism scholars often turn to other fields to acquire the necessary methodological knowledge and skills. Hopefully, this will change as journalism and the work of journalists continue to evolve within the rapidly changing information environment that shapes how people consume news, and cross-national analyses can help us understand the dynamics of such outcomes.

Appendix: Examples of Cross-national Secondary Data Datasets

Access and Use

While the datasets listed are intended to be shared with other researchers and are relatively easy to download, they have different guidelines and rules on access, conditions of use, and referencing. Generally, this means that the data can only be used for research or non-profit purposes, uses of the data should be cited or acknowledged, and the data files cannot be shared with others. The latter condition has implications for open-science practices. For example, researchers who use WVS and V-DEM data cannot store the data files in their own repositories.

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

Individual-level

Afrobarometer: https://www.afrobarometer.org
Asian Barometer Survey: https://www.asianbarometer.org
Eurobarometer: https://www.gesis.org/en/eurobarometer-data-service
Latinobarómetro: https://www.latinobarometro.org
Reuters Institute Digital News Report: https://www.digitalnewsreport.org
Worlds of Journalism Study: https://worldsofjournalism.org
World Values Survey: https://www.worldvaluessurvey.org

Country-level

Varieties of Democracy (V-Dem): https://www.v-dem.net
Global Party Survey: https://www.globalpartysurvey.org
World Press Freedom Index: https://rsf.org

References

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*, 1–19. https://doi.org/10.1177/2515245920960351

Bryan, M. L., & Jenkins, S. P. (2016). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review*, *32*, 3–22. https://doi.org/10.1093/esr/jcv059

Chan, M., Chen, H.-T., & Lee, F. L. F. (2019). Examining the roles of political social network and internal efficacy on social media news engagement: A comparative study of six asian countries. *The International Journal of Press/Politics*, *24*, 127–145. https://doi.org/10.1177/1940161218814480

Chan, M., Lee, F. L. F., & Chen, H.-T. (2021). Examining the roles of multi-platform social media news use, engagement, and connections with news organizations and journalists on news literacy: A comparison of seven democracies. *Digital Journalism*, *9*(5), 571–588. https://doi.org/10.1080/21670811.2021.1890168

Chan, M., & Yi, J. (2024). Social media use and political engagement in polarized times. Examining the contextual roles of issue and affective polarization in developed democracies. *Political Communication*, *41*(5), 743–762. https://doi.org/10.1080/10584609.2024.2325423

Chan, M., Yi, J., & Kuznetsov, D. (2022). Government digital repression and political engagement: A cross-national multilevel analysis examining the roles of online surveillance and censorship. *The International Journal of Press/Politics*, *29*, 371–393. https://doi.org/10.1177/19401612221117106

Donnellan, M. B., & Lucas, R. E. (2013). Secondary data analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis*, edited by T. D. Little (pp. 665–677) Oxford University Press.

Esser, F. (2013). The emerging paradigm of comparative communication enquiry: Advancing cross-national research in times of globalization. *International Journal of Communication*, *7*, 113–128.

Esser, F., & Hanitzsch, T. (2012). On the why and how of comparative inquiry in communication studies. In *Handbook of Comparative Communication Research*, edited by F. Esser & T. Hanitzsch (pp. 3–22). Routledge.

Fisher, C., Culloty, E., Lee, J. Y., & Park, S. (2019). Regaining control citizens who follow politicians on social media and their perceptions of journalism. *Digital Journalism*, *7*, 230–250. https://doi.org/10.1080/21670811.2018.1519375

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

Fletcher, R., Cornia, A., & Nielsen, R. K. (2020). How polarized are online and offline news audiences? A comparative analysis of twelve countries. *The International Journal of Press/Politics*, *25*, 169–195. https://doi.org/10.1177/1940161219892768

Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*, 2450–2468. https://doi.org/10.1177/1461444817724170

Garson, G. D. (2020). *Multilevel Modeling: Applications in STATA®, IBM® SPSS®, SAS®, R, & HLM™*. SAGE.

Goidel, K., Gaddie, K., & Ehrl, M. (2017). Watching the news and support for democracy: why media systems matter. *Social Science Quarterly*, *98*, 836–855. https://doi.org/10.1111/ssqu.12430

Hallin, D. C., & Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. Cambridge University Press.

Hamada, B. I. (2021). Determinants of journalists' autonomy and safety: Evidence from the Worlds of Journalism Study. *Journalism Practice*, *16*, 1715–1735. https://doi.org/10.1080/17512786.2021.1871861

Hamada, B. I., Abdel-Salam, A.-S. G., & Bebawi, S. (2024). Journalism practices in Western and Muslim majority countries: Culture matters. *The International Journal of Press/Politics*, 1–25. https://doi.org/10.1177/19401612241260730

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, *32*, 385–410. https://doi.org/10.1111/j.1468-2958.2006.00281.x

Hendrickx, J., & Opgenhaffen, M. (2024). Introduction: Understanding social media journalism. *Journalism Studies*, *25*, 919–930. https://doi.org/10.1080/1461670x.2024.2368861

Holbert, R. L., & Park, E. (2020). Conceptualizing, organizing, and positing moderation in communication research. *Communication Theory*, *30*, 227–246. https://doi.org/10.1093/ct/qtz006

Hox, J. J., Moerbeek, M., & ven de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications* (3rd ed.). Routledge.

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The International Journal of Press/Politics*, *25*, 493–516. https://doi.org/10.1177/1940161219900126

Kalogeropoulos, A., Negredo, S., Picone, I., & Nielsen, R. K. (2017). Who shares and comments on news? A cross-national comparative analysis of online and social media participation. *Social Media + Society*, *3*, 1–12. https://doi.org/10.1177/2056305117735754

Knutsen, C. H., Marquardt, Kyle L., Seim, B., Coppedge, M., Edgell, Amanda, Medzihorsky, J., Pemstein, D., Teorell, Jan, Gerring, J., & Lindberg, S. I. (2023). *Conceptual and Measurement Issues in Assessing Democratic Backsliding (Working Paper)*. V–Dem Institute.

Lee, F. L. F., Chan, M., Chen, H.-T., Nielsen, R., & Fletcher, R. (2019). Consumptive news feed curation on social media as proactive personalization: A study of six east Asian markets. *Journalism Studies*, *20*, 2277–2292. https://doi.org/10.1080/1461670x.2019.1586567

Little, A., & Meng, A. (2023). Subjective and objective measurement of democratic backsliding. *SSRN Electronic Journal*, 1–13. https://doi.org/10.2139/ssrn.4327307

Mede, N. G. (2022). Legacy media as inhibitors and drivers of public reservations against science: Global survey evidence on the link between media use and anti-science attitudes. *Humanities and Social Sciences Communications*, *9*(1), 1–11. https://doi.org/10.1057/s41599-022-01058-y

Nisbet, Erik C., Stoycheff, E., & Pearce, K. E. (2012). Internet use and democratic demands: A multinational, multilevel model of internet use and citizen attitudes about democracy. *Journal of Communication*, *62*, 249–265. https://doi.org/10.1111/j.1460-2466.2012.01627.x

Review Copy – Not for Redistribution
File Use Subject to Terms & Conditions of PDF License Agreement (PLA)

Pan, Z., & McLeod, J. M. (1991). Multilevel analysis in mass communication research. *Communication Research*, *18*, 140–173. https://doi.org/10.1177/009365091018002002

Pickering, W. S. F., & Walford, G. (Eds.). (2000). *Durkheim's Suicide: A Century of Research and Debate*. Routledge.

Slater, M. D., Snyder, L., & Hayes, A. F. (2006). Thinking and modeling at multiple levels: the potential contribution of multilevel modeling to communication theory and research. *Human Communication Research*, *32*(4), 375–384. https://doi.org/10.1111/j.1468-2958.2006.00292.x

Steensen, S., & Westlund, O. (2021). *What Is Digital Journalism Studies?* Routledge.

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and bayesian approaches. *American Journal of Political Science*, *57*, 748–761. https://doi.org/10.1111/ajps.12001

Toff, B., & Kalogeropoulos, A. (2020). All the news that's fit to ignore. *Public Opinion Quarterly*, *84*(S1), 366–390. https://doi.org/10.1093/poq/nfaa016

Vacca, R. (2022). *Introduction to multilevel analysis in R with lme4 and tidyverse*. https://raffaelevacca.github.io/Intro-multilevel-with-R/

Waisbord, S. (2019). The 5Ws and 1H of digital journalism. *Digital Journalism*, *7*, 351–358. https://doi.org/10.1080/21670811.2018.1545592