Assessing Local Journalism:

News Deserts, Journalism Divides, and the Determinants of the Robustness of Local News

PHILIP M. NAPOLI, MATTHEW WEBER, KATIE MCCOLLOUGH & QUN WANG

AUGUST, 2018

News Measures Research Project

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The economic challenges confronting local journalism have been well documented. Both of the revenue streams that local news organizations have traditionally relied upon (subscriptions and advertising) have been dramatically undermined as journalism production, distribution, and consumption have migrated online.

And yet, while we know that local journalism is suffering, we know relatively little about whether all communities are being affected in the same way.

- Are some types of communities suffering worse than others?
- Are there particular characteristics of individual communities that are related to the state of their local journalism?

These questions remain largely unanswered because, to date, most research on local journalism has involved detailed case studies examining the state of local journalism in a single community, or in a very limited number of communities. In addressing this gap in our understanding of local journalism, this study has three goals:

- To present a rigorous, replicable methodological approach to assessing the robustness of local journalism across a large number of communities
- To provide descriptive data on the robustness of local journalism that will provide indicators of the extent to which local communities are receiving journalism that is original, local, and that addresses critical information needs.
- To explore whether there are any relationships between the demographic and geographic characteristics of individual communities and the robustness of the local journalism that is available to those communities.

Drawing upon an analysis of over 16,000 news stories, gathered over seven days, across 100 randomly sampled U.S. communities, this study found that:

- Eight communities contained no stories addressing critical information needs.
- Twelve communities contained no original news stories.
- Twenty communities contained no local news stories.

In addition, this study found that:

- Only about 17 percent of the news stories provided to a community are truly *local* – that is actually about or having taken place within – the municipality.
- Less than half (43 percent) of the news stories provided to a community by local media outlets are original (i.e., are produced by the local media outlet).
- Just over half (56 percent) of the news stories provided to a community by local media outlets address a critical information need.

In addition, this study identified community characteristics that appear to be systematically related to the robustness of the local journalism that a community receives. The key characteristics identified in this anlaysis include:

- Distance from a large media market (the smaller the distance, the less robust is the local journalism).
- Number of universities (the more universities in a community, the more robust is the local journalism).
- Hispanic/Latino population (the greater the proportion of a community's population that is Hispanic/Latino, the less robust is the journalism in that community).

This study also found no significant relationship between a community's status as the county seat and the quantity or robustness of local journalism. This finding indicates that the presence of county government activity fails to generate any increases in journalistic production, which suggests that such government activity may no longer lead to a commitment of greater journalistic resources. Such a pattern would seem to reinforce contemporary concerns about the decline in local government reporting.

Overall, these findings provide some of the most comprehensive evidence to date of the magnitude of the *news deserts* problem confronting local communities. These findings, however, offer limited evidence of *journalism divides* – which we define as patterns in the availability of robust journalism that follow the geographic and demographic patterns that have characterized the digital divide.

INTRODUCTION

The economic challenges confronting local journalism have been well documented.[1] As news distribution and consumption have migrated online, it has been difficult for news organizations to attract the same kind of subscription revenue that they were able to attract for their print product. Attracting advertising revenues in the highly competitive online environment has proven equally difficult. Online platforms have siphoned off traditional forms of newspaper advertising such as classifieds. Social media platforms such as Facebook have proven capable of siphoning off not only national advertising dollars, but local advertising dollars as well.[2] These challenges have been particularly pronounced in the United States, where there is a greater reliance on the commercial model of news production than is the case in most other developed nations. *New York Times* Executive Editor Dean Baquet recently described local news as "the biggest crisis in American journalism."[3]

The consequences of these declines in local journalism are wide ranging. Research has shown, for instance, that the demise of local newspapers leads to increases in government costs, as the absence of the governmental watchdog allows local governments to operate under less scrutiny and thus, apparently, less efficiently.[4] Other studies have found that declines in the production and consumption of local journalism are related to declines in citizen knowledge, community enagement, and participation in elections.[5] Purveyors of hyperpartisan news and commentary have begun to treat the decline in local news outlets as an opportunity to create local news sites that have the appearance of traditional, mainstream local news outlets, with the ownership and sponsorship of the sites intentially left opaque.[6] Thus, the gap left by the demise of legitimate local news outlets is being filled by outlets with less journalistic, and more partisan, agendas. Given this range of consequences, it is perhaps not surprising that a recent editorial in the *Washington Post* described the crisis in local journalism as nothing less than "a crisis in democracy."[7]

And yet, while we know that local journalism is suffering, we know relatively little about whether all communities are being affected in the same way.

- ■ *Are some types of communities suffering worse than others?*

- ■ *Are there particular characteristics of individual communities that are related to the state of their local journalism?*

These questions remain largely unanswered because, to date, most research on local journalism has involved detailed case studies examining the state of local journalism in a single community, or in a very limited number of communities.[8] This tendency is most likely a reflection of the challenges associated with developing a methodology that can be applied to a larger sample of communities at a manageable cost. In research in which larger samples of communities are analyzed, the analytical framework is typically limited to a single technology (e.g., television, newspapers, hyperlocals, or citizen journalism aggregation sites), or to a particular issue or type of news (e.g., local government, state house, or election reporting).[9] This narrower technological or subject matter focus reflects the fact that the primary unit of analysis in such research is typically the individual media outlet or technology type, or the individual news story, rather than the community as a whole, and the totality of the news and information being produced for the community. As a result, we lack a robust understanding of the full magnitude of the problem affecting local communities, and if/how the problem varies across different types of communities.[10]

When we consider the individual community as the unit of analysis, concerns about "news deserts" inevitably arise.[11] Here, the concern is that communities may find themselves completely lacking in news serving their specific information needs. Some types of communities may be

affected more profoundly than others, with the declines affecting local journalism potentially varying according to particular demographic or geographic characteristics.[12] For instance, one report noted that large US cities such as New York, Washington, DC, and Los Angeles are employing an increasing proportion of the country's professional journalists, with smaller cities experiencing dramatic declines.[13] Other research has suggested that communities with large numbers of minorities, immigrants, or lower-income residents may be suffering disproportionately from the declines affecting local journalism.[14] Along related lines, another recent study raised the possibility of *journalism divides* – patterns in the relative availability in local journalism that reflect the patterns and characteristics associated with the well-known digital divide, in which lower-income, minority, and rural communities tended to suffer from lower levels of Internet access than the rest of the population.[15] Specifically, it may be the case that lower income and more ethnically diverse communities generally have access to less robust local news media; just as, historically, lower income and more ethnically diverse communities have had lower levels of access to the Internet.

With this state of affairs in mind, the goals of this paper are threefold:

1. To present a rigorous, replicable methodological approach to assessing the robustness of local journalism in a way that scales to the analysis of a large number of communities, and that can be used to assess differences in the state of local journalism either across communities or over time.

2. To provide descriptive data on the robustness of local journalism that will provide indicators of the extent to which local communities are receiving journalism that is original, local, and that addresses critical information needs.

3. To explore whether there are any relationships between the demographic and geographic characteristics of individual communities and the robustness of the local journalism that is available to those communities. This final goal would support efforts to determine whether certain types of communities are particularly vulnerable to the declines affecting local journalism.

METHOD

Research on local journalism has tended to focus on large metropolitan areas.[16] Our goal was to look outside of this context, to the smaller, more vulnerable communities across the U.S. Thus, our focus was on communities with populations ranging from 20,000 to 300,000 residents. Using U.S. Census data, these parameters resulted in a list of 493 communities. From this list, we selected a random sample of 100 communities. Then, we verified that sample and approximated the population based on population, income, demographic composition.

Identifying Local Media Outlets

Our next step was to generate a complete inventory of all of the local media outlets in each of these communities. Thus, our search criteria included local newspapers, as well as local radio stations, local television stations and local online-only news sources. Our inventory of local news sources was limited to those sources geographically located within each sampled community. Therefore, we excluded news sources that might produce news of relevance to the community but that were geographically located outside of the community. Clearly then, we employed a very strict geographic notion of local in our process for identifying local media outlets.

This inventory was created through a systematic process of consulting multiple media databases and directories – 11 in total (see Appendix A) and supplementing these database and directory scans with a multi-stage online keyword search protocol (see Appendix B). This multi-pronged approach reflects the fact that a comprehensive portrait of the media outlets serving local communities today can only be achieved via combining information from a broad array of sources. Even large-scale commercial media directories (e.g., Cision) were discovered to have coverage gaps when compared to the manual multi-database/directory search process that we employed.

Although our search criteria included print, radio, television, and online media, the content gathering and archiving was conducted exclusively online. Thus, the corresponding URL for each media outlet's home page was located and recorded for use in the archiving process described below. In this way, the journalistic outputs of daily and weekly newspapers, magazines, radio stations, television stations, and local cable channels all were assessed via their online content offerings, in the same way that the outputs of online news sources such as community journalism sites were assessed.

This approach (which is a reflection of the effort to create a realistically scalable methodology) runs counter to the common assertion that certain types of legacy media (e.g., local weekly print publications, ethnic media outlets) remain slow to utilize the internet as a means of disseminating their content. Our searches indicated that we are now at a point in the evolution of legacy media at which this generalization no longer holds true. The economic and strategic pressures and incentives to have an online presence, combined with the inherent economic imperative to distribute content production costs across as broad an audience base as possible,[17] suggest that the content available online can serve as a reliable indicator of the relative journalistic output across individual outlets, regardless of their "native" platform.

The key term here is *indicator,* as we are not seeking to produce a comprehensive inventory of journalistic output, only a set of indicators that are conceptually and methodologically robust and that can be employed in comparative analyses across communities or over time. In support of this position, the data gathered on the 100 selected communities identified only 37 out of 791 (5%) total media outlets that did not have a corresponding online presence, with the bulk of these being low-power and translator radio stations.[18]

Content Archiving

Our content data gathering relied on a partnership with the Internet Archive.[19] We used the Archive-IT community archiving platform to create our own web archive, as the locally-oriented web sites that are central to this study generally are not part of the Internet Archive's regular archiving work. The Archive-IT platform allows users to specify the websites for collection and to set parameters including the frequency and depth of crawling. In this case, we had 733 distinct URLs that we identified for content archiving.[20] In cases in which a media outlet's content was behind a paywall, subscriptions to those sites were obtained.

In terms of the depth of the content crawl, we focused on the home-page coverage of each media outlet and crawled to a depth of one. This means that in addition to archiving the home page, we archived each web page that was one hyperlink (or one click) from the home page. This approach draws from the premise in the journalism studies literature that news source front pages/home pages represent a meaningful indicator of the most important news events and issues affecting a community and thus represent a useful means of assessing media outlet performance.[21]

The last methodological question to address was the sampling used to create the web archive. For this project, we created a constructed week sample, selecting a Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday at random over a two-month period during the summer of 2016. Websites were crawled on July 27, August 2, August 6, August 28, September 9, September 21 and September 26 of 2016.

Using this methodology, we created a robust record of the home-page presentation of local news in 100 communities. The web crawl across the sample week generated a collection containing 1.6 million documents (html files, pdfs, images, audio files, etc.), 2.2 terabytes of total data, and, an archive of over 16,000 news stories.[22]

Content Analysis

The primary goal of our content analysis was to determine the extent to which journalistic output was serving the information needs of local communities. To do this in a way that could be reliably and efficiently accomplished at the scale at which we were operating, we focused on three criteria:

1. Whether the story was *original;*

2. Whether the story was *about the local community;*

3. Whether the story addressed a *critical information need.*

These criteria can be seen as fairly superficial – but fundamental –indicators of the complex notion of the "quality" of the journalism being produced within communities.[23] These criteria provide a relatively simple, economical, straightforward, and replicable set of indicators of journalistic performance that address the fundamental concern about whether journalistic sources are addressing communities' information needs.

The notion of critical information needs has been central to the ongoing discourse about the performance of local journalism.[24] This approach draws upon this discourse, and the research it has inspired.[25] Specifically, each story was content analyzed to determine whether it fit into one or more of the critical information needs categories identified in the literature review by Friedland et al. that was prepared for the Federal Communications Commission.[26] Friedland et al. provide eight categories of critical information needs. These categories are as follows:[27]

1. Emergencies and risks
2. Health
3. Education
4. Transportation systems
5. Environment and planning
6. Economic development
7. Civic information
8. Political life.

Descriptions for each of these categories can be found in Appendix C. Stories that did not address one of these categories were coded as a 9. This category is intended to capture less substantive story types (celebrity news, sports reports, etc.). For the purpose of the analyses presented here, categories 1-8 are collapsed in order to create a dichotomous variable (i.e., story does/does not address a critical information need) that reflects the frequently-employed distinction between "hard" and "soft" news.

Each story also was coded for whether it was *original* (i.e., produced by the media outlet rather than reprinted, linked, retweeted, or shared from elsewhere) and whether it was *about the local community.*

The emphasis here on original stories is intended to separate aggregation, linking, sharing, retweeting, and re-publication activities from original reporting.[28] This originality measure seeks to discern the robustness of local journalism sources by determining how active they are in *producing* news stories. A story was considered "original" if it had a byline by an outlet's reporter, or if it had no indicators that it was a reposting of, or hyperlink to, content produced elsewhere.

The emphasis on locality is employed to capure the extent to which the output of local journalism sources focuses on the local community. This measure is intended to address the extent to which local journalism is truly local, providing community members with news and information about, and directly relevant to, their communities. This measure reflects the long-standing localism principle, which has featured prominently in democratic theory perspectives on media, and in media policymaking. From this perspective, the extent to which citizens are able to engage in informed democratic participation in their communities is a function of the availability of local news and information about their communities.[29] For this variable, we opted for a strict geographic definition of community, where we identified an item as about the community only if the subject was an issue/event oriented around the specific municipality.

Together, these three variables reflect some of the primary concerns in ongoing critiques and analyses of the state of local journalism today:

1. That the economic pressures on local journalism create overwhelming incentives to aggregate and repurpose existing content rather than engage in original reporting[30]

2. That the changing technological and economic dynamics for news distribution and consumption are exacerbating the extent to which large-market or out-of-market news can infiltrate local communities, thereby undermining local journalism[31]

3. That the increasing challenges associated with attracting and retaining an audience for news are compelling local news outlets to neglect substantive topics in favor of an emphasis on "soft" news, celebrity, and sensationalism.[32]

For these reasons, we think these particular variables of focus represent a useful set of top-level indicators of how well local journalism is fulfilling its central purpose of facilitating informed participation and engagement in local community affairs.

Content analysis of the online news stories was conducted by a team of trained coders during the spring and summer of 2017.[33]

From these data, four dependent variables were constructed:

1. *Stories:* (the total number of news stories produced for the community)

2. *Original:* (the total number of original news stories produced for the community)

3. *Local:* (the total number of local news stories produced for the community)

4. ***CIN:*** (the total number of news stories addressing a critical information need produced for the community)

In addition, variables were created for each combination of variables 2-4 (i.e., counts of stories meeting two or more of the robustness criteria), leading to:

5. ***Original + Local***

6. ***Original + CIN***

7. ***Local + CIN***

8. ***Original + Local + CIN***

These combination measures were utilized only in the descriptive analyses and not in the multivariate analyses.

Community Characteristic Variables

For the multivariate analysis, we gathered data on a wide range of community characteristics that previous research has suggested may be related to the robustness of local journalism. A number of municipality characteristics were gathered from 2010 census data. From the census reports, we gathered data on the size of the population (POPULATION) in each community to control for the effects that community size is likely to have on journalistic output.[34] We also gathered data on the population density (DENSITY) of each community, as this also may be related to the robustness of journalistic output, with denser communities perhaps facilitating greater efficiencies in journalistic production.

Muncipality demographic data gathered for this analysis from 2010 census data included median household income (INCOME), the percentage of the population that is African-American (AA%), and the percentage of the population that is Hispanic/Latino (HL%). These data were gathered to explore whether the robustness of local journalism follows patterns in line with the digitial divide, which was often found to be a function of income and ethnic diversity.

Data were gathered on the number of four-year colleges/universities in a community (#UNIVERSITIES), since universities tend to be sources of activity and local news production. Data also were gathered on whether the municipality was a county seat (COUNTYSEAT), on the assumption that the greater governmental activity within such a community may generate higher levels of journalistic output. Finally, data were gathered on each community's distance from a large (Nielsen top 50) media market (DISTANCELMM), to explore the possibility that proximity to large media markets (and the one-way flows from large to small media markets that are characteristic of news flows) undermines the viability (and thus output) of local news outlets.[35]

Descriptive statistics for these community characteristics are presented in Appendix D.

RESULTS

As was noted above, one of the goals of this project is to provide a descriptive overview of the state of local journalism that reflected the analysis of a large number of communities. Toward this end,

Figure 1 focuses on the number of communities that generated no stories that met the various coding criteria (and combinations of coding criteria). This focus gives us a sense of the scope of the *news deserts* problem.

Of the 100 communities sampled for this analysis, *eight generated no news stories* for analysis. These communities represent news deserts in the most extreme sense. Residents in these communities received no news of any kind from their local media outlets. It should be noted that four of these eight communities had no media outlets, according to our search protocol.[36]


FIGURE 1:
ZERO-STORY COMMUNITIES BY STORY TYPE (N=100)
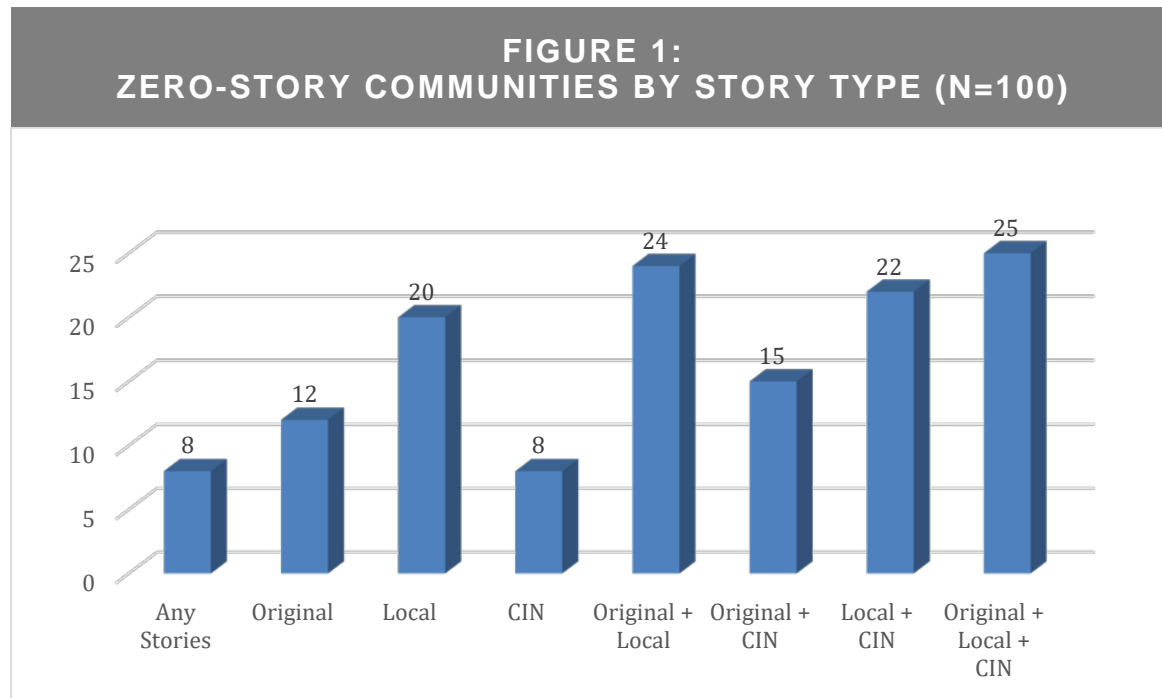
Similarly, *eight communities contained no stories addressing critical information needs. Twelve communities contained no original news stories. Twenty communities contained no local news stories.* Thus, a full 20 percent of the sample of communities was completely lacking in journalism about their communities. These communities represent an important dimension of the local news deserts that have become a growing concern in recent years. While many of these communities may be receiving news from their local media outlets, none of this news focuses on issues and events within the community.

Looking next at the extent to which communities receive stories that address more than one of our robustness criteria, the results continue to be somewhat troubling. Twenty-four communities in the sample had no stories that were both original and local; 15 had no stories that were original and that addressed a critical information need; 22 had no stories that were local and addressed a critical information need; and 25 had no stories that met all three criteria (original, local, and addressed a critical information need).

Figure 2 presents the aggregated results of the content analysis, showing the mean percentage across all communities in terms of each story robustness category and category combination.[37] This figure tells us what average proportion of the news stories in each community met the various robustness criteria. Perhaps the most striking finding in this table is that, on average, only about

17 percent of the news stories provided to a community are truly *local* – that is actually about or having taken place within – the municipality. Clearly, very little of what we might consider local news is (at least in the strict geographic sense) actually local.

**FIGURE 2:
MEAN PROPORTION OF TOTAL STORIES BY STORY TYPE (N=100)**
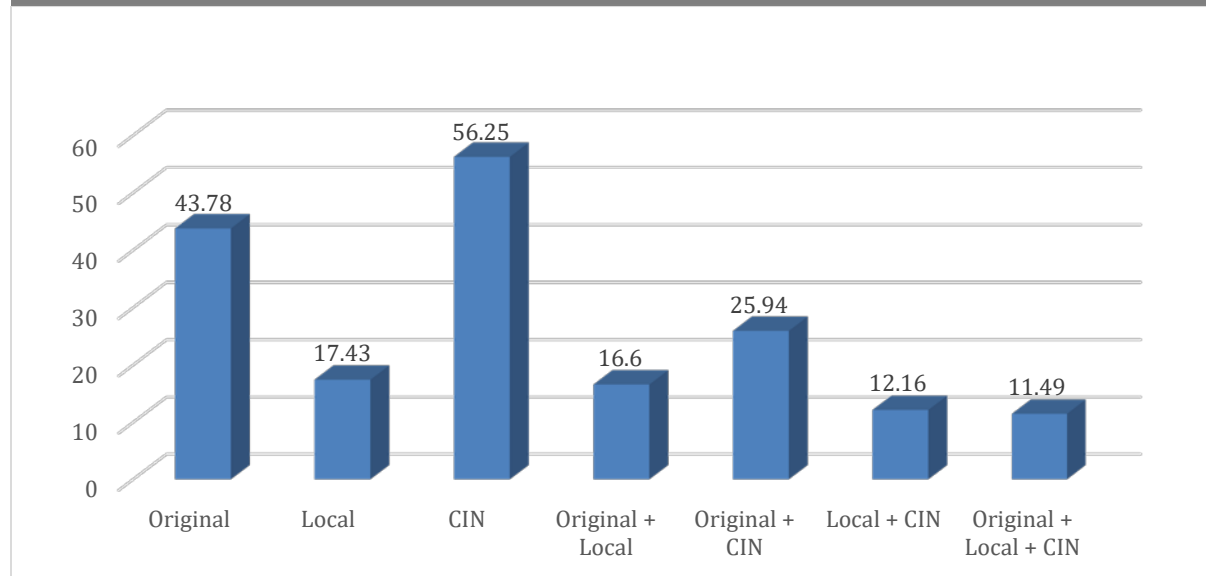
In terms of originality, the results are substantially better. Nonetheless, less than half (43 percent) of the news stories provided by these local media outlets are original. Thus, the majority of the news stories provided by local media outlets originate elsewhere. Local media outlets exhibited a strong tendency to hyperlink to stories produced by other outlets, or (as we often saw in our data) outlets post stories produced by other outlets directly to their site, but with attribution to that other outlet either very clear or discernible through our content analysis protocol.

Looking next at the extent to which news stories addressed critical information needs, we see that, on average, just over half (56 percent) of the news stories provided to the communities met this criterion. It is worth re-emphasizing that our critical information needs category was not locally oriented; thus, a nationally-focused political story was categorized as addressing a critical information needs just as a locally-focused story would be categorized as addressing a critical information need.

When we apply more stringent criteria to these news stories – isolating only those stories that meet two or all three of the robustness criteria – the proportions drop further. For instance, the average proportion of stories within a community that are original, local, *and* address a critical information need (the trifecta of robust local journalism according to our criteria) is less than 12 percent.

Multivariate Analyses

Finally, we sought to determine whether the robustness of local journalism in individual communities was a function of the geographic and demographic characteristics of those

communities. Such an analysis can help us to determine the types of communities that are suffering the most, in terms of the robustness of their local journalism.

For this analysis, we employed a type of regression called negative binomial regression.[38] Negative binomial regression is designed to accommodate "count" dependent variables such as those employed here (counts of the number of news stories that meet various criteria). For this analysis, we focused on four dependent variables:

1. The total number of stories in each community (STORIES)
2. The total number of original stories in each community (ORIGINAL)
3. The total number of local stories in each community (LOCAL)
4. The total number of stories addressing a critical information need (CIN)

As was described above, we gathered a range of community characteristic data to employ as independent variables. These included:

1. Population size (POPULATION)
2. Population density (DENSITY)
3. Median household income (INCOME)
4. Distance from a large (top 50) media market (DISTANCELMM)
5. The percentage of the community's population that is African American (%AA)
6. The percentage that is Hispanic/Latino (%H/L)
7. The number of universities in the community (#UNIVERSITIES)
8. Whether the community serves as the county seat (COUNTYSEAT).[39]

We ran a separate regression analysis for each of the dependent variables. To facilitate comparisons across each of the four regressions, they are presented in a single table (Table 1), with the coefficients of the statistically significant independent variables highlighted in boldface. Each of the four regression models was statistically significant.[40]

The first regression (STORIES) focuses on the overall story output (i.e., total number of stories) produced in each community. This variable provides a sense of the overall level of journalistic output, absent any assessment of each story's underlying characteristics. As this regression indicates – and as is the case across all four regressions – the POPULATION control variable is statistically significant. The significance of this variable indicates that overall journalistic output is, as we would expect, a function of the size of the population within a community.[41] Larger communities generate/support more reporting.

Continuing to look at regression #1, we see that DISTANCELMM and #UNIVERSITIES are both statistically significant in the expected direction. Thus, the further a community is from a top 50 media market, the greater the overall journalism output in that community. Similarly, the greater the number of four-year colleges/universities in a community, the greater the overall journalistic output in that community.

Looking next at the regression for ORIGINAL story output, POPULATION and #UNIVERSITIES again emerge as statistically significant in the expected directions. In addition, the percentage of the community population that is Hispanic/Latino (%HL) also is statistically significant in the negative direction. This indicates that the greater the proportion of a community's population that is Hispanic/Latino, the less original reporting that community receives.

TABLE 1:
COMMUNITY CHARACTERISTICS RELATED TO
JOURNALISM ROBUSTNESS (N=100)

	STORIES	ORIGINAL	LOCAL	CIN
Parameter	B	B	B	B
(Intercept)	3.99	3.6	2.9	3.06
[COUNTYSEAT=0]	-.30	-.38	-.65	-.22
[COUNTYSEAT=1]	0[a]	0[a]	0[a]	0[a]
POPULATION	**9.74E-6****	**1.06E-5****	**1.20E-5****	**9.31E-6****
DENSITY	-9.02E-5	-2.07E-5	-2.66E-5	-7.82E-5
INCOME	6.55E-6	1.2E-6	-2.66E-7	9.10E-6
DISTANCELMM	**.006***	.004	.005	**.007***
%AA	.003	.004	.002	.008
%H/L	-.006	**-.02***	**-.03****	-.003
#UNIVERSITIES	**.59***	**.62***	.58	**.63***
(Negative binomial)	1.30	1.60	1.73	1.38
Likelihood X^2	33.22**	27.61**	28.45**	33.88**

The third regression in Table 1 focuses on the level of LOCAL story production in a community. As is indicated in Table 1, POPULATION is once again statistically significant in the expected direction, and %H/L is once again negatively statistically significant. It appears that declines in locally oriented reporting also are a function of the extent to which a community's population is Hispanic/Latino.

Turning finally to regression #4, which focuses on the production of stories that address critical information needs (CIN), we see POPULATION and #UNIVERSITIES once again emerge as statistically significant in the expected directions. DISTANCELMM once again emerges as positively statistically significant, indicating that as distance from a large media market increases, the number of stories produced by local media outlets that address critical information needs increases as well.

It is worth noting that the COUNTYSEAT variable was not significant in any of the four regressions. This result is striking in that it suggests that the presence of county-level governmental activity within a municipality fails to generate any overall increases in journalistic output. This could be interpreted as an indicator of the extent to which this kind of local government reporting has fallen by the wayside so dramatically in recent years.

It is also important to note that potential journalism divide-related factors such as income, population density, and African-American population composition were not significantly related to any of the dependent variables. Thus, the robustness of journalistic output does not appear to be a reflection of the wealth of a community, the extent to which a community is rural vs. urban, or the size of the African-American population. However, the significance of the Hispanic/Latino variable (%H/L) across two of the four regressions suggests that the Hispanic/Latino population may be less well-served, particularly in terms of stories that are original or local.

The significance of the distance from a large media market (DISTANCELMM) across two of the four regressions suggests that the proximity of a community to such a market is a relevant factor that can negatively affect local journalistic output, particularly in terms of overall story output and in terms of story output addressing critical information needs.

CONCLUSION

This analysis has provided a sense of the extent to which the news provided to communities by their local media outlets is original, local, and addresses critical information needs – utilizing a community sample size and news story corpus that far exceed previous research on local journalism. As the descriptive data indicated, "news deserts" are fairly prominent across the sample, with eight of the communities containing no news stories, eight containing no stories addressing critical information needs, twelve containing no original stories, and 20 containing no local stories. One could argue that these communities that are completely devoid of news stories, and/or completely devoid of stories that are local, or original, or that address critical information needs, are emblematic of the new deserts that have become an increased concern as the economics of local journalism have become more challenging. Looking at journalistic output in the aggregate, less than half of the news provided by local media outlets is original; just over half addresses a critical information need; and only 17 percent is truly local. All told, these results would seem to exacerbate concerns about the robustness of local journalism in the U.S.

Turning next to the results of the multivariate analyses, the extent to which the robustness of local journalism is a function of the demographic and geographic characteristics of a community appears to be limited to the proportion of the community that is Hispanic/Latino and the community's distance from a top-50 media market. The negative significance of %H/L across two of the four regressions (LOCAL and ORIGINAL) may be indicative of greater economic challenges affecting foreign-language journalism in the U.S. (particularly alongside the non-significance of AA%). However, it is likely also the case that the interest level of the Hispanic/Latino population in non-local news (i.e., international news) is greater than the general population, which may also increase the extent to which local outlets serving these communities rely on content produced elsewhere.

The findings presented here provide some support for previous assertions that large media markets, and their associated substantial journalistic output, tend to have a negative effect on the journalistic output in nearby local communities. This is one of the key ramifications of the one-way flow phenomenon that has been shown to characterize journalism; furthermore, these negative effects are likely being exacerbated as journalism is increasingly disseminated and consumed via digital platforms.[42]   The process at work, then, is one in which the flow of large-market journalism into neighboring communities undermines the economics of local journalism production. In the case of this analysis, this phenomenon appears to have undermined overall story output as well as the production of stories addressing critical information needs. This latter pattern suggests that local media outlets may, to some extent, be ceding the more substantive elements of journalism to their large-market neighbors.

The fact that a community's status as a county seat failed to exhibit any significant relationship with journalistic output serves to highlight growing concerns that local governmental reporting may be particularly vulnerable in today's increasingly challenging journalistic environment. The fact that the presence of county government activity fails to generate any increases in journalistic production suggests that such activity may no longer lead to commitment of greater journalistic resources.

Future Research

There are a number of additional analytical directions that can be taken with these data. For instance, if we drill down into the distribution of news production across outlets, it would be possible

to explore the question of which types of outlets contribute most to a local news ecosystem. Previous research has indicated that local newspapers remain the most significant source of original local reporting.[43] Future research utilizing these data could explore if, or to what extent, this remains the case; along with the more general question, what is the extent to which journalistic output is concentrated within, or dispersed across, the range of outlets serving a community? This is a question that we will address in a subsequent report.

Future research may also consider if, or to what extent, the nature of the critical information needs addressed by local media outlets is a reflection of community characteristics, which may offer real-world indicators of the local importance of each critical information need. In this way, it might be possible to explore the extent to which local journalism is responsive to the particular information needs of each community.

In addition, network analysis can be employed to explore linking patterns across media outlets, and thus map the sourcing dynamics within local news ecosystems. By integrating data from named entity recognition software packages, it would also be possible to dig more deeply into geographic patterns in the distribution of local news coverage in order to identify the types of areas within communities that receive substantial news coverage and the types that do not.

This latter example also highlights an important next step if this type of analyses were to become part of a more regular, systematic assessment of the state of local journalism – exploring possible automated approaches to the content analysis conducted here. As automated content analysis continues to make strides in replacing the very time- and labor-intensive approach to content analysis utilized here, exploring the possibility of automated approaches to identifying whether a story is original, local, and addresses a critical information need seems worthwhile.

It is also important to note that the archive created for this analysis has been made publicly available,[44] and that the Internet Archive has continued to archive one day per month for all of the media outlets studied here, as the basis for an ongoing national sample of local media output. This archive could be utilized to perform a wide range of analyses of local journalism that analyze the content we have gathered in much different, or much more granular, ways. We hope that researchers find novel and valuable ways to make use of this substantial and current local journalism archive.

APPENDIX A:
MEDIA DATABASES/DIRECTORIES CONSULTED

The following were the key sources consulted in searching for news outlets within a given community.

Television

1. Association of Public Television Stations' Station Directory
2. FCC Broadcast Television License Database
3. NPR Labs Mapping and Population System

Radio

4. FCC AM and FM Broadcast License Database
5. Radio Locator

Newspapers

6. Library of Congress Directory of Newspapers

Online News Sites

7. Knight Foundation's Directory of Community News Sites
8. Columbia Journalism Review's Guide to Online News Startups
9. Online Newspaper Directory for the World
10. Michelle's List

Alternative Sources

11. Mondotimes

APPENDIX B:
MANUAL SEARCH PROTOCOL

In order to identify additional sources missed in the online search, a manual search protocol was developed in order to search for additional sources that were not readily accessible via traditional search.

1. Start with Wikipedia and search for each community
   a. May list community's media
   b. May provide community nickname(s) to utilize in search engine queries
   c. May describe large minority populations that could be useful in search for minority/foreign language media outlets

2. Go to Patch.com
   d. Enter community name into Find Your Patch pulldown menu

3. Google Search
   a. Key terms:
      i. "[Community Name][Community Nickname] News"
      ii. "[Community Name][Community Nickname] Journalism"
      iii. "[Community Name][Community Nickname] Hyperlocal"
      iv. "[Community Name][Community Nickname] Blog"
   b. Repeat in Spanish
      i. News = Noticias
      ii. Journalism = Periodismo

APPENDIX C:
CRITICAL INFORMATION NEEDS CATEGORIES

1. Emergencies and risks

Individuals, neighborhoods, and communities need access to emergency information on platforms that are universally accessible and in languages understood by the large majority of the local population, including information on dangerous weather; environmental and other biohazardous outbreaks; and public safety threats, including terrorism, amber alerts, and other threats to public order and safety. Further, all citizens need access to local (including neighborhood) information on policing and public safety.

2. *Health*

All members of local communities need access to information on local health and healthcare, including information on family and public health in accessible languages and platforms; information on the availability, quality, and cost of local health care for accessibility, lowering costs, and ensuring that markets function properly, including variations by neighborhood and city region; the availability of local public health information, programs, and services, including wellness care and local clinics and hospitals; timely information in accessible language on the spread of disease and vaccination; timely access to information about local health campaigns and interventions.

3. *Education*

Local communities need access to information on all aspects of the local educational system, particularly during a period when local education is a central matter for public debate, decision-making, and resource allocation, including: the quality and administration of local school systems at a community-wide level; the quality of schools within specific neighborhoods and geographic regions; information about educational opportunities, including school performance assessments, enrichment, tutoring, afterschool care and programs; information about school alternatives, including charters; information about adult education, including language courses, job training, and GED programs, as well as local opportunities for higher education.

4. *Transportation Systems*

All members need timely information about local transportation across multiple accessible platforms, including: information about essential transportation services including mass transit at the neighborhood, city, and regional levels; traffic and road conditions, including those related to weather and closings; timely access to public debate on transportation at all layers of the local community, including roads and mass transit.

5. *Environment and Planning***:**

Local communities need access to both short and long-term information on the local environment, as well as planning issues that may affect the quality of lives in neighborhoods, cities, and metropolitan regions, including; the quality of local and regional water and air, timely alerts of hazards, and longer term issues of sustainability; the distribution of actual and potential environmental hazards by neighborhood, city region, and metropolitan area, including toxic hazards and brownfields; natural resource development issues that affect the health and quality of life and

economic development of local communities; information on access to environmental regions, including activity for restoration of watersheds and habitat, and opportunities for recreation.

6. *Economic Development*

Individuals, neighborhoods, and communities need access to a broad range of economic information, including: employment information and opportunities within the local region; job training and retraining, apprenticeship, and other sources of reskilling and advancement; information on small business opportunities, including startup assistance and capital resources; information on major economic development initiatives affecting all local levels.

7. *Civic Information*

Communities need information about major civic institutions, nonprofit organizations, and associations, including their services, accessibility, and opportunities for participation in: libraries and community-based information services; cultural and arts information; recreational opportunities; nonprofit groups and associations; community-based social services and programs; and religious institutions and programs.

8. *Political Life*

In a federal democracy, citizens need information on local, regional, and county candidates at all units of governance, including: information on elected and voluntary neighborhood councils; school boards; city council and alder elections; city regions; and county elections; timely information on public meetings and issues, including outcomes; information on where and how to register to vote, including requirements for identification and absentee ballots; information on state-level issues where they impact local policy formation and decisions.

APPENDIX D:
DESCRIPTIVE STATISTICS

DESCRIPTIVE STATISTICS FOR COMMUNITY VARIABLES (N=100)				
Variable	**Minimum**	**Maximum**	**Mean**	**SD**
POPULATION	20256	291707	59408.05	50797.90
DENSITY	628.5	17346.30	3258.92	2830.43
INCOME	21301	156439	68775.30	25586.94
DISTANCELMM	4	342	57.78	61.86
AA%	0	86	10.78	15.28
HL%	1.0	96.0	15.73	19.05
#UNIVERSITIES	0	2	.21	.48
Variable	**Frequency (Y)**			
COUNTYSEAT (Y/N)	19			

ABOUT THE AUTHORS

Philip M. Napoli is the James R. Shepley Professor of Public Policy in the Sanford School of Public Policy, and a Faculty Affiliate in the DeWitt Wallace Center for Media and Democracy, at Duke University. He is the Co-Principal Investigator of the News Measures Researh Project.

Matthew Weber is an Associate Professor in the Hubbard School of Journalism and Mass Communication at the University of Minnesota, where he is a Cowles Endowed Fellow of Media Management. He is the Co-Principal Investigator of the News Measures Research Project.

Katie McCollough is an Assistant Professor of Media Studies at Augustana University. She has served as the lead Research Assistant for the News Measures Research Project.

Qun Wang is a Doctoral Candidate in the School of Communication and Information at Rutgers University, and has served as a Research Assistant for the News Measures Research Project.

ENDNOTES

[1] See, e.g., Abernathy, P. (2016). *The rise of a new media baron and the emerging threat of news deserts.* Chapel Hill, NC: University of North Carolina Press; Chen, D.W. (2017, Jan. 3). "In New Jersey, Only a Few Media Watchdogs are Left," *New York Times,* https://www.nytimes.com/2017/01/03/nyregion/in-new-jersey-only-a-few-media-watchdogs-are-left.html; Nielsen, R.K. (2015). *Local journalism: The decline of newspapers and the rise of digital media.* London: I.B. Tauris & Co.

[2] Tom Grubisich (2018, Jan. 4). "Why even top local news sites can't compete with Facebook on Ads," *Street Fight,* http://streetfightmag.com/2018/01/04/why-even-top-local-news-sites-cant-compete-with-facebook-on-ads/.

[3] Warren, J. (2017, May 31). "*New York Times* editor: Local news is the biggest 'crisis' in journalism," *Poynter,* https://www.poynter.org/news/new-york-times-editor-local-news-biggest-crisis-journalism.

[4] Kriston Capps (2018, May 30). "The hidden costs of losing your city's newspaper," *CityLab,* https://www.citylab.com/equity/2018/05/study-when-local-newspaper-close-city-bond-finances-suffer/561422/.

[5] George, L., & Waldfogel, J. (2006). Does the *New York Times* spread ignorance and apathy? Unpublished working paper, http://www.anderson.ucla.edu/faculty_pages/romain.wacziarg/mediapapers/GeorgeWaldfogel.pdf.; Hayes, D., & Lawless, J.L. (2014). As local news goes, so goes citizen knowledge, and participation in U.S. House elections. American University School of Public Affairs Research Paper No 2014-0004, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2452076; Hayes, D., & Lawless, J.L. (2018). The decline of local news and its effects: New evidence from longitudinal data, *Journal of Politics 80*(1), 332-336.

[6] Schwartz, J. (2018, May 1). "Baby Breitbarts to pop up across the country?" *Politico,* https://www.politico.com/story/2018/04/30/breitbart-tennessee-fake-news-560670.

[7] Waldman, S., & Sennott, C. (2018, April 11). "The crisis in local journalism has become a crisis of democracy," *Washington Post,* https://www.washingtonpost.com/opinions/the-crisis-in-journalism-has-become-a-crisis-of-democracy/2018/04/11/a908d5fc-2d64-11e8-8688-e053ba58f1e4_story.html?utm_term=.9867a3ac6bfc. See also Gondreau, H. (2017, Nov. 30). "Don't stop the presses! When local news struggles, democracy whithers," *Wired,* https://www.wired.com/story/dont-stop-the-presses-why-big-tech-should-subsidize-real-journalism/.

[8] Durkin, J., & Glaisyer, T. (2011). An information case study: Scranton. New America Foundation, http://oti.newamerica.net/publications/policy/Scranton; Durkin, J., Glaisyer, T., & Hadge, K. (2010). An information case study: Seattle. New America Foundation, http://oti.newamerica.net/publications/policy/seattle; Gloria, K., & Hadge, K. (2010). An information community case study: Washington, DC. New America Foundation, http://oti.newamerica.net/publications/policy/an_information_community_case_study_washington_dc; Morgan, F. (2011). The Research Triangle, North Carolina: A region of locally owned media outlets and entrepreneurs on the verge. New America Foundation, http://oti.newamerica.net/publications/policy/the_research_triangle_north_carolina; Pew Research Center (2010). How news happens: A study of the news ecosystem of one American city, http://www.journalism.org/2010/01/11/how-news-happens/; Ramos, D., Gunes, M.H., Mensing, D., & Ryfe, D. (2013). Mapping emerging news networks: A case study of the San Francisco Bay Area. In R. Menezes, A. Evsukoff, & M.C. Gonzalez, (Eds.), *Complex networks* (pp. 37–244). Berlin Heidelberg: Springer-Verlag.

[9] Becker, A., & Yanich, D. (2015). Citizens' critical information needs and local TV news. Report from the Local TV News Media Project, http://udspace.udel.edu/bitstream/handle/19716/21398/BeckerYanich%20Broadcast%20News%20CIN%20Model%2030%20July%202015_0.pdf?sequence=1; Fico, F., Lacy, S., Baldwin, T., Bergan, D., Wildman, S.S., & Zube, P. (2013). Newspapers devote far less coverage to county government than to city governance. *Newspaper Research Journal 34*(1), 104–111; Holt, K., & Karlsson, M. (2015). Random acts of journalism?: How citizen journalists tell the news in Sweden. *New Media & Society 17*(11), 1795–1810; Karlsson, M., & Holt, K. (2014). Is anyone out there? Assessing Swedish citizen-generated community journalism. *Journalism Practice 8*(2), 164–180; St. John, B., Johnson, K., & Nah, S. (2014). Patch.com: The challenge of connective community journalism in the digital sphere. *Journalism Practice, 8*(2), 197-212.

[10] For additional research addressing this gap in the literature, see the ongoing work of the Center for Innovation and Sustainability in Local Media (e.g., Stites, T. [2018, June 1], "New data tracks how fast news deserts are spreading," *Poynter,* https://www.poynter.org/news/new-data-tracks-how-fast-news-deserts-are-spreading.

[11] Abernathy (2016); Ferrier, M., Sinha, G., & Outrich, M. (2016). Media deserts: Monitoring the changing media ecosystem. In M.I. Lloyd & L.A. Friedland (Eds.), *The communication crisis in America, and how to fix it* (pp 215-232). New York: Palgrave Macmillan; Stites, T. (2011, December 8). Layoffs and cutbacks lead to a new world of news deserts. *Nieman Lab*, http://www.niemanlab.org/2011/12/tom-stites-layoffs-and-cutbacks-leadto-a-new-world-of-news-deserts/.

[12] See, e.g., Ferrier, et al. (2016).

[13] Tankersley, J. (2015, April 23). Why the PR industry is sucking up Pulitzer winners. *The Washington Post.* Retrieved from http://www.washingtonpost.com/blogs/wonkblog/wp/2015/04/23/why-the-prindustry-is-sucking-up-pulitzer-winners/.

[14] Bair, M. (2017, June 19), "News deserts in immigrant communities," *MediaShift,* http://mediashift.org/2017/06/news-deserts-immigrant-communities/; Stites, 2018.

[15] Napoli, P.M., Stonbely, S., McColloughet, K., & Renninger, B. (2017). Local journalism and the information needs of local communities, *Journalism Practice 11*(4), 373-395. For a discussion of the digital divide, see Pick, J.B., Sarkar, A., & Johnson, A. (2015). United States digital divide: State level analysis of spatial clustering and multivariate determinants of ICT utilization. *Socio-Economic Planning Sciences, 49,* 16-32.

[16] See Ali, C., Radcliffe, D., Schmidt, T.R., Donald, R. (in press). Searching for Sheboygans: On the future of small market newspapers. *Journalism* (describing a "knowledge gap" in the study of small market news organizations).

[17] Hamilton, J. (2004). *All the news that's fit to sell: How the market transforms information into news.* Princeton, NJ: Princeton University Press.

[18] Translator stations generally just relay the signal of a larger nearby station. Thus any news provided by such stations is already accounted for in our capturing of the news output of the "parent" station. Low power stations may be more likely to be producing journalism that is not captured by our online archiving protocol.

[19] https://archive.org/.

[20] It is worth noting that, in some instances, more than one outlet shared the same URL. This typically occurred in relation to radio stations. Because radio stations often are part of large ownership groups, it was sometimes the case that an individual radio station's home page would be (or would redirect to) the home page for the entire station group. In such instances, the home page was analyzed each time it arose in connection with an individual community.

[21] For a more detailed discussion, see Napoli, et al. (2017).

[22] The complete archive is available online at https://archive-it.org/collections/7520. Note that the actual archive today is larger than described in the text. The archive was created with specific boundaries for the purposes of this research, but the Internet Archive decided to continue to data collection effort on a monthly basis to maintain the record of local news websites.

[23] Lacy, S., & Rosenstiel, T. (2015). Defining and measuring quality journalism. Report from the News Measures Research Project, http://mpii.rutgers.edu/wp-content/uploads/sites/129/2015/04/Defining-and-Measuring-Quality-Journalism.pdf.

[24] See, e.g., Knight Commission (2009). *Informing communities: Sustaining democracy in the digital age,* http://www.knightcomm.org/read-the-report-and-comment/; Waldman, S. (2011). *The information needs of communities: The changing media landscape in a broadband age.* Federal Communications Commission, Washington, D.C., http://www.fcc.gov/infoneedsreport.

[25] E.g., Becker and Yanich (2015); Napoli, et al. (2017).

[26] Friedland, L., Napoli, P.M., Ognayanova, K., Weil, C., & Wilson, E.J. (2012). *Review of the literature regarding critical information needs of the American public.* Washington, DC: Federal Communications Commission.

[27] See Friedland et al. (2012) for extensive and detailed descriptions and discussions of each of these critical information needs categories. .

[28] See, e.g., Pew Research Center (2010).

[29] See, e.g., George, & Waldfogel (2006).

[30] See, e.g., Anderson, C.W., Bell, E., & Shirky, C. (2012). Post-industrial journalism: Adapting to the present, http://towcenter.org/research/post-industrial-journalism/.

[31] George & Waldfogel (2006).

[32] See, e.g., Patterson, T. (2013). *Informing the news: The need for knowledge-based journalism.* New York: Vintage.

[33] Prior to training, two pilot tests on a sample of data were conducted by the researchers in order to identify data-gathering challenges and difficulties interpreting or applying the coding categories. These pilots refined the training and content analysis protocols. Coders underwent two trial sessions using data samples (followed by assessment and debriefings) in order to identify any challenges or uncertainties associated with the content analysis protocol and code book. In identifying and coding individual news stories, coders were instructed and trained to exclude social media feeds, widgets, photo galleries, event calendars, outlet promotional announcements, and advertisements from their coding activity. Coders also coded only those stories with publication dates that matched the dates in the constructed week sample. Google Translate was used to facilitate coding of foreign-language content. A dedicated communication channel facilitated with the app Slack was established and maintained during the entirety of the coding process, so that coders could share questions or challenges that emerged during the coding process and obtain immediate feedback and guidance.

Using Riffe, et al.'s (2013) formula for determining an intercoder reliability sample, a random sample of cases was selected for an intercoder reliability check by an additional trained coder. Using Krippendorff's alpha, intercoder reliability was .81 for the Original variable, .84 for the Local variable, and .73 for the Critical Information Needs variable. See Riffe, D., Lacy, S., & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research (3rd ed.).* New York: Routledge.

[34] Napoli, et al. (2017).

[35] Napoli, et al. (2017).

[36] In our analyses, we have elected to treat the communities sampled in the same way that survey researchers treat survey respondents. That is, the individual communities sampled and analyzed remain anonymous in our reporting. We have taken this approach in order to maintain a focus on the issue of *community characteristics* and to not prompt discussions, critiques, or defenses of individual communities. We believe is is the broader patterns exhibited in this research, not the identities of the individual communities, that are of the greatest significance.

[37] The percentages presented in this figure thus represent the mean of the percentages generated by each story type for each community, rather than overall story type percentages, in keeping with our focus on the individual community as the unit of analysis.

[38] Because the dependent variables are count variables, the use of traditional OLS regression is not appropriate. Instead, the use of Poisson regression was explored. However, none of the four dependent variables followed the necessary Poisson distribution, with Kolmogorov-Smirnov Zs ranging from 3.6 to 6.3 ($p < .01$). These results are a reflection of the overdispersion of the dependent variables. Thus, negative binomial regression was employed for each of the four dependent varies (STORIES, ORIGINAL, LOCAL, CIN). Given the substantial number of zero counts across the dependent variables (ranging from eight in the case of STORIES and CIN to 20 for LOCAL) we recognize that a zero inflated negative binomial regression may represent a further improvement over the negative binomial regression approach employed here. However, initial exploration of a zero inflated negative binomial regression indicated no significant differences from the negative binomial regression; however, additional exploration of the zero inflated negative binomial regression is ongoing. Given the substantial number of zero counts across the dependent variables (ranging from eight in the case of STORIES and CIN to 20 for LOCAL) we recognize that a zero inflated negative binomial regression may represent a further improvement over the negative binomial regression approach employed here. However, initial exploration of a zero inflated negative binomial regression indicated no significant differences from the negative binomial regression; however, additional exploration of the zero inflated negative binomial regression is ongoing.

[39] Collinearity diagnostics indicated no problematic levels of multicollinearity amongst the independent variables (tolerance statistics ranging from .49 to .81).

[40] Likelihood χ^2s ranged from 27.61 to 33.88 ($p < .01$). Note that negative binomial regression does not produce an R^2 similar to that produced in OLS regression.

[41] In previous research, focusing on the number of media outlets serving each state (Napoli, et al., 2017), we found a curvilinear relationship with population (suggesting that the extent to which the number of media outlets increases as a state's population increases tapers off as population size increases; see Authors, in press). Here, however, in focusing on journalistic output, it is worth noting that scatterplots revealed no such relationship.

[42] See, e.g., Hindman, M. (2009). *The myth of digital democracy.* Princeton, NJ: Princeton University Press.

[43] Pew Research Center (2010).

[44] Available at https://archive-it.org/collections/7520.