Published in final edited form as:

J Commun. 2009 September ; 59(3): 514. doi:10.1111/j.1460-2466.2009.01433.x.

Newspaper Coverage of Cancer Prevention: Multilevel Evidence for Knowledge Gap Effects

Michael D. Slater,
The Ohio State University

Andrew F. Hayes,
The Ohio State University

Jason B. Reineke,
The Ohio State University

Marilee A. Long, and
Colorado State University

Erwin P. Bettinghaus
Klein-Buendel, Inc

Abstract

Prior research on knowledge gap effects, in health as well as in other domains, has focused largely on assessing individual-level differences in exposure to news based on self-report of media use. Inherent inferential limitations of this approach are addressed by testing the hypothesis that the relationship between education and cancer prevention knowledge will be moderated by regional differences in U.S. news coverage of cancer prevention. The study also tests, using these methods, findings by Kwak (1999) suggesting that the importance of attention to relevant news in predicting knowledge decreases as information available in the news increases. Using a representative national sample of newspaper coverage to assess regional differences in cancer prevention coverage, a representative national probability sample to assess respondent education and cancer prevention knowledge, and multilevel analyses of the relationship between regional coverage differences and knowledge of persons in those regions, support is found for both of these propositions.

News coverage of health issues has been shown to predict associated knowledge, attitudes, and behavior in a variety of health domains, including substance use (Stryker, 2003) and cancer (Yanovitzky & Blitz, 2000). In particular, recent research has suggested the presence of knowledge gap effects in health domains—that is, the strength of the relationship between socio-economic status (SES) and health knowledge increases as the availability of relevant information in the social environment increases (Viswanath & Finnegan, 1996). In particular, one analysis has indicated that greater education predicts greater knowledge of the causes of cancer, as does attention to health news (Viswanath et al., 2006); another provides evidence that higher SES individuals are more likely to respond to media reports of celebrity diagnoses of cancer by seeking information than are lower SES individuals (Niederdeppe, 2008).

The present research seeks to advance these lines of inquiry by addressing key methodological limitations of much knowledge gap research. Such research typically depends upon respondents' self-report of news media use or attention to health information. However,

associating knowledge or information-seeking with self-reports of media use, as in Viswanath, et al., 2006, is somewhat problematic with respect to claims of causality. For example, persons who hold health as a priority (Lau, Hartman, & Ware, 1986; Slater & Flora, 1991) are probably more knowledgeable, more likely to adopt more healthy behaviors, and to report using more health-related media than are people less interested in health. These associations could generate spurious relationships, assuming that higher SES individuals may be less worried about paying rent and feeding family members and more able to give higher priority to health than lower SES individuals. In this case, one would expect higher SES individuals to have greater health knowledge and also report greater use of health-relevant media, for reasons having more to do with SES status than likely effects of greater media use.

In this study, we use a unique, nationally representative data set of news coverage to identify regional differences in the amount of news coverage regarding cancer prevention and to assess the knowledge gap effect—i.e., whether a relationship between news coverage and health knowledge is contingent on people's level of education (Viswanath & Finnegan, 1996). In so doing, we provide a conservative test comparatively free of individual-level problems of reverse causality and third-variable explanations associated with self-reports of media exposure.

News Coverage of Cancer

There is a growing literature on news coverage of specific cancers, particularly breast cancer. For example, a study of Canadian mainstream and ethnic media found that coverage tracked cancer incidence more closely than mortality and over-represented breast cancer overall (Hoffman-Goetz & Friedman, 2005). Two studies comparing African-American newspapers and mainstream U.S. newspapers found parallel results (Cohen et al., in press; Stryker, Emmons, & Viswanath, 2007).

A content analysis of a nationally representative sample of U.S. local and national news confirmed that coverage of cancer more closely tracked incidence rates rather than mortality rates. Results also showed a tendency to over-report breast cancer and under-report lung and other cancers relative to both incidence and mortality. Of particular interest for the present study is that news coverage tended to emphasize treatment and causes of cancer, and appeared less attentive to cancer prevention and detection (Slater, Long, Bettinghaus, & Reineke, in press). That study, however, did not examine whether the amount of cancer prevention coverage influenced cancer knowledge.

Several previous efforts have sought to link content analysis data with survey data from NCI's Health Information National Trends Survey (HINTS), a national, random-digit-dial sample of more than 6,000 respondents (Nelson et al., 2004). One such study compared a content analysis of large-market newspapers and ethnic newspapers to HINTS survey data in an effort to link coverage, self-reported attention, and cancer prevention behaviors (Stryker, Moriarty, & Jensen, 2008). The authors concluded that increased coverage of cancer prevention could impact cancer prevention knowledge. The major limitation of that study, however, was that survey and content analytic findings were related to one another descriptively, rather than through inferential statistical tests. In this study, we focus on coverage of cancer prevention to follow up the Stryker, et al. (2008) results, as our study provides the opportunity to assess their descriptive claims through inferential analysis. In addition, we believe prevention coverage is of particular substantive interest given that such coverage in general is inadequate (Slater et al., in press), so variation in coverage is likely to have a meaningful impact.

At least one study has statistically tested the link between news content and health knowledge and possible knowledge gap effects. Niederdeppe (2008) used HINTS in combination with an automated content analysis of Associated Press news stories about celebrity cancer diagnoses

or deaths from the Lexis-Nexis data base. The HINTS survey included a question about how long ago the respondent had searched for health information about cancer. Respondents who had searched within a week of media reports of celebrity cancer instances were thereby identified. Results indicated that while the media reports did not significantly increase information-seeking overall, higher SES individuals were more likely to seek cancer information after a media report of a celebrity cancer instance than were lower SES individuals.

These results, by using a measure of potential media exposure independent of respondents' self-reports, are an important contribution to our understanding of knowledge gap processes. However, there are two major limitations. One is that media reports of celebrity diagnoses are only one of several common types of cancer news story (Slater et al., in press). The other is that the Niederdeppe (2008) study focused on information-seeking behavior, rather than actual differences in knowledge. Knowledge gap research is primarily concerned with actual knowledge, although it is certainly likely that information-seeking is predictive of knowledge.

In the present study we use an approach developed to evaluate tobacco-prevention campaigns (Farrelly et al., 2002; Emery et al., 2005); we associate data on regional differences in media exposure to cancer prevention news stories with individual-level data on prevention knowledge. Based on descriptive results from a previous study (Stryker et al., 2008), we predict the following:

Hypothesis 1. People who live in regions with greater concentrations of cancer prevention information in cancer news stories will possess greater knowledge about cancer prevention than people who live in regions with lesser concentrations of cancer prevention information in cancer news stories.

Testing Knowledge Gaps Using Regional Data on News Content

The knowledge gap hypothesis argues that news media exacerbate knowledge disparities between high and low SES individuals, because high SES individuals have greater access to and make more efficient use of information from these media (Viswanath & Finnegan, 1996). A premise of the knowledge gap hypothesis in this context, therefore, is that education and cancer health knowledge are positively related; prior research using the HINTS data set supports this premise (Viswanath et al., 2006).

Consistent with prior knowledge gap research (e.g., Viswanath et al., 2006; Niederdeppe, 2008), we also would expect that the relationship between availability of relevant health information in regional newspapers and corresponding health knowledge will be moderated by indicators of SES, such as education. Given the Viswanath et al. (2006) finding that cancer knowledge was associated primarily with education and not with income, we predict the following with respect to the regional news coverage differences:

Hypothesis 2. The relationship between the proportion of cancer prevention news content in a given region and cancer prevention knowledge will increase contingent on respondent education.

It seems plausible that attention might also moderate the relationship between news coverage and health knowledge. When news reports on cancer prevention are rare, more attentive readers may be more likely than less attentive readers to gain knowledge. However when these news stories are more common, the effect of attention may be smaller because less attention is required to find the news stories (see Kwak, 1999, for relevant findings from the political communication domain). Alternatively, it may be that attention to health news increases the impact of greater amounts of media coverage regarding cancer prevention, perhaps because the coverage is more carefully scrutinized (see Niederdeppe, Frosch, & Hornik, 2008, for

evidence for such an effect with respect to information-seeking behavior). Therefore, we pose the following question:

Research Question 1. Is the relationship between attention to health news and knowledge of cancer prevention moderated by the relative amount of cancer prevention news content, and if so, in what way?

Methods

This study combined data from a content analysis of local newspaper coverage that we believe to be the most nationally representative gathered to date, and a national, random-sample survey of cancer-related media use, information behavior, and cancer knowledge.

Content Analysis

The content analysis sample used two, constructed, 28-day months (Riffe, Lacy & Fico, 1998) from 2002 and 2003. For each month, a week was constructed from each three-month season, so every day during the year had an equal probability of inclusion and seasons as well as days of the week were counterbalanced for equal representation in the sample. The 210 Designated Market Areas (DMAs[®]), which cover more than 99% of the U.S. population (Standard Rate and Data Service, 2000), were divided into six strata of increasing market size, so that each stratum contained media markets covering about one-sixth of the U.S. population. On each day of the constructed month, three newspapers from each stratum were selected for study in the following way. A DMA belonging to the stratum was randomly selected, the major newspaper in that DMA was chosen (or was randomly selected if there was more than one major newspaper), and then two additional daily newspapers were selected, one from above the median circulation size for that DMA on that day and one from below the median. A total of 18 newspapers were selected per sampling day (with some exceptions for DMAs in which three papers were not available on a sampling day), and 956 local newspapers were selected for the full sample. Full details on the sampling strategy are provided elsewhere (Long, Slater, Boiarsky, Stapel, & Keefe, 2005).

Local television news stories were also sampled and studied. However, the occurrence of cancer prevention news on television was so rare as to make incorporation into these analyses inappropriate.

We double-coded a random selection of newspaper pages in order to test reliability in identifying cancer stories. Because of the large number of news items that any given edition of a newspaper prints, we randomly selected 10% of the pages with news content from each newspaper in the reliability sample. We used multiple coders at times, and therefore calculated Cohen's kappa for pairs of coders. We also retested reliability several times to minimize risk of coding drift. Initial kappas ranged from .64 to .86. At 25% completion, the kappa was .82; at 50% completion, kappas ranged from .83 to .95; and at 75% completion, the kappa was .74. Details on story identification procedures are summarized in Slater et al. (in press). A total of 633 cancer news stories were identified.

Presence or absence of a prevention focus in the cancer news story was determined by reading the headline and lead of the news story, which defines the primary news focus of the story. A prevention focus was defined as content that addressed protective behaviors, such as using sunscreen, eating a healthy diet, limiting alcohol, and exercising regularly; it did not include articles about screening and diagnostic activities, such as getting Pap smears, PSA tests, or mammograms. Two coders tested the reliability of identifying cancer prevention news stories on 15% of the sample; they obtained a kappa of .76 for primary cancer topic. Details on coding procedures may be found in Slater et al. (in press).

One challenge in this study was to define the regional units used to test regional differences in prevention news content and prevention knowledge given the strengths and limitations of our data set. We crossed our six-market strata with four geographic groupings (northeast, southeast, midwest, and west) to capture both market size and regional differences; this approach resulted in 23 regional units (geographic groupings were based on standard federal distinctions with minor exceptions, e.g., grouping Baltimore with Pittsburgh given the former's geographic proximity to Pennsylvania and apparent dissimilarity to southern cities of similar market size). There were no cities in the southeast large enough for inclusion in the largest market strata (which included New York, Chicago, and Los Angeles), so we had 23, rather than 24 regional units. This approach allowed us to group markets of similar size (e.g., rural, small town, larger city, major city) in the same section of the country. We deemed this method of grouping markets to be superior to the other logical approach of grouping markets by state because divisions by state would ignore market size; since large-market and small-market papers vary greatly in terms of staff size, specialization, and sheer volume of news, it seemed important to address this variable in our design. For example, grouping media markets by state would lump rural upstate New York with New York City, agricultural California with Los Angeles, etc.

Media coverage of cancer was not sufficient to make more fine-grained regional divisions (e.g., by county or individual DMA) feasible. Conversely, ignoring sections of the country and simply using market strata (or vice versa) seemed excessively crude; moreover, such approaches would not provide enough upper-level units for an appropriate multi-level modeling analysis.

Therefore *region* refers to areas of similar market size within each of the four major geographic divisions of the United States, so that, for example, highly rural Massachusetts, New York, and Pennsylvania would be grouped together as would be Philadelphia and Boston, consistent with the six market size groups within the northeastern United States (see Table 1 for details on assignment of markets to upper level groups).

Regional newspaper emphasis on cancer prevention coverage, the focal independent variable in the analysis, was operationalized as the number of cancer prevention stories divided by the total number of cancer stories identified in that region. Newspaper size differs within market strata; a small newspaper in a large DMA might have 18 pages of editorial content, while a major city paper in that same DMA might have 50 pages. Examining the relative emphasis on prevention as a percentage of the newspaper's total coverage of cancer permitted us to adjust for size differences and make more meaningful comparisons between regions. Moreover, it seemed plausible to us that, in larger papers, most readers read fewer of the total stories given the sheer volume of coverage, and thus relative emphasis on cancer prevention would be a more sensible indicator of the reader's experience regarding cancer information in that paper. Table 1 provides the total number of cancer-related stories in the sample in each region as well as the percent of those stories that were cancer prevention-focused. There were significant differences among the 23 regions in the percentage of cancer coverage that was prevention-focused, $\chi^2(22) = 36.68, p < .05$, with the percentage of prevention-focused cancer coverage varying between 0 and 36.4%.

The HINTS Survey

The 2003 Health Information National Trends Survey (HINTS) is a random-digit-dial, probability sample telephone survey, $n = 6369$, conducted by the National Cancer Institute. Numbers were generated from a sampling frame of all telephone exchanges in the country, between October 2002 and April 2003. Respondents had to be age 18 or older, and African-Americans and Hispanics were oversampled. The response rate was 33%, including adjustments for the estimated incidence of residential telephone numbers (see Nelson et al.,

2004; full details are available on line,
http://hints.cancer.gov/docs/HINTS_2003_final_report.pdf).

Variables used in this analysis from the HINTS survey included respondents' level of education, age, sex, attention to medical and health news in newspapers, and cancer prevention knowledge. HINTS respondents were 60.4% female with a mean age of 47.74 years ($SD = 17.42$). Education was measured on a four-point scale: not completing high school (1); high school graduate or GED (2); some college—i.e., three or fewer years of college (3); or college graduate—i.e., four years or more of college (4). The resulting variable had a mean of 2.77 and a standard deviation of 1.02. As discussed below, education was coded for analysis using a sequential coding system rather than treating it as an interval-level scale. Attention to medical and health news in newspapers was measured by respondents' answers to the question "How much attention do you pay to information about health or medical topics in newspapers?" Response options and numerical codes included "not at all" (0), "a little" (1), "some" (2) or "a lot" (3), ($M = 1.58$, $SD = 1.11$).

Cancer-prevention knowledge was operationalized using responses to the open-ended question "Can you think of anything people can do to reduce their chances of getting cancer?" Respondents who provided an answer were probed until they indicated that they could not think of any other approaches to reducing cancer risks, or until they had given nine responses. Information on what responses could be considered correct was obtained from the National Cancer Institute. For each respondent the number of correct answers given was tallied to indicate cancer prevention knowledge, $M = 1.97$, $SD = 1.40$.

In addition to individual responses to the survey, the HINTS data set includes the DMA in which each respondent is located; these data made it possible to link the HINTS data set to the region/market size variable described earlier. Table 1 provides the number of respondents to the HINTS survey in each of the 23 regions.

Missing Survey Data

Of the 6,369 respondents to the HINTS survey, 232 (3.6%) either did not respond to or were not asked at least one of the questions used as predictors in the models reported below. More specifically, all 232 of these respondents had no information recorded about level of education. Furthermore, 14 of these 232 had no newspaper attention response recorded, and 26 did not have age data. We used a hot-deck imputation procedure (Hawthorne & Elliott, 2005; Little & Rubin, 2002; Nordholdt, 1998; Reilly, 1993) to impute missing values for these variables. This procedure replaces a respondent's missing value with the value from a randomly chosen respondent in the same "deck," with decks defined for this study by the combination of respondent sex and region, two variables that were complete for all respondents. For example, if a male from region 2 was missing education information, his education value was imputed by randomly selecting another male from region 2 and using that man's education level as the value for the missing one. This procedure ensures that the distribution of responses post-imputation will be the same, within random sampling error, as the distribution pre-imputation. To control for the possible effects of missing values on the outcome variable, we created a dummy variable by coding whether a case was imputed on any variable in the model (0 = no, 1 = yes); we then used the dummy variable as an additional predictor in the models reported below.

Results

We analyzed the data with a multilevel Poisson model in HLM 6.0 (Scientific Software International, 2004) using quasipenalized maximum likelihood and a random intercept. The Poisson model was chosen because the outcome variable is a discrete count and bounded at 0

(for those respondents who didn't correctly provide any means of preventing cancer) rather than continuous and unbounded. Although Poisson models frequently exhibit overdispersion when applied to real data (meaning the conditional variance is typically larger than the conditional count rather than equal to the conditional count, as the pure Poisson model assumes), including a random intercept is recommended as a means of reducing or eliminating the overdispersion when observations (here, survey respondents) are nested within higher order clusters (in our case, region). Furthermore, allowing the intercept to vary randomly accommodates the non-independence that is typical in multilevel data problems such as ours, increasing the accuracy of estimated standard errors (Hayes, 2006; Raudenbush & Bryk, 2002). We also used a robust standard error estimator to further reduce the likelihood that model misspecification would invalidate hypothesis tests resulting purely from inaccurate standard errors. See Rabe-Hesketh and Skrondal, (2005, p. 181-216) for relevant analytical details, Long (1997) and Garner, Mulvey, and Shaw (1995) for a discussion of count models, and Hayes (2006) and Raudenbush and Bryk (2002) for introductory and more advanced treatments, respectively, of multilevel models.¹

Prior to analysis, we first examined the legitimacy of treating education and level of attention as interval-level variables. This was accomplished by regressing cancer prevention knowledge on education (in the same kind of multilevel Poisson model), with education coded using an ordinal, sequential, dummy coding system (see Darlington, 1990, pp. 236-237 for details). For a four-level ordinal variable, this system requires three dummy codes chosen such that the first dummy codes the contrast between the lowest and second lowest ordinal category (S1 in the models reported below), the second codes the contrast between the second lowest and second highest category (S2 in models below) and the third codes the contrast between the second highest and the highest category (S3 in models below). The same procedure was used by regressing knowledge on attention, coded as a set of three sequential dummy codes. We found that for education, the coefficients for the three dummy codes differed statistically from each other, meaning that a single step on the ordinal scale was not associated with a consistent increase in cancer prevention knowledge. In other words, a single coefficient for education would not describe the relationship between education and knowledge as well as one that allowed for knowledge to change nonlinearly (as a step function) with education. For this reason, we retained the sequential dummy coding system in the main analyses reported below. However, for attention, we found no differences between the coefficients for the sequential dummy codes, meaning that it was sensible to treat it as an interval-level variable and estimate the effect of attention with a single coefficient rather than three.

We first tested whether there was any regional variation in cancer prevention knowledge by estimating an empty model containing a random intercept with no predictors. However, we modified the standard procedure slightly by including the missing variable dummy described above at the individual level, grand mean centered, to control for any regional variation in the frequency of imputation of missing data. To remain consistent, we followed this procedure in all of the analyses that follow. Of interest is the variance component for the intercept, which quantifies the variance in cancer prevention knowledge across regions (adjusting for regional differences in the use of imputation at the individual level). As can be seen in Table 2 (model

¹To verify that our results were not an artifact of a particular choice of software implementing particular estimation algorithms, we also estimated the models described here using the **gllamm** procedure in Stata version 10 (StataCorp, 2007). The results in all models were largely the same, with similar coefficients and *p*-values. We report only the results from HLM here. For interested readers, the Stata commands used were

```
generate cons = 1
```

```
eq ri: cons
```

```
gllamm know ivs, family(poisson) i(area) robust eqs(ri)
```

where *ivs* was the list of independent variables in the model (which varied across models; see Table 2 and text), *know* was cancer prevention knowledge, and *area* was the variable coding region.

1), there was indeed significant variation in cancer knowledge across regions, as the variance component for region (the random intercept in the model) was statistically different from zero, $\chi^2(22) = 42.723, p < .01$.

Hypothesis 1 predicts a positive relationship between cancer prevention knowledge in a region and cancer prevention news content within that region. As can be seen in Table 2 (model 2), in regions where cancer prevention coverage was greater, so too was knowledge of how cancer can be prevented. Note that regional prevention coverage predicted knowledge even after controlling for the influence of sex, age, education, and attention to health information in the newspaper (all individual-level variables, grand mean centered) on such knowledge (model 3). Greater knowledge of cancer prevention was also associated with higher education (replicating the standard knowledge gap phenomenon) and attending more to health information in the newspaper, but not with age. Furthermore, males had lower knowledge of cancer prevention methods than did females. These individual differences consistently predict cancer prevention knowledge in all models we report.

Next, we considered the possibility that the relationship between cancer prevention content and cancer prevention knowledge was spurious, attributable to other regional and market size differences correlated with content. We focused on regional/market-size differences in average education and average attention to health-related news (for example, it is possible that there is less prevention content in smaller markets and also lower average education, or a greater interest in health-related issues in the western United States). Both of these variables were constructed by averaging responses to the attention and education questions from respondents within region. As can be seen in Table 2 (model 4), in regions with higher average education, cancer prevention knowledge was higher on average. Furthermore, controlling for average regional education and attention eliminated the relationship between cancer news content and knowledge. Interestingly, this effect of regional variation was larger than the individual-level effect of education, but the individual-level effect remained positive and statistically significant. In other words, respondents from more educated regions were more knowledgeable about cancer prevention than respondents from less educated regions, and so too were the relatively more educated from within a region relative to those less educated in that same region. Therefore, our analyses supported Viswanath et al.'s (2006) finding that education predicts cancer knowledge, as we found it based on regional/market size differences in education as well as at the individual level. However, after applying educational controls, we did not find convincing support for Hypothesis 1, that differences in regional/market size cancer prevention news content influenced cancer prevention knowledge independent of other contingent factors.

Model 5 in Table 2 was used to test Hypothesis 2, which tests the classic knowledge-gap assertion as to whether increased content translates into increased knowledge more so among the more educated. This model specifies a cross-level interaction between regional cancer content and individual-level education, and, as noted, provides a novel test of the knowledge gap hypothesis. Because education is coded with three sequential codes, three cross-level interactions are specified, one for each contrast representing increasing levels of education. As can be seen, regional cancer prevention content interacts with education, but the differential effect occurs only among those who have not attended college versus having attended some college ($S2 \times$ Prevention Content in Table 2). Differential cancer prevention content seems to have no differential effect among the least educated (no high school diploma versus a high school graduate) or most educated (some college versus having a college degree). Figure 1 plots this interaction (with “relatively little” content defined as 0%, “relatively moderate” as 10%, and “relatively high” as 20%). Observe that the slope for content is positive among the more educated (those who attended at least some college), but flat or even somewhat negative among the less educated (those whose education went no further than high school). Reframed

somewhat differently (focusing on content rather than education as the moderator), the gap in knowledge as a function of education is somewhat greater in higher content communities compared to lower content communities.

A similar approach was used to answer Research Question 1 by including a cross-level interaction between regional cancer content and individual attention to health news (Model 6 in Table 2). Indeed, the relationship between attention and knowledge varied across regions that differed in cancer-prevention content. Although greater attention to news stories about health topics was generally associated with more cancer prevention knowledge, the negative coefficient in Model 6 means that the relationship between attention and knowledge was smaller (but still positive) in communities with a greater proportion of their cancer stories focused on cancer prevention. This interaction is depicted graphically in Figure 2. Although the interaction is fairly subtle, it is statistically different from zero in the direction consistent with Kwak's (1999) finding in the political realm.

We estimated a final model to examine whether the interactions observed in Models 5 and 6 were unique, i.e., existed independent of each other. Model 7 includes cross-level interactions between regional cancer prevention content and both education and attention. As can be seen, these interactions persisted even after controlling for each other.

Discussion

These findings contribute to the knowledge gap literature in several ways. First, we provide the first test of the knowledge gap hypothesis using both a national sample of news content and a national sample of survey data. Our results provide inferential tests to confirm Stryker et al.'s (2008) more descriptive findings. This study, though not free from limitations as noted below, offers a test of knowledge gap effects that is not subject to concerns about the validity of individual self-reports regarding news exposure. Similarly, this design is not subject to third-variable history effects that might confound tests of the effects of communication campaigns or news coverage of emerging issues on knowledge gaps.

Second, the examination of ordinal effects suggests that the moderating effect of education is not linear, but is a function of whether the respondent had attended at least some college. Future knowledge gap research might examine whether this non-linear moderator effect exists in other contexts in health and in political communication.

Third, these findings extend knowledge gap research in the health domain to the effects of routine news coverage of health topics such as cancer prevention. These results suggest that knowledge gap effects in health do not merely arise from specific efforts to communicate about health issues in campaigns, which typically have been the focus of knowledge gap research in the health domain, but are a pervasive function of news use and availability in the United States.

Therefore, efforts to reduce knowledge-related health disparities need to address differences in access to and use of information available through the news as well as through access to health providers. From this perspective, initiatives to supply African-American newspapers with health information via specialized press services (Cohen et al., in press) seem well-conceived, to the extent that we can presume the populations served by these papers are lower on average in socio-economic status than the population as a whole, or at least that their readers may be in a position to convey information to others who may be of relatively lower socio-economic status.

Fourth, these findings replicate Kwak's (1999) findings that attention can reduce knowledge differences associated with differences in media coverage. In particular, our evidence suggests that attention is especially important when the news coverage of interest is relatively rare in a

given media market. This replication is of particular interest given that we find these results in a different domain (health instead of political communication) and using different methods—in our case, a multilevel study using both a nationally representative content analysis and survey.

The pattern of the interaction is inconsistent with Niederdeppe et al.'s (2008) findings regarding attention and information-seeking, in which attention increases the impact of greater coverage. We regard this inconsistency as due to the difference in the outcome variable (knowledge in our study, information-seeking in the Niederdeppe et al. [2008] study). It is likely that an underlying concern about health would lead to both attention to and seeking of cancer information, resulting in a reasonably strong covariance between the latter two variables. Greater exposure to cancer-related coverage, then, might be more likely to occasion information-seeking about cancer among these already-motivated, attentive readers. Such a process would be consistent with the strong relationship between attention and information-seeking, and the modest relation between coverage and information-seeking, found in the Niederdeppe et al. (2008) study.

In contrast, attention to health news and our outcome variable, cancer prevention knowledge, are unlikely to be as strongly related as are attention and cancer information-seeking. Knowledge presumably is an outcome of information-seeking. It is therefore at least one step removed from attention, and the relationship between attention and prevention knowledge therefore is likely to be weaker. Moreover, since, as we show, prevention coverage is relatively uncommon, cancer news may occasion cancer information-seeking but that seeking would rarely be about prevention topics. Therefore, it seems plausible that, in the case of an under-covered topic such as cancer prevention, attention is particularly relevant in that it leads to such coverage being noticed, a necessary prerequisite to affecting prevention knowledge.

Substantively, the finding that increased attention to the news seems to compensate at least in part for the effect of reduced news content on knowledge (or, conversely, increased information may compensate for less attention) may be encouraging. It is not clear whether the reason for this attention effect is greater attention to whatever news is in fact available to the news consumer, or greater motivation to engage in information-seeking. In either event, this result does suggest that efforts to raise the salience of health topics might help reduce knowledge disparities. It may be that for low SES populations preventive health concerns may be a less immediate priority than day-to-day maintenance needs; however, it appears that the greater the information in the environment, the less essential such attention becomes. Therefore, providing information through various community channels to low SES populations makes sense in terms of addressing unequal access to mediated information and less academic training for processing such information, and compensating for what perhaps may be less attention to such information given other pressing life needs.

Limitations to this study must also be noted. As in any study of this type, we cannot be certain that we have eliminated all alternative explanations for differences between regional/market size groups at level two that might co-vary with differences in cancer prevention coverage. We also recognize, as discussed above, that we are assuming that the percentage of cancer prevention stories is as informative as the actual amount of coverage. In other words, we focus on how likely it is that people will find a story focusing on cancer prevention when they encounter a news story on cancer. We believe this is a reasonable focus, since it permits comparisons across market size. However, it reflects our assumption that absolute amounts of coverage can be misleading because in larger papers much of this coverage is likely to be missed by most readers. To the extent to which this assumption is inaccurate, however, our analyses would be biased in a conservative direction by adding error.

In addition, although our 23 level-two region/market size categories were appropriate given our design and emphasis on national random samples, they are nonetheless relatively crude and heterogeneous given the size and diversity of the United States. As a result, we regard these analyses as conservative and our parameter estimates for findings as lower-bound estimates. The importance of the findings, as noted above, is not in provision of robust parameter estimates for the population but in triangulating socially important knowledge gap findings using an alternative methodology. Future research might focus on DMA-level analyses using Lexis-Nexis data bases to provide more fine-grained results and more robust parameter estimates, though at a sacrifice in representativeness of news coverage studied (notably, through loss of small-market newspapers studied here).

This study's limitations, overall, are counterbalanced by our ability to address limitations in previous research associated with the use of self-report measures of media exposure, limited content analysis samples, and the use of descriptive approaches to link content analyses with survey responses. In sum, the new evidence we provide regarding knowledge gap effects at the regional/market size level suggests that knowledge disparities due to news content differences deserve increased attention from both theoretical and substantive perspectives.

Acknowledgments

This research was supported by grant CA109694 from the National Cancer Institute to the fifth author and grant AA10377 from the National Institute on Alcohol Abuse and Alcoholism to the first author.

References

- Cohen, EL.; Caburnay, CA.; Luke, D.; Rodgers, S.; Cameron, GT.; Kreuter, M. Cancer coverage in general audience and black newspapers. *Health Communication*. In: Darlington, RB., editor. Regression and linear models. New York: McGraw-Hill; 1990. In press
- Emery S, Wakefield MA, Terry-McElrath Y, Saffer H, Szczypka G, O'Malley PM, Johnston LD, Chaloupka FJ, Flay B. Televised state-sponsored anti-tobacco advertising and youth smoking beliefs and behavior in the United States, 1999-2000. *Archives of Pediatrics and Adolescent Medicine* 2005;159:639-645. [PubMed: 15996997]
- Farrelly M, Heaton C, Davis K, Messeri P, Hersey J, Haviland M. Getting to the truth: Evaluation national tobacco countermarketing campaigns. *American Journal of Public Health* 2002 June;92(6):901-907. [PubMed: 12036775]
- Garner W, Mulvey EP, Shaw EC. Regression analysis of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin* 1995;118:392-404. [PubMed: 7501743]
- Hawthorne G, Elliott P. Imputing cross-sectional missing data: Comparison of common techniques. *Australian and New Zealand Journal of Psychiatry* 2005;39:583-590. [PubMed: 15996139]
- Hayes AF. A primer on multilevel modeling. *Human Communication Research* 2006;32:385-410.
- Hoffman-Goetz L, Friedman D. Disparities in coverage of cancer information in ethnic minority and mainstream print media. *Ethnicity and Disease* 2005;15:332-340. [PubMed: 15825981]
- Kwak N. Revisiting the knowledge gap hypothesis: Education, motivation, and media use. *Communication Research* 1999;26:385-413.
- Lau RR, Hartman KA, Ware JE. Health as a value: Methodological and theoretical considerations. *Health Psychology* 1986;5(1):25-43. [PubMed: 3720718]
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2. Hoboken, NJ: Wiley; 2001.
- Long, JS. Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage Publications; 1997.
- Long M, Slater MD, Boiarsky G, Stapel L, Keefe T. Obtaining nationally representative samples of local news media outlets. *Mass Communication and Society* 2005;8:299-322.
- Niederdeppe J. Beyond knowledge gaps: Examining socioeconomic differences in response to cancer news. *Human Communication Research* 2008;34:423-447.

- Niederdeppe J, Frosch DL, Hornik RC. Cancer news coverage and information-seeking. *Journal of Health Communication* 2008;18:181–199. [PubMed: 18300068]
- Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, Rimer BK, Viswanath KV, Weinstein N, Alden S. The Health Information National Trends Survey (HINTS): Development, design, and dissemination. *Journal of Health Communication* 2004;9:443–60. [PubMed: 15513791]
- Nordholt ES. Imputation: Methods, simulation experiments, and practical examples. *International Statistical Review* 1998;66:157–180.
- Rabe-Hesketh, S.; Skrondal, A. Multilevel and longitudinal modeling using Stata. College Station, TX: Stata Press; 2005.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2. Thousand Oaks, CA: Sage Publications; 2002.
- Reilly M. Data analysis using hot deck multiple imputation. *The Statistician* 1993;42:307–313.
- Riffe, D.; Lacy, S.; Fico, F. Analyzing media messages: Using quantitative content analyses in research. Mahwah, NJ: Erlbaum; 1998.
- Scientific Software International. HLM version 6: Linear and nonlinear modeling [Computer software]. Lincolnwood, IL: Author; 2004.
- Slater MD, Flora JA. Health lifestyles: Audience segmentation analysis for public health interventions. *Health Education Quarterly* 1991;18:221–223. [PubMed: 2055779]
- Slater MD, Long M, Bettinghaus EP, Reineke JB. News coverage of cancer in the U.S.: A national sample of newspapers, television and magazines. *Journal of Health Communication*. In press.
- Standard Rate and Data Service. SRDS TV & cable source. Des Plaines, IL: SRDS, Inc; 2000.
- StataCorp. Stata Statistical Software: Release 10 [Computer software]. College Station, TX: StataCorp LP; 2007.
- Stryker J. Media and marijuana: A longitudinal analysis of news media effects on adolescents' marijuana use and related outcomes. *Journal of Health Communication* 2003;8:305–328. [PubMed: 12907398]
- Stryker JE, Emmons KM, Viswanath K. Uncovering differences across the cancer control continuum: A comparison of ethnic and mainstream cancer newspaper stories. *Preventive Medicine* 2007;44:20–25. [PubMed: 16949141]
- Stryker JE, Moriarty CM, Jensen JD. Effects of newspaper coverage on public knowledge about modifiable cancer risks. *Health Communication* 2008;23:380–390. [PubMed: 18702002]
- Tichenor PJ, Donohue GA, Olien CN. Mass media flow and differential growth in knowledge. *Public Opinion Quarterly* 1970;34:159–170.
- Viswanath K, Breen N, Meissner H, Moser RP, Hesse B, Steele WR, Rakowski W. Cancer knowledge and disparities in the information age. *Journal of Health Communication* 2006;11:1–17. [PubMed: 16641071]
- Viswanath, K.; Finnegan, JR. The knowledge gap hypothesis: Twenty-five years later. In: Bureson, B., editor. *Communication yearbook*. Vol. 19. Thousand Oaks, CA: Sage; 1996. p. 187–227.
- Yanovitzky I, Blitz C. Effect of media coverage and physician advice on utilization of breast cancer screening by women 40 years and older. *Journal of Health Communication* 2000;5:117–134. [PubMed: 11010345]

Figure 1.
Estimated Cancer Prevention Knowledge as a Function of Respondent Education and Regional Newspaper Cancer Prevention Content

Figure 2.
Estimated Cancer Prevention Knowledge as a Function of Respondent Attention to Health and Medical News and Regional Newspaper Cancer Prevention Content

Table 1

Regions, DMAs within Region in Sampling Frame, Hints Sample Size in Region, and Percentage of Cancer-Related Stories that were Prevention Focused

Region	DMAs	Size	Area	Hints <i>n</i>	Cancer Prevention Content in Sample (% of total cancer coverage)	Number of Cancer Stories in Sample
1	New York	1	NE	345	12.0	50
2	Los Angeles	1	W	277	5.3	57
3	Chicago	1	MW	172	3.9	52
4	Philadelphia, Boston, Washington DC	2	NE	508	10.5	38
5	Seattle/Tacoma, San Francisco/San Jose/Oakland	2	W	259	12.5	24
6	Dallas/Ft Worth, Atlanta, Houston	2	S	316	8.1	37
7	Detroit	2	MW	93	0	19
8	Tampa/St Petersburg, Miami/Ft. Lauderdale, Orlando/Daytona Beach/Melbourne	3	S	253	0	24
9	Indianapolis, St. Louis, Minneapolis-St. Paul, Cleveland	3	MW	314	13.6	22
10	Phoenix, Denver, San Diego, Portland OR, Sacramento/Stockton/Modesto	3	W	343	4.9	41
11	Pittsburgh, Baltimore	3	NE	110	0	9
12	Hartford/New Haven, Buffalo, Harrisburg/Lancaster/Lebanon/York, Providence/New Bedford	4	NE	190	4.8	21
13	Charlotte, Raleigh-Durham, Nashville, Greenville/Spartanburg/Asheville/Anderson, San Antonio, Birmingham, Memphis, New Orleans, Norfolk/Portsmouth/Newport News, West Palm Beach/Ft. Pierce, Oklahoma City, Greensboro/High Point/Salem, Louisville	4	S	520	5.7	35
14	Kansas City, Cincinnati, Milwaukee, Columbus OH, Grand Rapids/Kalamazoo/Battle Creek	4	MW	259	18.5	27
15	Salt Lake City, Albuquerque/Santa Fe	4	W	98	0	8
16	Jacksonville, Little Rock/Pine Bluff, Tulsa, Charleston/Huntington WV, Richmond/Petersburg, Austin, Mobile/Pensacola, Knoxville, Lexington, Roanoke/Lynchburg, Paducah/Cape Girardeau/Harrisburg/Mount Vernon, Shreveport, Ft. Meyers/Naples, Huntsville/Decatur, Chattanooga, Columbia, Jackson	5	S	407	4.4	45
17	Las Vegas, Fresno, Visalia, Tucson, Honolulu, Spokane	5	W	130	0	5
18	Dayton, Flint/Saginaw/Bay City, Wichita/Hutchinson Plus, Toledo, Green Bay/Appleton, Des Moines/Ames, Omaha, Springfield MO, Champaign/Springfield/Decatur, Madison, South Bend, Davenport/Rock Island/Moline, Cedar Rapids/Waterloo/Dubuque	5	MW	348	15.0	20

Region	DMAS	Size	Area	Hints <i>n</i>	Cancer Prevention Content in Sample (% of total cancer coverage)	Number of Cancer Stories in Sample
19	Wilkes/Barre/Scranton, Albany/Schenectady/Troy, Syracuse, Rochester, Portland/Auburn, Burlington/Plattsburgh	5	NE	196	36.4	11
20	Tri Cities TN-VA, Waco/Temple/Bryan, El Paso, Baton Rouge, Savannah, Harlingen/Weslaco/Brownsville/McAllen, Charleston, Greenville/New Bern/Washington, Tyler/Longview, Tallahassee FL/Thomasville GA, Augusta, Florence/Myrtle Beach, Montgomery, Ft. Smith/Fayetteville/Springdale/Rogers, Macon, Lafayette, Amarillo, Columbus GA, Corpus Christi, Columbus/Tupelo/West Point, Monroe LA/El Dorado AR, Beaumont/Port Arthur, Wheeling WV/Steubenville OH, Wichita Falls TX/Lawton OK, Lubbock, Bluefield/Beckley/Oak Hill, Albany, Odessa/Midland, Wilmington, Panama, Biloxi/Gulfport, Sherman/Ada, Salisbury, Abilene-Sweetwater, Clarksburg-Weston, Gainesville, Hattiesburg/Laurel, Dothan, Jonesboro, Harrisonburg, Greenwood/Greenville, Bowling Green, Meridian, Jackson TN, Charlottesville, San Angelo, Victoria	6	S	517	0	43
21	Colorado Springs/Pueblo, Reno, Monterey/Salinas, Santa Barbara/Santa Maria/San Luis Obispo, Eugene, Yakima/Pasco/Richland/Kennewick, Boise, Bakersfield, Chico-Redding, Medford/Klamath Falls, Anchorage, Palm Springs, Idaho Falls/Pocatello, Billings, Missoula, Yuma/El Centro, Great Falls, Grand Junction/Montrose, Twin Falls, Eureka, Butte/Bozemon, Cheyenne/Scottsbluff/Casper-Riverton, Bend, Fairbanks, Juneau, Glendive	6	W	294	9.1	11
22	Johnstown/Altoona, Erie, Binghamton, Bangor, Utica, Elmira, Watertown, Presque Isle	6	NE	113	18.2	11
23	Evansville, Youngstown, Lincoln/Hastings/Krny Plus, Ft. Wayne, Peoria/Bloomington, Sioux Falls, Fargo/Valley City, Traverse City/Cadillac, La Crosse/Eau Claire, Duluth/Superior, Rockford, Wausau/Rhineland, Topeka, Terre Haute, Sioux City, Columbia/Jefferson City, Joplin MO/Pittsburg KS, Minot/Bismark/Dickinson, Rochester/Mason City/Austin, Quincy/Hannibal/Keokuk, Rapid City, Marquette, Parkersburg, St. Joseph, Lafayette, Mankato, Ottumwa-Kirkville, Lima, Alpena	6	MW	307	8.7	23

Note: NE = Northeast, MW = Midwest, S = South, W = West

Table 2

Multilevel Poisson Models of Cancer Prevention Knowledge

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Level-1 Variables							
Intercept	0.679	0.678	0.257	0.259	0.259	0.261	0.257
Age			-0.001	-0.001	-0.000	-0.000	-0.000
H.S. Graduate vs. No H.S. Diploma (S1)			0.256***	0.257***	0.260***	0.256***	0.258***
Some College vs. High School Grad (S2)			0.172***	0.169***	0.168***	0.169***	0.168***
College Degree vs. Some College (S3)			0.217***	0.216***	0.215***	0.216***	0.215***
Male			-0.081***	-0.081***	-0.082***	-0.081***	-0.082***
Attention to Health/Medical News			0.068***	0.069***	0.069***	0.069***	0.069***
Any missing data	-0.318***	-0.318***	-0.303***	-0.303***	-0.306***	-0.306***	-0.307***
Level-2 Variables							
Prevention Focus Content		0.003*	0.002*	0.001	-0.002	0.001	-0.003
Mean Regional Education				0.179**	0.179**	0.179**	0.180**
Mean Regional Attention				-0.111	-0.115	-0.113	-0.119
Cross-Level Interactions							
Prevention Content × S1					-0.003		-0.002
Prevention Content × S2					0.008**		0.009**
Prevention Content × S3					0.000		0.001
Prevention Content × Attention						-0.001**	-0.002**
Random Effects							
Intercept	0.0017**	0.0014*	0.0004	0.0001	0.0001	0.0001	0.0001
Residual	0.9841	0.9842	0.9028	0.9031	0.9019	0.9032	0.9019

In all models, all predictors are grand mean centered except the three sequential dummy codes for education, in order to simplify interpretation of the interactions in models 5, 6, and 7. All hypothesis tests are based on robust standard error estimators from the unit-specific model.

* $p < .05$

** $p < .01$

*** $p < .001$