Sample Size for Newspaper Content Analysis in Multi-Year Studies

By Stephen Lacy, Daniel Riffe, Staci Stoddard, Hugh Martin, and Kuang-Kuo Chang

J&MC Q

This study examines the most efficient method of sampling content from five years of daily newspaper editions. Selecting nine constructed weeks (nine issues from a Monday, nine from a Tuesday, etc.) from five years is more efficient than the ten constructed weeks—two from each year—suggested by previous research on populations of a year's newspaper content. This rule holds provided the variables being measured do not have large variances.

During the half century since researchers began examining the methods of sampling used in selecting media content units for analysis,[1] a number of studies have appeared that explore the comparative efficiency of different types of probability sampling in describing a year's content of daily newspapers, weekly newspapers, television newscasts, and news magazines.[2] Most often, these have compared simple random samples of different sizes with stratified samples of different sizes.

This study continues in that vein, seeking the most efficient constructed week sample size for use across an *extended* time frame—such as a five-year period. Content analysts conducting studies that seek trends across time—"a decade, a presidential term, an editorship, a period of social unrest, the entire history of a magazine"[3]—must often wrestle with the question of how much data to collect to represent such time frames reliably and with optimal use of resources.

Background

In their study of changes in writing styles used by journalists and novelists during the 1885-1989 period, Danielson, Lasorsa, and Im randomly sampled thirty front-page sentences for each of 105 years for two major dailies. They examined ten sentences from each of three editions per paper per year.[4] Random selection of the sentences notwithstanding, the question arises as to whether examining each of the 105 years was necessary and whether three editions per paper per year were adequate. Could Danielson, Lasorsa, and Im have used a sampling system that selected every *nth* year? Should they have drawn more editions per year for their sample?

J&MC Quarterly
Vol. 78, No. 4
Winter 2001
836-845

Stephen Lacy is professor and director of the School of Journalism at Michigan State University; Daniel Riffe is professor in the E.W. Scripps School of Journalism at Ohio University; Staci Stoddard is a research attorney for the Michigan Court of Appeals; Hugh Martin is assistant professor in the Henry W. Grady College of Journalism and Mass Communication at the University of Georgia; and Kuang-Kuo Chang is a doctoral candidate in the Michigan State Mass Media Ph.D. program.

Clearly, the Danielson-Lasorsa-Im strategy of examining content for each of the 105 years yields a remarkable research design able to embrace an entire century. Content analysts have, however, developed alternative approaches in order to infer across long periods, without examining each and every year in the period. Gil Fowler's readability study of three major non-sensationalist dailies—the *Chicago Tribune, New York Times*, and Memphis *Commercial Appeal*—across a sixty-three-year period (1904 to 1966), used only three data points: 1904, 1933, and 1966.[5] Stevenson's earlier readability study of the *New York Times, New York Tribune*, and *Washington Post* encompassed an eighty-nine-year period (1872 to 1960), but also drew an "irregular" systematic sample (i.e., the skip interval was not consistent) and looked at issues from 1872, 1895, 1925, and 1960.[6]

While such approaches permit a general assessment of change across fairly broad time periods, they are not sensitive to fluctuations that may occur *between* data points. Of course, some researchers would counter that by drawing particularly large samples at each data point, they gain a richness not available in the "shallower" (three editions per year) but "broader" (each of 105 years) sampling design used, for example, by Danielson, Lasorsa, and Im.

At the heart of this issue is the recurring question of how many sampled edition dates are needed to adequately represent the population during a particular period in time—not just a single year, but a number of years. How many sample dates must be drawn to represent a presidential administration, an editorship or a five-year organizational planning cycle?

Why not just draw a simple random sample of all editions published during that administration or planning cycle? Indeed, why not a random sample of all issues during the 105 years studied by Danielson, Lasorsa, and Im; the sixty-three-year period studied by Fowler; or the eighty-nine-year expanse studied by Stevenson? One answer is that previous studies, beginning with Stempel's in 1952,[7] have shown that the cyclic nature of media content can render simple random sampling inefficient (i.e., more editions must be sampled), compared to other types of sampling. For example, daily papers vary from day to day during a week because of the advertising cycle, and simple random sampling can over-sample large-news hole Wednesday and Sunday editions and under-sample scanty Saturday editions.[8]

One solution to the problem of systematic content variation in daily newspapers is stratified sampling that yields constructed weeks.[9] Constructed week samples involve identifying all Mondays, and randomly selecting one Monday, then identifying all Tuesdays, and randomly selecting one Tuesday, etc., to "construct" a week that ensures that each source of cyclic variation—each day of the week—is represented equally. This stratified sampling presumably controls for sources of "systematic variation."[10]

In comparing different sampling methods, Riffe, Aust, and Lacy[11] found that one constructed week was adequate for representing a six-month "population" of editions for a daily newspaper, but that two constructed weeks were better. By contrast, they needed a simple random sample of twenty-right editions to adequately represent the six-month period. By extension, they reasoned, it would take a minimum of two constructed weeks to reliably represent an entire year's content, a finding consistent with Stempel's 1952 work.[12]

In subsequent examinations of the effect of media cycles on sampling, Lacy, Riffe, et al. explored whether stratified (constructed) sampling was more effective than simple random sampling with other news media. For weekly newspapers, they found "stratified sampling has some efficiency compared to random sampling, but the influence of cycles in content is not as strong in weeklies as in dailies." They said that randomly selecting twelve issues per year stratified by month is preferable to sampling of fourteen issues chosen totally at random.[13]

Maximum efficiency for sampling nightly network newscasts was achieved when two days stratified by month were selected; it took thirty-five days selected via simple random sampling to match the efficiency of the twenty-four-day-per-year sample.[14] With a population of one year of weekly magazine issues, researchers found that randomly selecting one issue stratified by month, or twelve issues per year total, was more effective than simple random selection of fourteen issues.[15]

In a study of sampling efficiency of monthly consumer magazines, Lacy, Riffe, and Randle concluded that a constructed year worked best for a five-year period, randomly selecting one issue from each of the months from the period.[16] Just as the newspaper stratified sampling method of constructed weeks takes advantage of systematic content variations by days of the week, consumer magazine stratified sampling takes advantage of systematic variations by season and month.

Research Question

As researchers using longitudinal research designs examine changes in media content over time, questions about the most efficient sampling procedures become increasingly important. When the time period under examination is greater than a single year, what kinds of adjustments—if any—need to be made to drawing two constructed weeks from each year? If generalizing to a five-year period of a daily newspaper's content is the goal, should the selected sample be ten constructed weeks, as suggested by Riffe, Aust, and Lacy (five years X two constructed weeks per year)? In short:

> What is the minimum number of randomly constructed weeks needed for accurate inference to a population of five years of a daily newspaper's editions?

Method

Testing sampling efficiency for inferring to a five-year population of newspaper editions required three steps: (1) calculating population parameters for the five-year population; (2) drawing random stratified samples in sets of 50 for various numbers of weeks and calculating how well they estimate the population parameters; and (3) determining which size (number of weeks) stratified samples were most effective. Because research has already established that stratifying samples by days of the week (constructed weeks) is more efficient than simple random samples,[17] this study only used constructed week sampling. Existing research indicates that two constructed weeks per year (for a total of ten across five years) would be efficient and representative. The question this research sought to answer is whether fewer than ten constructed weeks randomly drawn from all five years would be equally effective in estimating population parameters. Such an increase in efficiency would be important to researchers with limited resources.

Comparisons among sample sizes are based on the distribution of sample means predicted by the Central Limits Theorem: 68% of sample means (for a particular variable) should fall within one standard error of the population mean; 95% should fall within two standard errors. For example, for a given sample size (say, a simple random sample of n=16) and test variable, a researcher would draw 50 random samples of that size. If 34 (68%) and 48 (96%) of the 50 means fell within the ranges predicted by the theorem, the method was considered adequate for yielding a representative sample.

In previous research of this type, a decision rule was used: a sample size was efficient if both its percentages equaled or exceeded expected percentages, provided the next larger sample size did not drop below either expected percentage. For example, if 95% of 50 sample means for six constructed weeks were within plus or minus two standard errors of the population mean and 68% of these means were within one standard error of the population mean, the size was considered effective unless the percentage dropped under 95% for the 50 samples of seven constructed weeks. If it did drop, the effectiveness of the six-constructed week samples was considered an anomaly, and the acceptable level became the next sample size (perhaps eight weeks) that met the criteria.

This study used four variables to answer the research question: number of photographs, number of information graphics, number of stories, and number of stories by staff members. Photographs were defined as any illustration of an event or person produced with a camera. Information graphics were defined as any drawing or graphic that presents information, which included maps, tables, figures, and cartoons other than those on the comics pages. Stories were defined as any text item that has a headline regardless of length. Staff-written stories were any stories written by someone identified as a member of the staff by a byline. These four variables were selected because they had been used in previous content studies and because they were expected to show a degree of variation in their means and standard deviations that would provide a more demanding test than would using a single variable.

To create a five-year population, three students used microfilm to code the five years of the *Lansing State Journal* from 1990 to 1994. These years and this newspaper were selected because of availability at the beginning of the study. The five years had a total of 1,820 issues, five of which were missing from the microfilm. With 1,820 issues, it is highly improbable that these missing data would have changed the conclusions of this study. However, the variable population means were substituted for the missing issues data.

TABLE 1
Population Distribution for Five Years of Newspaper Editions (1,820 issues from 1990-1994)

	Population Mean	Population Standard Deviation	Coefficient of Variation
Number of Photographs	22.07	9.42	.427
Number of Graphics	15.40	9.22	.600
Number of Stories	62.35	10.60	.170
Number of Staff Stories	15.04	10.26	.682

The frequencies of each variable were counted for each issue, summed across the five years and divided by the number of editions (1,820) to determine the population means. Because this study involved the simple counting of easily identified units, no reliability check was required. In a study such as this, exact identification of the number of units is not essential as long as the totals are a good estimate of the population parameter. Small variations from the real population parameters will not alter the reliability of the conclusions about the representativeness of different sample drawn from the created population.

Using a random number generator, 50 samples of six-, seven-, eight-, and nine-constructed weeks each were drawn for a total of 200 samples (50 x four time periods). The sample means and standard errors were calculated for each of the four variables for all 200 samples (800 sample means and standard errors). Each sample was tested to see if the population mean fell within one and two sample standard errors of the sample mean. The percentage of samples for each set of 50 was then compared to the percentages predicted by the Central Limits Theorem to determine if the sampling method would produce a representative sample.

Results

The first section of the results section will discuss the nature of the population of newspaper content being studied. The second will address the research question concerning sampling efficiency for a five-year period.

The Population. The assumption of variability among tested variables was supported by Table 1 data showing the population means, standard deviation, and coefficient of variation (CV). The CV is the standard deviation divided by the mean and represents the variability of units in a population or sample. The higher the coefficient, the more variable the cases. The CVs for the four variables in Table 1 were .17 for number of stories, .43 for the number of photographs, .60 for number of graphics, and .68 for number of stories by staff members.

Riffe, Lacy, and Fico discuss the potential impact of content variability on sampling.[18] They warned about the impact large variances in

TABLE 2

Population Distribution for Each of Five Years of Newspaper Editions (1,820 Issues from 1990-1994)

	No. of Photos		No. of Graphics		No. of Stories		No. of Staff Stories	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1990	24.7	11.4	15.5	6.1	69.8	15.4	30.2	14.4
1991	21.7	10.2	24.7	6.7	65.1	6.9	10.7	2.4
1992	21.1	9.2	16.3	11.1	57.1	8.0	10.8	2.6
1993	21.7	7.6	11.5	8.4	59.9	7.3	11.1	2.7
1994	21.1	7.7	9.0	2.7	59.7	7.3	12.3	2.9

content can have on the representativeness of samples from that content. If the CV for a variable exceeds .5, they suggest that a researcher should increase the size of the sample.

Data in Table 2 show the means per issue for each variable in each of the five years. The mean number of photographs declined slightly after 1990, but remained almost equal during the last four years. The standard deviation declined during the first three years and was almost equal the last two. This stability suggests homogeneity across years and an anticipation that fewer than ten constructed weeks might be efficient.

The number of information graphics was more volatile with an increase from a mean of 15.5 in 1990 to a mean of 24.7 in 1991, and a dropping off each of the following three years to 9.0 in 1994. The standard deviation varied greatly as well. The mean number of stories per issue was more variable than the number of photographs, but not as erratic as the number of graphics. The overall trend from 1990 to 1994 was a decline in number and in standard deviation. The mean number of stories declined but became less variable.

The number of staff stories showed a different pattern: it was very high the first year, but it also varied tremendously. The mean dropped by two-thirds from 1990 to 1991, and the standard deviation dropped by 83%. An examination of the data suggests that the differences could represent a combination of the greater use of part-time reporters in 1990 and the staff reductions in 1991 from the advertising slump.[19]

The Samples. The research question asked: What is the minimum number of randomly constructed weeks needed for accurate inference to a population of five years of issues of a daily newspaper? As noted, previous research suggests that two constructed weeks from a newspaper will be representative of a year's issues of newspapers.[20] Thus, ten weeks—two from each of five years—should be representative of the five-year content population. The research question can be restated as: Does a sampling procedure that draws fewer than ten constructed weeks yield a valid sample for five years of daily newspapers?

TABLE 3

The Percentage of 50 Samples where the Samples Mean Falls within Plus or Minus One or Two Standard Error of the Population Mean for Five Years of Newspaper Editions

Constructed Weeks	# of Photos 1 SE	# of Photos 2 SE	# of Graphics 1 SE	# of Graphics 2 SE	# of Stories 1 SE	# of Stories 2 SE	#of Staff 1 SE	#of Staff 2 SE
Nine	74%	96%	72%	98%	78%	96%	74%	94%
Eight	68%	94%	72%	96%	72%	92%	74%	96%
Seven	74%	96%	76%	96%	72%	94%	72%	96%
Six	60%	98%	66%	94%	70%	92%	70%	94%

Note: The Central Limits Theorem predicts that 95% of random sample means will be within plus or minus two standard errors of the population mean, and that 68% will be within plus or minus one standard error of the population mean. The underlined percentages indicate the sampling percentages exceeded these critical values.

Data in Table 3 present the results for 50 samples for each of six, seven, eight, and nine constructed weeks for all four variables. Underscored percentages meet or exceed those predicted by the Central Limits Theorem. Six weeks is obviously an insufficient sample because four of the eight percentages are below the criteria (68% for one standard error and 95% for two).

Seven constructed weeks improved performance considerably because seven of the eight sets of 50 samples exceeded the expected proportions. Interestingly, the one set of samples that failed to meet the predicted proportions was the variable with the least amount of variation—number of stories.

Eight constructed weeks proved to be even less efficient than seven weeks because two of eight sets of samples failed to meet the expected proportion. Only 94% of the 50 samples for number of photos in the newspaper were within two standard errors and number of staff stories fell within two standard errors only 92% of the time.

Nine constructed weeks from five years of daily newspaper issues almost meets the required performance level, except that the 50 samples for number of staff stories failed to meet the 95% cut-off point. Only 94% of the samples had a sample mean that fell within two standard errors of the population mean. In effect, one sample more than expected fell outside the expected two-standard error cut-off point. This most likely reflects the relatively high variability of this variable. With high standard deviation in the population, there is a greater chance that the mean of a particular sample will lie fairly far from the population mean. Looking at the sample in this set, three of the samples missed having the sample mean within two standard errors of the population mean by less than .05.

Conclusion

Whether or not a researcher should use a nine-week constructed sample for a five-year period depends on the nature of the newspaper content analysis project, the variability in the sample for the selected variables and the resources available. This study indicates that there is a high probability that a nine-constructed-week sample will provide a valid representation of the content in a daily newspaper during that period. Even though one variable failed, this probably reflects the use of the limited 50-sample sets to test the Central Limits Theorem when the variable shows high variation. The Central Limits Theorem is based on an assumption of a large number of sample means.

This recommendation needs a proviso, however. If the sample shows high variance for certain variables, then ten constructed weeks, two from each year, would be a more conservative approach. Riffe, Lacy, and Fico[21] defined high variability as a CV greater than .5. A researcher would start by selecting the nine constructed weeks during the five-year period and then calculate the CV. If the CV exceeded .5, the researcher would then create a tenth constructed week to increase the probability that the sample was representative of the five years.

This study suggests three approaches to a long-term study of newspaper content. Assuming a 100-year period and variables with a CV under .5, the first approach would be to sample two constructed weeks from each year, for a total of 200 weeks or 1,400 issues. This would be the most likely way to provide a valid description of the content over 100 years.

A second approach would be to select two constructed weeks from each year but to avoid sampling each year. This raises an issue of how to select the years within the 100-year period. One way would be to select randomly, but this could likely lead to an uneven distribution of years that would not allow the researcher to represent trends and could easily be misleading about the nature of content. A better solution would be to pick a constant interval, say every two years. This would allow an analysis of trends and it would be efficient (two weeks from 50 years instead of 100 years), but there is a chance of missing unusual years, such as this sample's 1990. Unfortunately, there is no published evidence about what type of interval would work best.

The third approach is one suggested here, with nine constructed weeks taken from each five-year period. This procedure is likely to include variations across years because all years are included during the period. Over 100 years, this procedure would yield a total of 180 constructed weeks (or 1,260 issues) instead of the 200 constructed weeks (or 1,400 issues) when two weeks were selected for each year.

This research also indicates that a researcher studying a time period shorter than five years (for example, four years in a presidential term) should use two constructed weeks from each year in order to generate an adequate representation of newspaper coverage during that period. Questions raised here about the limits of fewer than two constructed weeks a year with highly variable content suggest this approach.

In summary, selecting nine constructed weeks from five years is more efficient than the ten—two from each year—suggested by previous research on populations of a year when the variables being measured have small variations. However, if variations are large, ten constructed weeks should be used.

NOTES

1. A. Mintz, "The Feasibility of the Use of Samples in Content Analysis," in *Language of Politics*, ed. H.D. Lasswell, N. Leites, R. Fadner, J.M. Goldsen, A. Gray, I. L. Janis, A. Kaplan, D. Kaplan, A. Mints, I. de sola Pool, and S. Yakobson (New York: George W. Stewart, 1949), 127-52; Guido H. Stempel III, "Sample Size for Classifying Subject Matter in Dailies," *Journalism Quarterly* 29 (summer 1952): 333-34; J. Davis and L. W. Turner, "Sample Efficiency in Quantitative Newspaper Content Analysis," *Public Opinion Quarterly* 15 (winter 1951): 762-63; Roy L. Jones and Roy E. Carter Jr., "Some Procedures for Estimating 'News Hole' in Content Analysis," *Public Opinion Quarterly* 23 (fall 1959): 399-403.

2. Daniel Riffe, Charles F. Aust, and Stephen R. Lacy, "The Effectiveness of Random, Consecutive Day and Constructed Week Samples in Newspaper Content Analysis," *Journalism Quarterly* 70 (spring 1993): 133-39; Stephen Lacy, Kay Robinson, and Daniel Riffe, "Sample Size in Content Analysis of Weekly Newspapers," *Journalism & Mass Communication Quarterly* 72 (summer 1995): 336-45 ; Daniel Riffe, Stephen Lacy, Jason Nagovan, and Larry Burkum, "The Effectiveness of Simple Random and Stratified Random Sampling in Broadcast News Content Analysis," *Journalism & Mass Communication Quarterly* 73 (spring 1996): 159-68; and Daniel Riffe, Stephen Lacy, and Michael Drager, "Sample Size in Content Analysis of Weekly News Magazines," *Journalism & Mass Communication Quarterly* 73 (autumn 1996): 635-44.

3. Riffe, Lacy, and Drager, "Sample Size in Content Analysis of Weekly News Magazines."

4. Wayne A. Danielson, Dominic L. Lasorsa, and Dai S. Im, "Journalists and Novelists: A Study of Diverging Styles," *Journalism Quarterly* 69 (summer 1992): 436-46.

5. Gilbert L. Fowler Jr., "The Comparative Readability of Newspapers and Novels," *Journalism Quarterly* 55 (autumn 1978): 589-92.

6. Robert L. Stevenson, "Readability of Conservative and Sensational Papers Since 1872," *Journalism Quarterly* 41 (spring 1964): 201-206.

7. Stempel, "Sample Size for Classifying Subject Matter in Dailies."

8. Riffe, Aust, and Lacy, "The Effectiveness of Random, Consecutive Day and Constructed Week Samples."

9. Jones and Carter, "Some Procedures for Estimating 'News Hole.'"

10. Riffe, Lacy, and Drager, "Sample Size in Content Analysis of Weekly News Magazines."

11. Riffe, Aust, and Lacy, "The Effectiveness of Random, Consecutive Day and Constructed Week Samples."

12. Stempel, "Sample Size for Classifying Subject Matter in Dailies."

13. Riffe, Aust, and Lacy, "The Effectiveness of Random, Consecutive

Day and Constructed Week Samples."

14. Riffe, Lacy, Nagovan, and Burkum, "The Effectiveness of Simple Random and Stratified Random Sampling in Broadcast News Content Analysis."

15. Riffe, Lacy, and Drager, "Sample Size in Content Analysis of Weekly News Magazines."

16. Stephen Lacy, Daniel Riffe, and Quint Randle, "Sample Size in Multi-Year Content Analyses of Monthly Consumer Magazines," *Journalism & Mass Communication Quarterly* 75 (summer 1998): 408-417.

17. Stempel, "Sample Size for Classifying Subject Matter in Dailies," and Riffe, Aust, and Lacy, "The Effectiveness of Random, Consecutive Day and Constructed Week Samples."

18. Daniel Riffe, Stephen Lacy, and Frederick G. Fico, *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (Mahwah, NJ: Lawrence Erlbaum Associates, 1998), 100.

19. The variation in this five-year sample is probably not unusual for any five-year period. During such a period, newspaper budgets can be affected by a variety of variables from outside the newspaper—reader demand, advertising demand, and even general economic variations. The variability among variables in this study reflects the reality of newspaper resource allocation.

20. Stempel, "Sample Size for Classifying Subject Matter in Dailies," and Riffe, Aust, and Lacy, "The Effectiveness of Random, Consecutive Day and Constructed Week Samples."

21. Riffe, Lacy, and Fico, *Analyzing Media Messages*, 100.