

**Procedures for Assembling the Fourth Data Set**  
*Congress, the Press, and Political Accountability*  
R. Douglas Arnold

The fourth data set is designed to determine whether the volume of newspaper coverage affected what citizens knew about their representatives. This data set was constructed by linking information about how extensively the 67 newspapers in the third data set covered particular representatives with information about citizens' knowledge of their local representatives, as recorded in the autumn 1994 survey conducted by the National Election Studies. The unit of analysis is the individual citizen. Added to the usual attitudinal data about each citizen is information about how a local newspaper covered that citizen's representative during 1993 and 1994. The original 1994 NES data set had 1,795 respondents. I have information about local newspaper coverage for 675 of these respondents.

The data set is far from ideal. The foundation for the analysis is the autumn 1994 survey conducted by the National Election Studies. This biennial survey of American citizens is the gold standard for information about what citizens know about candidates for the House, Senate, and presidency. The problem is finding an appropriate way to link the NES data set with my own data on newspaper coverage. Ideally, one would like to know which newspaper each respondent read and then link information about how that newspaper covered a representative with information about the respondent's opinions. Unfortunately, although the NES survey asked how many times a week a respondent read a newspaper, it did not ask the name of the newspaper. So, I have been forced to assume that the local newspaper for which I have data is the same paper that a citizen read. It is not a crazy assumption. In many districts, there was a single dominant paper that most newspaper readers read. Errors in matching are more common in cities with competing newspapers and in far-flung districts that encompass several media markets. If two newspapers served the same media market, I assumed that a respondent read the larger one.

The inevitable errors in linking my data with NES data introduce noise into the merged data set. Although the noise will surely mask weak signals in the data, it should not overpower strong signals. One reason is that competing same-city newspapers do not cover local representatives that differently. In any event, the noise works against confirming hypotheses about the impact of newspaper coverage on what citizens know about representatives.

The fourth data set contains 675 respondents in 100 congressional districts. Nine representatives in these districts did not run for reelection and nine representatives ran unopposed, so there were 559 respondents in 82 congressional districts who faced an incumbent running against a challenger. The fourth data set contains all information that NES collected about each respondent and all information from the third data set about how the dominant newspaper in the respondent's district covered the local representative. The newspaper information is about the volume of coverage in 33 newspapers, not about the content of the coverage. We know how many articles per month mentioned a local representative; we know nothing about what messages those articles conveyed.