Correlational linkage analysis (Frequently Applied Designs)

AUTHOR
Laia Castro Herrero, Theresa Gessler, Silvia Majoo-Vazquez

KEYWORDS
correlational linkage analysis, second order linkage analysis, time series analyses, aggregated data, content analysis, content data, survey data, mixed method, media effects, public opinion climate, media context

BRIEF DESCRIPTION
Correlational or second-order linkage analyses (Schulz, 2008) correlate content data points and survey data at the aggregate level. They are generally used to infer the impact of public opinion climate, the media context or media use on individual attitudes, cognitions and behaviors. Correlational linkage analyses make use of data collected at different points in time to be able to describe patterns of change and stability over time and to compensate for the reduced number of observations resulting from aggregating individual-level data. They often employ manual and automated content analysis, descriptive and inferential statistical analyses, and time series analysis.

FIELD OF APPLICATION/THEORETICAL FOUNDATION
Linkage analyses have extensively been used in the fields of political communication (Soroka, 2002), EU studies (Brosius et al., 2019a), and more recently, social media and social movements. Studies that employed second-order linkage analyses are related to theories of agenda setting (McCombs & Shaw, 1972), framing (Vliegenthart et al., 2008), or media bias and tone (Brosius et al., 2019b) (see chapter Content Analysis in Mixed Method approaches for a detailed account of applications and advantages of using linkage analyses).

EXAMPLE STUDIES
In this data entry we describe two studies that regress survey data on media content data with additional weights to better model news media effects. The first study (Boomgaarden & Vliegenthart, 2007) weighs media coverage of a particular topic (immigration) by issue prominence and circulation of the newspapers considered in the study. The second one (Vliegenthart et al., 2008) further introduces a publication recency moderator to account for how close in time a given news story was published from when survey data was collected and individuals may have been exposed to such piece of information.

REFERENCES
of Public Opinion Research, 348–357.

**Table 1. Data matching in correlation linkage analyses**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Relationship of theoretical interest</th>
<th>Sample</th>
<th>Time frame</th>
<th>Content-analytical constructs</th>
<th>Linkage strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boomgaard &amp; Vliegenthart (2007)</td>
<td>News media reporting about immigration-related topics on aggregate share of vote intention for anti-immigrant parties</td>
<td>(a) 157,968 articles collected through computer-assisted analysis, dealing with immigration and published in the five most-read Dutch national newspapers</td>
<td>1990-2002</td>
<td>Visibility of immigration-related topics in news</td>
<td>(1) The authors calculate a visibility score per article by computing:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.1.) an average person’s log probability that s/he is exposed to news about immigration through a given article. This is done by using the frequency with which this article mentions immigration-related topics ( f(t,a) ), both in the headline ( f_h(t,a) ), in which case the frequency is weighed by 8, and in the body of the text ( f_b(t,a) ), in which case the frequency is multiplied by 2.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.2.) 1.1. is weighed by circulation of the newspaper where the article is published ( c(a) ).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.3.) 1.1. is weighed by whether the article is placed in the front page or other to account for how prominently the topic is featured ( f_p(a) ).</td>
</tr>
</tbody>
</table>
Notationally, the equation can be written as follows:

\[ V(t, a) = \sum_{i} \log(j_i(t, a) + f_i(t, a)) \times c(a) \times fp(a) \]

(2) In a second step, V(a) are aggregated for all articles in all outlets by month (the time unit to link content and survey data).

(3) Final immigration visibility scores (independent variable) are linked to monthly percentage of people that reported intending to vote for an anti-immigration party (dependent variable) through time series analysis. The authors run ARIMA models, successively adding controls for extreme right leadership peaks (Fortuyn’s entrance in the political arena and assassination), immigration levels, unemployment rates, the interaction between the both and finally, the media visibility variables.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Relationship of theoretical interest</th>
<th>Sample</th>
<th>Time frame</th>
<th>Content-analytical constructs</th>
<th>Linkage strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>period t in each country c.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) 9,649 hand-coded articles that mentioned the EU at least twice (at least one of these references in the headline or in the lead of the article) were then analysed to investigate the framing of the EU. Approximately 50 articles per country were coded for each 6-month period.

(c) Self-reports on EU support from the bi-annual standard Eurobarometer.

(2) Framing scores are then assigned to each article (benefit and disadvantage frames 0-2, conflict framing ranged from 0 to 3)

(3) Mean framing scores per time period–country combination (fs(t,c)) are multiplied by visibility scores (vs(t,c)) to capture the overall salience of the frames (beyond its presence) as follows:

\[ S(t, a) = fs(t, c) \times vs(t, c) \]

(4) OLS regressions with panel corrected standard errors are run with benefit, disadvantage and conflict framing as main independent variables, and aggregated-level support for the EU as dependent variable.