

Data Quality for Measuring Political Protest and Government Change

Charles Lewis Taylor
Virginia Tech University

Abstract

For analyzing protest and government change, data need to be both reliably measured and valid for the task at hand. Data fundamentalists insist that nothing less than total capture of every event that has taken place is sufficient for analysis. That scenario is, of course, an impossibility. Rather, methods to measure the statistical probability that a sample is appropriate for its targeted universe allow empirical analysis to proceed.

Keywords: Reliability, validity, event indicators, political protest, government change

1. Introduction

Analyzing government change and political protest can contribute to understanding causes of conflict, which in turn may lead to its mitigation. For this analysis to be effective, however, reliable empirical data for testing theory are as important as the theoretical frameworks they are to test. Otherwise, we obtain only “grand theories firmly planted in mid-air,” as Karl W. Deutsch used to say. Assuring the quality of data, therefore, is itself a serious component of scholarship that raises the argument above the level of ideology. Measuring political phenomena is fraught with dangers derived from imprecise or false concepts as well as from clerical error. Assumptions and procedures in data coding affect analytical results and are therefore inherently theoretical in nature. Clear-headed thought in developing political measurements is thus required if progress is to be made in political understanding.

Data are judged by their validity and their reliability. Validity refers to cogency or soundness in reason. Are the procedures and rules for coding able to withstand counterarguments? Are they convincing? Do they work for the theoretical purpose at hand? Reliability refers to trustworthiness and accuracy. Are the data dependable? Can they be used with confidence? Are they sufficient for the questions being asked of them? Validity is one among several exercises in theory building. It requires speculative and theoretic contemplation. Reliability, on the other hand, is more a matter of meticulously disciplined work. It involves simply getting it right. If a dataset is to be effective for analysis, it must be constructed both with carefully designed concepts and closely controlled accounting.

2. How Good Do the Data Need to Be?

This question is meaningful only in terms of how the data are to be used. Are they suitable for the purpose to which they will be put? Considerably more detail will be needed for a fine-grained study than for an all-embracing comparison. Examining the exact sequence of

Charles Lewis Taylor, Professor, Department of Political Science. Virginia Polytechnic Institute and State University. Email: clt@vt.edu

conflict events during an eight-day riot will require far more complete data than investigating the differing levels of aggregate violence in a number of societies over a century. Data fundamentalists find this concept difficult to accept. They assume that sufficient diligence will produce absolute accuracy and believe that nothing less is acceptable for any analysis. Not only is there a truth out there, but it is a truth that can be determined with exactitude. Inevitably, personal perceptions of truth become the standard by which results are judged. This situation leads either to endless argument or to scholarly nihilism. The alternative that avoids this dilemma is to analyze human behavior in probabilistic terms. Working with partial knowledge and estimating the probable likelihood of error is usually the best we can expect. The development of the statistical approach to scientific analysis was based upon this understanding.

Charles Brockett missed this vital point in his evaluation of event data reported in the *World Handbook of Political and Social Indicators: Third Edition*. He claimed to discover “gross and systematic errors” in the data because he found reports of some events that were not included. He asserted that analyses employing this source were flawed due to their use of “meaningless data” to measure mass political violence. Although he made these charges on the basis of very short-term observations (over scattered years in the 1970s) of one data series (political deaths) for a very small number of cases (five Central American countries), he reified his figures into a benchmark of real truth against which to measure the entire 30 years, 139 countries, and 18 event series of the *World Handbook*. “A prodigious undertaking, providing annual events data for 1948-77, it is not good enough,” he concluded.¹ Good enough for *what* he did not address.

In addition to ignoring the fact that such a tiny, non-random sample is unrepresentative of the whole population of events, Brockett seems to have misunderstood the general nature of statistical analysis. In their devastating reply, Dixon, Muller, and Seligson made the point that

documented evidence of underreporting does not necessarily imply a precise accounting. Brockett’s investigation of deaths in Central America is a case in point. While he does supply convincing evidence of underreporting of deaths, he fails to provide enough detailed information to permit an accurate correction.²

Dixon et al. used a censoring technique that identified moderate rather than the excessive levels of violence claimed by Brockett.³ Because underreported cases could be reasonably expected to rank above the threshold that they determined, they set the score for these cases equal to that threshold. This technique allowed observing how the underreported cases affected their analysis. When they compared the effects of the adjusted data to those of the unadjusted data in a model predicting violence from land inequality, the estimates changed slightly and the overall fit declined marginally, but the general pattern remained pretty much the same. Making allowances for error and employing the best data available did not overturn the results. Dixon, Muller, and Seligson noted that “blanket charges of inaccuracy without supporting evidence can do little to advance inquiry” and that evidence of significant

¹ Charles D. Brockett, “Measuring Political Violence and Land Inequality in Central America,” *American Political Science Review* 86 (1992): 169; Charles Lewis Taylor and David A. Jodice, *World Handbook of Political and Social Indicators: Third Edition* (New Haven: Yale University Press, 1983).

² William J. Dixon, Edward N. Muller, and Mitchell A. Seligson, “Inequality and Political Violence Revisited: Response,” *American Political Science Review* 87 (1993): 990.

³ It would have been possible, they say, to eliminate cases plagued by documented irregularities, but this would have produced serious distortions since irregularities are located precisely in the times and places where disorder is known to be frequent. The optimal solution, of course, would have been to correct inaccuracies with better data, but that is easier said than done.

underreporting “is better used as a remedy than an indictment.”⁴

There are no methods by which to obtain a metric of everything that has happened. Historians have of course known this for centuries. There are no perfect scholars, and even if there were, no complete sources for political events exist. No scribe could ever record them all. Events happen more quickly and in more places than can be accommodated by available clerks. Hence, no dataset can be assured of containing every single event of conflict within its scope, whether it be a quantitative assessment for a couple of years of a country’s history taken from police blotters and local newspapers or a carefully compiled descriptive report drawn from several separate historical archives.

3. So What Do We Do?

A dataset needs to be characterized not so much by completeness as by comparability across time and space. Wooley points out that any dataset of political events is in reality a sample drawn from the universe of actual events that occurred, and that no absolute benchmark exists by which the quality of any particular sample can be judged.⁵ If we follow the logic of sampling theory, of course the more densely packed the dataset, the greater its likelihood of representing the actual patterns that exist in the total event population. Nevertheless, there is always the overwhelming probability that more events occurred than were reported, however large the sample.

The first requirement for probabilistic analysis is the reasonable assurance that the data sample is unbiased. A sample may or may not be random with regard to the various criteria that are important to the investigation at hand. Moreover, it will be drawn from an unknown population. A careful examination is required of the sources as well as the protocol by which the data are to be coded or otherwise manipulated. Analytical interest, of course, is in the patterns of events, the relationships among these patterns, and the connections between them and other relevant variables. Are the data a sufficient representation of the underlying population of events for testing probabilistic models that explain these patterns and relationships? The answer depends upon the requirements of the particular model. Dixon, Muller, and Seligson found no difference in relationship when they modified their data, but not all adjustments and retesting would allow statistical results to stand. It is the responsibility of the analyst to assess empirically the effects that data quality will have upon the analysis.

The questions of validity and reliability are of course related to the quality of the sources. How comprehensive is the source? What biases does it contain? Are events from reports recognized (whether by human coders or by automated coding) in terms relevant to theoretical concerns? Answering the last question requires developing base codes consistent with the theoretical framework: What needs to be measured and what is the logic for doing so? Finally, for large datasets that are most efficiently collected by automatically coding digital data, reliability has to do with the convergence of machine and human coders in assigning events to the proper event type and in identifying the associated attributes of these events.

Many of the early cross-national political event datasets were based upon print media. Crews of individual coders worked together to create and refine coding rules that could be

⁴ Dixon, Muller, and Seligson, “Inequality and Political Violence,” 990.

⁵ John T. Wooley, “Using Media-Based Data in Studies of Politics,” *American Journal of Political Science* 44 (2000): 156-173.

applied to journalists' reports of activities related to political protest and government change. The participants then checked inter-coder reliability by means of tests on sample material. Such a project was, of course, labor intensive and made even more so by the use of multiple sources, including sources in local languages. Adding to the required work, several scholars have sought local and international news sources. These, they argue, provide additional perspectives on political processes.⁶ Multiple sources increase the problem of dealing with duplicate reports of the same event, however. Simply adding event reports wherever found does not alone solve the comparability problem, but sources containing more event reports do offer the opportunity of more comparable reporting. Althaus, Edy, and Phalen made the case for using wire services rather than newspapers. Newspapers, they said, have a "news hole" of a specific size and the events reported have to fit into this space. Wire services have no news hole. Their continuous coverage can more easily expand or contract to cover what is actually happening. Wire services therefore include far more news than newspapers can afford to print.⁷

Reuters News Agency was singled out as a particularly good source that covers much that is missed by other news agencies.⁸ Althaus et al. argued that the multiple sources available through Reuters Textline Service in Nexis-Lexis do reduce bias and generate more density. In a test, they found Reuters Textline dwarfed several putative benchmarks in Northern Ireland, Poland, and the United States in the number of protest events that it covered. They found that the Radio Free Europe chronology included only 31.2 per cent of the events found in the Reuters' sources; Uncensored Poland News (Solidarity Newsletter) covered 3.7 per cent; the Bew and Gillespie Northern Ireland chronology covered 10.9 per cent; and the New York Times covered 11.4 per cent.⁹ Almeida and Lichbach found that internet sources provided much more complete coverage than the New York Times, Global Newbank, and the Seattle Times.¹⁰

Rucht and Ohlemacher rejected wire services on the basis of their then-inaccessibility and of the flood of information that would have to be screened, but over time, various schemes have been developed for machine-aided coding and eventually for direct computer coding.¹¹ Significant among these are systems developed by Philip Schrodtt and by Virtual Research Associates (VRA).¹² Eventually, agreement between human and machine coding has begun to approach 90 per cent, which is at a level about equal to the best results among trained human coders.

A major problem for human coders is the "black box" of the coder's mind. It is difficult

⁶ Henrik Sommer and James R. Scarritt, "The Utility of Reuters for Events Analysis in Area Studies: The Case of Zambia-Zimbabwe Interactions, 1982-1993," *International Interactions* 25 (1999): 1-31; Christian Davenport and Patrick Ball, "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995," *The Journal of Conflict Resolution* 46 (2002): 427-450.

⁷ Scott L. Althaus, Jilly A. Edy, and Patricia F. Phalen, "Using Substitutes for Full-Text News Stories in Content Analysis: Which Text is Best?," *American Journal of Political Science* 43 (2001): 707-724.

⁸ Mark Irving Lichbach, "The International News About Governability: A Comparison of the *New York Times* and Six News Wires," *International Interactions* 10 (1984): 311-340; Phillip A. Huxtable and Jon C. Pevehouse, "Potential Validity Problems in Event Data Collection," *International Studies Notes* 21 (1996): 8-19; Sommer and Scarritt, "The Utility of Reuters."

⁹ Althaus et al., "Which Text is Best?"

¹⁰ Paul D. Almeida and Mark Irving Lichbach, "To the Internet, From the Internet: Sources of Data About Antiglobalization Protest," *Mobilization* 8 (2003): 249-272.

¹¹ Dieter Rucht and Thomas Ohlemacher, "Protest Event Data: Collection Uses and Perspectives," in *Issues in Contemporary Social Movement Research*, eds. Ron Eyerman and Dario Diani (Beverly Hills: Sage, 1992), 76-106.

¹² Philip A. Schrodtt, "CAMEO: Conflict and Mediation Event Observations: Event and Actor Codebook," <http://eventdata.psu.edu/data.dir/cameo.html>; Virtual Research Associates, Integrated Data Events Analysis (IDEA), <http://vranet.com/data.aspx>.

to account for exactly how and why any single individual may respond to an event report. Coding manuals, coder training, and inter-coder reliability checks reduce the problem of random reactions, but coding errors do occur and are difficult to identify. In contrast, automated coding is highly consistent and thereby transparent in terms of error patterns. It is also flexible in that once errors are identified, codes can be altered and data rerun, creating a new dataset. In effect, automation replaces numerous moderately and low-skilled human coders with a few highly skilled programmers and protocol developers.

The Virtual Research Associates Reader is one project involved in the automated coding approach that makes use of Reuters Business Briefs. Participants include Doug Bond, Joe Bond, and Charles Taylor. Current examination of the validity of these data as dynamic indicators of intrastate conflict processes is encouraging. To be sure, media selection and misrepresentation present difficulties for conflict analysis, and although the source of the data is imperfect, it does provide a valid trace of conflict events in the aggregate. The question is not whether international news sources literally describe all events that occur but whether indices drawn from them represent cross-national patterns. The VRA project presents the results of a random sampling approach to assessing coding reliability for 50 conflict event types anywhere in the world.

Given the limitation of resources in the real world, there is always a trade-off between intricate measurements of narrower scope that usually produce a larger number of event reports for any specific time and place, and broader studies that allow greater comparative analysis. The broader the scope in time or space, of course, the greater the possibility for overlooking events of interest because the resources that can be devoted to each detail are relatively fewer. New developments, however, continue to mitigate even this problem.

Further Reading

- Barker-Plummer, B. "News as a Political Resource." *Critical Studies in Mass Communication* 12 (1995): 306-324.
- Barranco, J., and D. Wisler. "Validity and Systematicity of Newspaper Data in Event Analysis." *European Sociological Review* 40 (1999): 170-192.
- Beissinger, Mark R. *Nationalist Mobilization and the Collapse of the Soviet State*. London: Cambridge University Press, 2002.
- Boyd-Barrett, Oliver. *The International News Agencies*. Beverly Hills: Sage, 1980.
- Boyd-Barrett, Oliver, and Daya Kishan Thussu. *Contra-Flow in Global News: International and Regional News Exchange Mechanisms*. London: John Libbey, 1992.
- Boyd-Barrett, Oliver, and Terhi Rantanen, eds. *The Globalization of News*. Thousand Oaks, CA: Sage, 1999.
- Danzger, M. Herbert. "Validating Conflict Data." *American Sociological Review* 40 (1975): 570-584.
- Eisinger, P. "The Conditions of Protest in American Cities." *American Political Science Review* 67 (1973): 11-23.
- Franzosi, R. "The Press as a Source of Socio-Historical Data." *Historical Methods* 20 (1987): 5-15.
- Gans, H. *Deciding What is News*. New York: Pantheon, 1979.
- Hocke, Peter. "Determining the Selection Bias in Local and National Newspaper Reports on Protest Events." In *Acts of Dissent*, edited by Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt, 131-163. Berlin: Edition Sigma Rainer Bohn Verlag, 1998.
- Hug, Simon. "Selection Bias in Comparative Research: The Case of Incomplete Data Sets." *Political Analysis* 11 (2003): 255-274.
- Hug, Simon, and Dominique Wisler. "Correcting for Selection Bias in Social Movement Research." *Mobilization* 3 (1998): 141-161.
- Kielbowicz, R., and C. Scherer. "The Role of the Press in the Dynamics of Social Movements." *Research in Social*

- Movements, Conflicts and Change* 9 (1986): 71-96.
- King, Gary, and Will Lowe. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization* 57 (2003): 617-642.
- Koopmans, Ruud. "The Use of Protest Event Data in Comparative Research: Cross-National Comparability, Sampling Methods and Robustness." In *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans and Friedhelm Neidhardt. Berlin: Rainer Bohn Verlag, 1998.
- Koopmans, Ruud, and Dieter Rucht. "Protest Event Analysis – Where to Now?" *Mobilization* 4 (1999): 123-130.
- Leetaru, Kalev, and Philip A. Schrodt. "GDELT: Global Data on Events, Location and Tone, 1979-2012." Paper presented at the International Studies Association Meeting, San Francisco, April 2013. <http://eventdata.psu.edu/data.dir/GDELT.html>
- McCarthy, John D., Clark McPhail, and Jackie Smith. "Images of Protest." *American Sociological Review* 61 (1996): 478-499.
- McFarlane, Thomas, and Iain Hay. "The Battle for Seattle: Protest and Popular Geopolitics in The Australian Newspaper." *Political Geography* 23 (2003): 211-232.
- Mueller, Carol McClurg. "International Press Coverage of East German Protest Events, 1989." *American Sociological Review* 62 (1997): 820-832.
- Oliver, Pamela E., and Daniel J. Myers. "How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events." *American Journal of Sociology* 105 (1999): 38-87.
- Oliver, Pamela E., and Gregory M. Maney. "Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions." *American Journal of Sociology* 106 (2000): 463-505.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (2010): 651-660.
- Rucht, Dieter, and Friedhelm Neidhardt. "Methodological Issues in Collecting Protest Event Data: Units of Analysis, Sources and Sampling, Coding Problems." In *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans and Friedhelm Neidhardt. Berlin: Rainer Bohn Verlag, 1998.
- Shapiro, Gilbert, and John Markoff. *Revolutionary Demands: A Content Analysis of the Cahiers de Doléances of 1789*. Stanford: Stanford University Press, 1998.
- Snyder, David, and William R. Kelly. "Conflict Intensity, Media Sensitivity and the Validity of Newspaper Data." *American Sociological Review* 42 (1977): 1105-1117.
- Swank, Eric. "In Newspapers We Trust: Assessing the Credibility of News Sources That Cover Protest Campaigns." *Research in Social Movements, Conflicts and Change* 22 (2000): 27-52.
- Tarrow, Sidney. *Struggle, Politics and Reform: Collective Action, Social Movements, and Cycles of Protest*. Ithaca: Cornell University Press, 1989.
- White, R. "On Measuring Political Violence in Northern Ireland." *American Sociological Review* 58 (1993): 75-85.

Bibliography

- Almeida, Paul D. and Mark Irving Lichbach. "To the Internet, From the Internet: Sources of Data About Antiglobalization Protest." *Mobilization* 8 (2003): 249-272.
- Althaus, Scott L., Jilly A. Edy, and Patricia F. Phalen. "Using Substitutes for Full-Text News Stories in Content Analysis: Which Text is Best?" *American Journal of Political Science* 43 (2001): 707-724.
- Brockett, Charles D. "Measuring Political Violence and Land Inequality in Central America." *American Political Science Review* 86 (1992): 169-176.
- Davenport, Christian, and Patrick Ball. "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995." *The Journal of Conflict Resolution* 46 (2002): 427-450.
- Dixon, William J., Edward N. Muller, and Mitchell A. Seligson. "Inequality and Political Violence Revisited: Response." *American Political Science Review* 87 (1993): 983-993.

-
- Huxtable, Phillip A., and Jon C. Pevehouse. "Potential Validity Problems in Event Data Collection." *International Studies Notes* 21 (1996): 8-19.
- Jenkins, J. Craig, Charles Lewis Taylor, Marianne Abbott, Thomas V. Maher, and Lindsey Peterson. "Global Conflict Data: The World Handbook of Political and Social Indicators IV." Mershon Center for International Security Studies. www.sociology.osu.edu/worldhandbook
- Lichbach, Mark Irving. "The International News About Governability: A Comparison of the *New York Times* and Six News Wires." *International Interactions* 10 (1984): 311-340.
- Rucht, Dieter, and Thomas Ohlemacher. "Protest Event Data: Collection Uses and Perspectives." In *Issues in Contemporary Social Movement Research*, edited by Ron Eyerman and Dario Diani, 76-106. Beverly Hills: Sage, 1992.
- Schrodt, Philip A. "CAMEO: Conflict and Mediation Event Observations: Event and Actor Codebook." <http://eventdata.psu.edu/data.dir/cameo.html>
- Sommer, Henrik, and James R. Scarritt. "The Utility of Reuters for Events Analysis in Area Studies: The Case of Zambia-Zimbabwe Interactions, 1982-1993." *International Interactions* 25 (1999): 1-31.
- Taylor, Charles Lewis, and David A. Jodice. *World Handbook of Political and Social Indicators: Third Edition*. New Haven: Yale University Press, 1983.
- Virtual Research Associates. Integrated Data Events Analysis (IDEA), <http://vranet.com/data.aspx>.
- Wooley, John T. "Using Media-Based Data in Studies of Politics." *American Journal of Political Science* 44 (2000): 156-173.