Advance Access publication April 2, 2007

Political Analysis (2007) 15:365–386
doi:10.1093/pan/mpm007

Improving Data Quality: Actors, Incentives, and Capabilities

Yoshiko M. Herrera
Department of Government, Harvard University,
Davis Center, #S301, 1730 Cambridge Street, Cambridge, MA 02138
e-mail: herrera@fas.harvard.edu (corresponding author)

Devesh Kapur
Centre for Advanced Study of India, University of Pennsylvania,
3600 Market Street, Suite 560, Philadelphia, PA 19104
e-mail: dkapur@sas.upenn.edu

This paper examines the construction and use of data sets in political science. We focus on three interrelated questions: How might we assess data quality? What factors shape data quality? and How can these factors be addressed to improve data quality? We first outline some problems with existing data set quality, including issues of validity, coverage, and accuracy, and we discuss some ways of identifying problems as well as some consequences of data quality problems. The core of the paper addresses the second question by analyzing the incentives and capabilities facing four key actors in a data supply chain: respondents, data collection agencies (including state bureaucracies and private organizations), international organizations, and finally, academic scholars. We conclude by making some suggestions for improving the use and construction of data sets.

It is a capital mistake, Watson, to theorise before you have all the evidence. It biases the judgment.
—Sherlock Holmes in "A Study in Scarlet"

Statistics make officials, and officials make statistics."
—Chinese proverb

1 Introduction

Modern capital markets and political science have at least one thing in common: a dependence on data. But the resemblance stops there. When data quality declines in capital markets or when investors and analysts become insufficiently critical about price/earnings

Authors' note: For generous comments at many stages in the paper, the authors would like to thank Dawn Brancati, Bear Braumoeller, Kanchan Chandra, Jorge Dominguez, Errol D'Souza, Richard Grossman, Ana Grzymala-Busse, Andrew Kydd, David Laitin, Daniel Posner, Jasjeet Sekhon, Hillel Soifer, Jessica Wallack, and Steven Wilkinson and the Comparative Politics Research Workshop at Harvard University, and the anonymous reviewers from *Political Analysis*. The authors take full responsibility for any errors. An earlier version of this paper was presented at the American Political Science Association Annual Meetings, Boston, MA, August 2002.

© The Author 2007. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

ratios and revenues, debacles like Enron and Worldcom can happen. In cases like these, executives felt greater incentives to meet short-term targets for earnings growth than they did to produce accurate data. The consequences: shareholder lawsuits, regulatory and accounting reforms, jail sentences for executives, and investors' losing their shirts. When data quality slips in political science or when political scientists are insufficiently critical about the way their data were created or how they should be used, very little happens. Inattentiveness to data quality is, unfortunately, business as usual in political science.

We propose a heightened critical attention to data construction and a new way of looking at it: as an operation performed by *data actors* in a data supply chain. We know that data do not "grow on trees," yet we must occasionally remind ourselves that data are produced by people and entities according to their own *incentives* and *capabilities*. Despite strong disciplinary consensus about the behavioral effects of incentives, their effect on data actors has been woefully understudied by political scientists. Like all organizations, those that produce data are prone to problems of agency, bureaucratic incentives, shirking, and multiple principals and goals, all of which are likely to shape their output, that is, data. By turning our critical gaze inward, to the creation of the everyday data we take for granted, we hope to show the necessity of focusing on data quality, discipline-wide.

Ideally, we would like to make routine in the discipline such questions about data quality as, Who produced the data? Why? What were the producers' incentives and capabilities? Did they work as an independent agency, or were they influenced by external actors? Did the producers have incentives to shape the data rather than just report it? Such critical questioning is long overdue.

Although we advocate greater critical attention to the construction of data sets, we want to emphasize that our aim is *not* to question the utility of "large-*N* studies," where the large number of observations is critical to reliably address problems related to bias and measurement error. However, we do believe that there are serious weaknesses in many data sets used in cross-country regressions currently in vogue in political science. Therefore, addressing the strategic construction and use of data speaks directly to the validity of results.

The paper is divided into two sections. We first outline some problems with existing data set quality, including issues of validity, coverage, and accuracy; and we discuss some ways of identifying problems as well as some consequences of data quality problems. Subsequently, we examine how the incentives and capabilities facing four key actors in a data supply chain affect data quality: respondents, states (including bureaucracies and politicians), international organizations (IOs), and finally, academics. We conclude by making some suggestions for improving the use and construction of data sets.

2 Problems with Data Sets and Why They Matter

Problems of data quality are manifest and significant in a wide range of settings, from information collected by IOs and governments to the data sets compiled by individual scholars. They affect all sorts of indicators, from those more difficult to measure like identity variables, to the more objective indicators such as economic variables.

The *measurement* of data *quality*, however, has barely begun. Our framework for measuring it has three elements: validity, coverage, and accuracy. Validity refers to the relationship between theoretical concepts and collected information; coverage refers to the completeness of data sets; and accuracy refers to the correctness or avoidance of errors in data sets. We end this section of the paper by covering some ways to recognize quality problems and a brief discussion of consequences.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

2.1 *Validity*

Validity is at the heart of data quality because the objective of information collection in social science research is to enable one to draw inferences and test theories. If the connection between what is actually measured and what is purported to have been measured is tenuous (or absent altogether, in some cases), then the empirical enterprise breaks down. Gary Goertz (2005) has outlined three levels of social science research that provide a useful framework for thinking about validity: concepts, dimensions, and data. We can consider validity in terms of the relationship between each of these levels. For example, take democracy as the concept of interest to us. Depending on our definition of the concept, dimensions might include fairness of elections or civil liberties, and data for the first dimension might include the incumbency win rate or the margin of victory, whereas rights enumerated in the constitution, such as universal suffrage or the number of protests, might serve as data for the second dimension. Scholars might disagree on the definition of the concept itself and subsequently which dimensions should be used to measure it. They also might disagree on the data to be used for any particular dimension. This framework suggests that the starting point for assessing the validity of data sets must begin with the definition of concepts.

Unfortunately, many important concepts in political science remain undertheorized. There is still little theoretical agreement on basic definitions of concepts such as "rule of law," "corruption," and "identity." Consider "caste" for instance, a concept that many people believe plays an important role in social, political, and economic outcomes in India. Is caste a self-understanding or a socially ascribed category? An ethnic distinction or a class distinction? The answers to these definitional questions indicate different dimensions and types of data that would be needed to assess the real-world presence or absence of castes. Even "objective" variables such as gross national product (GNP) are not immune to such conceptual complexities, although decades of standardization of the System of National Accounts have led us to largely forget the tremendous amount of coordinated effort that went into defining GNP.

Despite the fundamental importance of concept-appropriate choices for measurement, too little attention has been paid to the construction of some of the most widely used indices and data sets. Some authors, notably Munck and Verkuilen (2002a), have suggested general standards for assessment of data sets and outlined a framework for evaluation that specifically draws attention to issues of conceptualization, measurement, and aggregation. And, the issue of measurement validity has been addressed by Adcock and Collier (2001) in the APSR. Unfortunately, however, much more attention to these methodological issues is needed in practice.

The Polity data series, one of the most widely used indices of democracy and authoritarianism in political science, offers a typical case of concept validity problems accompanied by a widespread absence of scrutiny by users. The analysis by Gleditsch and Ward (1997) of the third edition of Polity warned that "the analytical composition of the well-known democracy and autocracy scores is not upheld by an empirical analysis of the component measurements." Moreover, they argued that "democracy, as measured by the Polity indicators, is fundamentally a reflection of decisional constraints on the chief executive. The recruitment and participation dimensions are shown to be empirically extraneous despite their centrality in democratic theory" (Gleditsch and Ward 1997, 361). Our intention is not to single out Polity.[1] Although this finding about a data set that many of us take for granted is important, it is hardly unique.

[1]The data sets we use as examples in this paper were chosen not because they are particularly error prone, but rather because they are among the most widely used in political science. Discussion of their shortcomings is thus both relevant and illustrative for the entire field.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Another case of troubled concept validity was covered in a symposium on identity in the American Political Science Association (APSA) *Comparative Politics Newsletter* (Symposium 2001). The authors pointed out that although identity researchers predominantly rely on the constructivist paradigm, quantitative indices such as the Ethno-Linguistic Fragmentation index (ELF) remain primordialist.[2] The same can be said for the continued use of the very limited race and ethnicity categories on the U.S. census to measure "diversity." There appears to be a frustrating disconnect between conceptual and methodological advancements on the one hand and the continued use of theoretically outdated dimensions on the other.

Measurement validity addresses the next level: the relationship between dimensions and collected data. Despite the fact that measurement validity is a basic lesson in any introductory data analysis course in political science, the use of imprecise or concept-inappropriate indicators remains widespread in the field. This is evident in overt cases where data simply do not match a dimension. But there are many more subtle cases such as level-of-analysis problems where, for example, national data may be substituted for regional data or where recent annual data are not available and thus old data are used repeatedly. For example, caste data were last collected in India on the 1931 census, but, as the most current data available, these 1931 data continue to be used to explain contemporary phenomena.

A related issue in measurement validity is the problem of consistency, comparability, or reliability across countries. In brief, what is measured in one country, although it may go by the same name, may not be what is being measured in another country. For example, data purporting to measure "human capital" mainly depend on measures of education. However, the most frequently used measure, "years of schooling," cannot distinguish between years spent in a *madraasa* in Pakistan or a magnet school in the United States. Moreover, the production of precise numbers to code survey responses masks the incomparability that occurs when identical questions are interpreted differently by respondents.[3]

2.2 *Coverage*

A second major component of data quality is issue coverage—that is, the presence or absence of the data needed for a given research question. In many cases data on key variables of interest to scholars and governments are either incomplete or simply not collected at all, especially for certain types of countries.

In the worst cases, meaningful work on many important questions cannot be done at all. For most countries in the world, variation within countries cannot be analyzed since key political indicators, such as substate or regional measures of democracy, rule of law, and corruption, are not available. Similarly, beyond macroeconomic data, we lack information on several important economic indicators. We all recognize that a significant part of production and trade in less developed countries (LDCs) is carried out in the informal sector, yet there is a dearth of data on this vital part of the economy.

Some endemic coverage gaps are specific to certain parts of the world. Demographic data older than 20 years, such as the size and growth rate of the population, cannot be

[2]Efforts are underway to address this problem. For example, alternatives to the ELF include: the politically relevant ethnic group data set by Posner (2004); a constructivist data set on ethnic voting by Chandra et al. (2005); attempts to measure identity more generally (Abdelal et al. 2006); and an index of ethnonationalist mobilization (Cederman and Girardin 2005).

[3]There have been important recent attempts to address the problem of cross-cultural comparability of survey questions. See King and Wand (2004) and King et al. (2004).

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

unambiguously determined in more than a few African countries, with the margin of error often near 20%. The same is true of social statistics, such as those relating to literacy, school enrollment ratios, and poverty levels (Chander 1988). Closed societies also limit the availability of information. And finally, with the increasing use of online statistics and the prominence of the English language among Western social scientists, statistics that are not in English are more likely to be ignored than those that have been translated into English.

2.3 *Accuracy*

The final consideration of data quality is accuracy or the avoidance of outright errors at the level of data collection and presentation. Some errors are the result of methodological reforms whose new measurements indicate changes despite real-world constancy, and others are the result of biased data due to the subjectivity of respondents.

Apparent changes in data are sometimes due to changes in methodology. Measured infant mortality in the Soviet Union rose in the 1970s. According to Velkoff and Miller, however, Soviet infant mortality in all probability remained flat; what changed was the way in which it was measured (Velkoff and Miller 1995). Similarly, one reason why the growth of services may be a statistical artifact is the increased level of outsourcing in manufacturing firms. For instance, if General Motors spins off its design unit, the data will show a decline in manufacturing and an increase in services, even though little has changed in the real economy. And since many transactions in services are in the (unreported) informal sector, an economy that sees a shift from the informal to the formal sector will see faster growth in measured services compared to the actual change.

The subjectivity of respondents has been amply documented in survey research and poses obvious problems for data quality. Though underacknowledged, such bias is no less rife among the population of "experts" whose responses underpin widely used data sets like the Freedom House democracy ratings and Transparency International's corruption index. The generous Freedom House scores toward certain Central American countries in the 1980s may have reflected cold war, that is, anti-communist, understandings of democracy among experts; similarly, Transparency International largely measures *bureaucratic* corruption, rather than overall corruption, due to the types of people who give assessments. Close examination of these indices reveals that measures that rely on expert opinion can be biased by factors that affect the population of experts.

This criticism is not directed against using expert respondents to construct indices. Our intention, rather, is to emphasize the need to be circumspect and explicit about the subjective construction of such quantitative data sets, and thereby to better understand underlying biases and ultimately *improve* the construction and use of such data. Ostensibly objective data sets that quantify complex concepts such as "democracy," "governance," and "rule of law" are often based either on subjective surveys or on indexes whose weights are also subjective. That analysis is subjective is not a problem per se, but that it is often taken or imagined to be objective obscures the challenges of using data wisely to apprehend real-world phenomena.

2.4 *Recognizing Quality Problems*

How then does one identify problems with data quality? The two likeliest ways are by looking for discrepancies among sources or inconsistencies within publication series and by looking into external citation of problems.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Often one need only be a careful reader to uncover discrepancies either within the data produced by a single organization or between different organizations claiming to measure the same thing. The International Monetary Fund (IMF)'s primary statistical publication, *International Financial Statistics*, provides many instances where the data of the same year in books from different years do not match. Similarly, there are sometimes unexplained discrepancies between the print and electronic versions. This problem is by no means unique to the IMF. The World Bank offers data on GNP per capita growth rates for countries where underlying GNP data do not exist; they also report the share of agriculture in gross domestic product (GDP) for countries with nonexistent GDP estimates (Kapur, Lewis, and Webb 1997). Moreover, there is no evidence that these anomalies have ever been corrected. Another way to spot quality problems is to look for discrepancies between organizations: between 1981 and 1986, the IMF's GDP estimates for Zaire were about 60% of those of the World Bank.

Unfortunately, many government statistical offices do not fare much better than the IMF and World Bank, and there is no indication that the quality of statistics is improving over time. In India, the Central Statistical Organization (CSO) produces data on GNP and other macromeasures of the economy. On the other hand, the National Sample Survey Organization (NSSO) provides micromeasures of the economy through surveys on consumption, education, and so on. In principle, the consumption data estimated by the macroapproach of the CSO and the microdata aggregated from household surveys conducted by the NSSO should be equal, although some variations are inevitable. A few decades ago that was the case. More recently, the discrepancy between NSSO and CSO data has grown increasingly substantial: 1999/2000 NSSO figures showed consumption at just half the level of the CSO estimates. The weaknesses of India's national accounts data are also evident in the growing discrepancy between the expenditure and production estimates of GDP. A recent World Bank report points out that choosing between these estimates is not easy and that "the only conclusion that can be made confidently is that [India's] statistical architecture, once a model for other developing countries, needs more consistency checks" (World Bank 2000, para. 1.19). Whether or not India's people are getting poorer, "its statistics unquestionably are" (Aiyar 2001).

A second way to recognize quality problems is to review the data's external citation by scholars. Reviews and analyses of existing data sets are on the rise, a trend we strongly encourage. Munck and Verkuilen (2002a), for example, have evaluated nine data sets on democracy.[4] Some analyses have been cautionary. Assessing the latest, fourth edition of the Polity series, Treier and Jackman (2006) concluded that "skepticism as to the precision of the Polity democracy scale is well-founded, and that many researchers have been overly sanguine about the properties of the Polity democracy scale in applied statistical work." Others have been more forceful in their criticism. In assessing the Bretton Woods institutions, T. N. Srinivasan (1994, 4) stated bluntly: "publications of international agencies, such as the Human Development Report [of the United Nations Development Programme (UNDP)] and World Development Indicators of the World Bank, give a misleading, if not altogether false, impression of the reliability, comprehensiveness of coverage, comparability and recency of the data, and fail to warn the unwary users of the serious deficiencies in the data".

[4]Munck and Verkuilen (2002a) was followed by three discussion pieces as well as a response by the authors: see Coppedge (2002), Marshall et al. (2002), Munck and Verkuilen (2002b), and Ward (2002). For another evaluation of democracy measures, see Collier and Adcock (1999). For a painstaking analysis of trade statistics, see Yeats (1990) and Rozanski and Yeats (1994); on comparisons of governance indices, see Kaufmann, Kraay, and Zoido-Lobatón (1999a, 1999b) and Kaufmann, Kraay, and Mastruzzi (2002); on rule of law, see Berkowitz, Pistor, and Richard (2003); and on ethnicity, see Laitin and Posner (2001) and Wilkinson (2002).

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

2.5 *Consequences*

Problems of low data quality, that is, problems with validity, coverage, and errors, will affect the quality of political science research. Where concepts are not clearly defined, we should expect a lot of variance in both choices of dimensions, as well as inconsistencies in measurement of data across time and space. These quality problems will also affect the analysis and conclusions that can be drawn from the data. And, when data sets are used in quantitative analysis, there are also technical consequences. In terms of research results, several technical issues are relevant to the construction of data sets: measurement bias, measurement error and correlation of errors, and pooling or aggregation of measures.

Measurement *bias* is conceptually separate from measurement *error.* Where the measures themselves are biased, there are a host of complex issues and the consequences depend on how the measures are biased and how the models are parameterized.[5] The consequences of *measurement error* depend on where the errors are located and with what they are correlated. It is worth briefly considering the following types of errors.

1. *Measurement error in the dependent variable*: In this case the regression coefficients will have larger variances, leading to greater uncertainty regarding inference validity.
2. *Measurement error in uncorrelated independent variables*: As long as the independent variable is not correlated with any other independent variable, it will result in a biased coefficient for that variable and the coefficient will be attenuated toward zero. In other words, if one is certain that the independent variables are not correlated, measurement error in one such variable will make the estimate of that variable's effect biased downward, but the estimates of the other variables will be unaffected.
3. *Measurement error among correlated independent variables*: If the independent variables are correlated, then even random, unbiased measurement error in one single variable will lead to biased coefficients, and the direction of the bias is difficult to determine; in some cases the coefficients may even have the wrong sign (see Achen 1985). In other words, if independent variables are correlated and they almost always are in nonexperimental settings, then measurement error in only one variable can make the estimates of that variable's effect as well as other variables' effects inconsistent.
4. Measurement error in *independent variables correlated with* measurement error in *the dependent variable*: If this occurs then the correct specification assumption is violated and in general all the coefficients are biased.

Given these issues, the cross-country pooling of data and in particular the combination of data from Organisation for Economic Co-operation and Development (OECD) countries with LDC data may be problematic if it entails correlated measurement error or bias. If measures associated with LDCs have greater measurement error than the data from OECD countries (for reasons outlined below) and if the measurement error is correlated with other variables of interest, and perhaps with the dependent variable, then the results may be biased and inconsistent. And it is worth repeating that this is the case even if the measurement error itself is not biased.[6]

[5]For a more general discussion of measurement bias see White (1994).

[6]We hasten to add that the discussion of the consequences of correlated measurement error is in regard to ordinary least squares– and maximum likelihood–type estimators, two very commonly used models in political science.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

3 Data Actors and the Data Supply Chain: Incentives, Capabilities, and Consequences

Problems with data quality have not gone entirely unnoticed. Methodologists and statisticians are working to devise technical fixes for various problems in large data sets.[7] And a variety of scholars have individually endeavored to improve upon existing data sets[8] or to suggest novel indicators and measures.[9] These painstaking efforts at evaluation and corrections have so far received too little attention. The uncritical use of problematic data sets, without regard to these attempts at improvement, continues relatively unabated. Despite well-known problems, high-profile data sets like the Polity series, retain, in the words of Treier and Jackman (2006, 22), "near-canonical status.". All of which leads to a big question: *why do these problems with data set quality persist?*

Our answer to this question focuses on two factors: the incentives and capabilities of data actors. Data collection is of course costly, a factor which alone could explain some of the quality problems. But resources and budgets are not the only problem. Incentive structures facing both producers and users of data sets are an important part of the explanation as well: the incentives and capabilities of actors and institutions in the data supply chain have significant yet underacknowledged consequences for data quality.

Figure 1 schematically represents the supply chain of data production. It begins with original respondents—individuals, households, firms, and government agencies. The data collection agencies—state statistical institutions and private firms—are the next links in the data chain. State agencies can be both respondents and suppliers of data. As we move upstream, these data are supplied to IOs and nongovernmental organizations (NGOs), which have emerged as critical repositories of comparable cross-national data sets. Academic scholars receive and share data with IOs, but sometimes also receive data directly from either state statistical offices or private data collection firms. Although academics also collect data directly from respondents, the substantial costs of putting together large data sets means that their involvement is usually indirect by way of technical advice and assistance to IOs, NGOs, and data collection agencies. Similarly, IOs and NGOs also assist and therefore influence data collection agencies. This explains the dotted lines going back toward data collection agencies. Below we discuss each of these data actors in terms of incentives, capabilities, and consequences, summarized in Table 1.

3.1 *Respondents—Incentives*

The incentives for respondents include opportunity costs, fear of punishment, political support, and material gain. Opportunity costs come into play when the incentives to respond at all are weak. This is often the case when respondents see no direct benefit in participating, as when households are asked to complete census forms or firms are surveyed, without statutory provisions mandating participation. Census participation is encouraged by the threat of formal punishment in countries where answering the questionnaire is mandated by law.

[7] See, for example, the preceding discussion, as well as Treier and Jackman (2006) on adjustments to the Polity IV series. For attempts to address contextually specific effects across contexts, see Wong and Mason (1991), King and Wand (2004), and King et al. (2004).

[8] Examples of works attempting to update and amend the correlates of war data set include Bueno de Mesquita (1981, 21) and Slantchev (2004).

[9] There are far too many works to name here, but, for an example, see Mishler and Rose (2001) on measurement of political support in transitional regimes or Rose (2002/2003) on measurement of the informal economy in transitional regimes.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

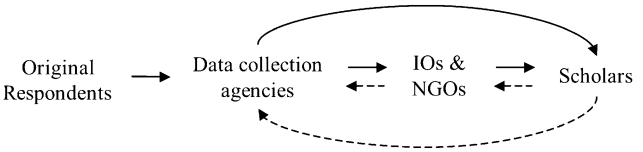


Fig. 1 Supply chain of data production.

Ironically, economic deregulation and political liberalization can reduce incentives if deregulation removes the legal obligation to respond. This was the case with the 1989 USSR census compared to the 2002 Russian census. Participation was mandatory in the former, voluntary and, not surprisingly, lower in the latter. Before liberalization in India in 1991, licensing requirements mandated that firms fill out surveys. With delicensing and the abolition of the government agency formerly responsible for the surveys, the response rate fell as the new agency lacked any statutory powers to compel responses (Nagraj 1999).

Mistrust of surveyors or fear of punishment for participation can be at work in both liberal as well as authoritarian regimes. In an environment where respondents do not trust surveyors or the state, they may be reluctant to respond openly to questions if they fear that that information might be used against them. Although this lack of trust is more likely in authoritarian regimes, it can also be a problem in democracies where privacy concerns may be primary.

Pressure to comply with state directives or the need to secure political support may provide incentives for respondents to deliberately misreport data. The same logic that motivates households in China to underreport their number of children for fear of prosecution also moved firms in the USSR to overestimate production in order to fulfill planning targets. Similarly, in China, an audit probe of 100 state-owned enterprises in 2003 found that 81 had falsified their accounts, 69 of which reported nonexistent profits. Even allowing for selection bias in the firms audited, can we trust the data reported by the 300,000-odd firms in the state sector and, in turn, China's overall economic statistics (Kynge 1999)?

Material gain is another incentive that affects respondents. In many countries, especially where the boundary between the tax authorities and the statistical office appears fluid, private entrepreneurs will understate earnings and output to avoid taxes. This is not only the case in places like China or Russia, and tax avoidance is not the only possible material incentive. In countries with capital controls and exchange-rate distortions, trade data are especially likely to be manipulated by firms, through underinvoicing of exports and overinvoicing of imports. And the spate of corporate accounting scandals in the United States testifies to the power of incentives on data integrity—in this case, the linkage between reported profit earnings and fat annual bonuses. Beyond economics, data on identity groups are also subject to material incentives: for example, the wide array of compensatory (affirmative action) measures in India has moved many to strategically misrepresent their caste origin in order to exploit state benefits.

When incentives pull actors in different directions in different countries, cross-national data sets are susceptible to particularly skewed results. Data on global fishery catches collected by the Food and Agricultural Organization (FAO) are a good example of this. Most fishermen tend to underreport their catches, and consequently, most countries can be presumed to underreport their catches to the FAO. Yet the catch statistics reported by China to the FAO continued to climb from the mid-1980s until 1998. Watson and Pauly (2001) found that the difference had less to do with fish than with the structure of domestic incentives in China, especially the link between promotions of fisheries officials for

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Table 1 Actors, incentives, capabilities, and consequences in the data supply chain

Actors	*Incentives*	*Capabilities*	*Data quality problems: validity, coverage, and accuracy*
Respondents (households, firms, and state employees)	• Opportunity costs • Fear of punishment (mistrust of surveyors) • Political support • Material gain	• Time • Knowledge • Level of education/ literacy • Access to surveys • Level of health	• Lack of response • Intentional misreporting • Selection bias in responses
Data collection agencies (state bureaucracies or private firms)	• Internal organizational/ professional norms • Material gain • External pressure (from governments, society, and IOs)	• Human capital • Financial resources from governments or IOs or researchers	• Lack of data collection or incomplete collection • Unintentional errors • Intentional misreporting/ manipulation of data • Selection bias in responses
IOs and NGOs	• Internal organizational/ professional norms • Support of donor states • Cooperation of respondent states	• Human capital • Financial resources from donor states	• Lack of data collection • Selection bias in responses
Academia	• Rewards for publication quantity • Rewards for theoretical contribution • Costs of data collection/ improvement • (For junior scholars) support of tenured scholars	• Time • Research funding • Existing data sets • Skills and technology for quantitative analysis	• Lack of new data sets • Continued use of low-quality data sets • Misuse of data that do not match dimensions or concepts

reported production increases. Statistics can thus be fishy in different ways depending on the different incentives for reporting across multiple countries.

3.2 *Respondents—Capabilities*

Respondents' resources and capabilities primarily consist of time, knowledge, level of education, access to surveys, and level of health. Respondents who work or are otherwise busy may have less time to answer surveys; this is true across countries and may be

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

a problem for the sample if certain types of people respond less frequently. Knowledge is another resource that varies, leading not only to variance in accuracy of responses but also to variance in response rates, if less knowledgeable people are less willing to participate. And knowledge may be related to level of education, as, for example, illiterate people would be less able to fill out written surveys. Access to surveys might also vary insofar as surveyors tend to be concentrated in larger urban areas rather than remote or rural locations. As respondents' capabilities vary, so will their responses, and if the capabilities are not evenly distributed in populations of interest, there may be selection bias in the responses.

3.3 *Respondents—Consequences*

The incentives and capabilities of respondents can result in nonresponses, intentional misreporting, and selection bias. Overcoming these factors, where possible, will depend on giving respondents more resources and positive incentives for participation. Unfortunately, changing incentives and capabilities is likely to involve expensive structural and institutional change, and is therefore a complicated, long-term problem. Selection bias can at least be compensated for by a range of statistical techniques and technical solutions, such as targeting samples, but one has to be able to identify it first.

3.4 *Data Collection Agencies—Incentives*

Data collection agencies include state statistical offices as well as private firms and NGOs charged with producing statistics. The bureaucrats who staff these agencies may face internal organizational incentives, or external political and economic incentives, such as support of IOs or material gain.

Internal organizational incentives may include factors as basic as professionalism. Agencies where both workers and management care about professionalism and reputation will tend to uphold international statistical norms. The quality of work will be higher when statisticians want to be recognized for meeting international professional standards. These professional norms are not insignificant considering the generally low status and low pay of public-sector statisticians around the world—and may explain high-quality state statistics in relatively poor countries such as Ecuador.

Such high professionals standards are, alas, rarely the case. Since many governments are inept, corrupt, and venal, especially in nondemocratic or poor countries, why would we expect their statistics departments to be substantially different? In other words, if the public sector in most LDCs is dysfunctional, in large part because of the inability or unwillingness to discipline shirking, we ought to expect similar behavior in those parts of the public-sector bureaucracy responsible for collecting data. Such situations, where even the principals are engaged in shirking, may lead to unintentional errors or incomplete data at best or intentional misreporting at worst.

The integrity of a national statistical agency's data is also affected by the independence of the agency from its government, usually the executive. Compared to the large literature on central bank independence, little analysis has been done on the relative independence of national statistical agencies. Historically, state statistics developed to meet the specific needs of governments and hence were biased toward serving government goals. This problem of government pressure continues in many countries, especially nondemocratic ones. In China, for instance, it is still quite difficult for public organizations to exist independently of the Communist Party. Consequently, local party leaders are the direct superiors of local National Bureau of Statistics functionaries, making it difficult for statisticians to act independently of the Party's wishes.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Even in democracies, state statistics may be subject to political pressure. In the United States, recent scandals over the manipulation of the costs of a prescription drug plan or intelligence on Iraq have called into question the independence of politically sensitive data. In federal states generally, subnational governments may have incentives to misreport or manipulate data submitted to federal or national governments in order to maximize transfers from the federal government.[10] Censuses may be particularly prone to such pressures because in many countries the allocation of state largesse, as well as political representation, is based on census data.

In some cases, the political implications of certain data may simply render data collection impossible. Many countries omit census questions regarding ethnicity or religion due to potential political fallout over results: for example, France does not ask the race or ethnicity of its citizens, and entire censuses have been stopped in countries such as Lebanon, Nigeria, and Pakistan because of fears that the results would favor certain groups.

IOs can also offer incentives to skew data. Central banks and finance ministries of countries undergoing an IMF program have an incentive to minimize their fiscal deficit data to meet IMF program targets, whereas European Union members have a similar incentive to meet the Maastricht criteria.

3.5 *Data Collection Agencies—Capabilities*

The capabilities of data collection agencies primarily consist of human capital and financial resources from governments, IOs, and scholarly researchers. Human capital is critical to the production of high-quality data. However, attracting high-quality individuals to work in government statistical agencies is a difficult task. Few would rank positions in state statistical agencies at the top of prestige hierarchies. In Russia, for example, the best statisticians (who have not gone to work for IOs) go to the Ministry of Finance or the Central Bank rather than the State Statistical Committee (Goskomstat). The latter's staff is overwhelmingly (90%) female, underscoring the well-known link between gender and occupational status. Russia is not alone on this issue: Rawski (2000) cites the Chinese case, where "the country's statistical agencies complain that firms assign often untrained staff to compile statistics, look for chances to cut positions assigned to statisticians, and refuse to submit standard reports." And China is much better able to compel compliance than most other countries.

In India's case, statisticians in the federal bureaucracy are recruited through an exam and interview conducted by a statutory autonomous body, the Union Public Service Commission. By any yardstick, the number of applicants taking exams for jobs in the federal government is extremely high (Table 2). However, as Table 2 indicates, in the case of the Indian Statistical Service, the number of applicants was the lowest and the application-to-post ratio the second lowest. Furthermore, it was the only service where the recommendation-to-post ratio was less than one, implying that qualified candidates were unavailable. If a country of a billion people which otherwise does not lack qualified professionals cannot find fifty qualified statisticians annually to staff its statistical bureaucracy, what does that say about the statistical capabilities of other poor countries that are much less well endowed?

In addition to human capital, data collection agencies and especially state statistical offices compete for financial resources from governments, IOs, and researchers. More

[10]On incentives for revenue forecasts among U.S. states, see Wallack (2006).

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Table 2 Statistical capabilities in the government of India

Service/exam	*No. of posts*	*No. of applicants*	*Application-to-post ratio*	*Recommendation-to-post ratio*
Civil Services[a]	411	309,507	753	1
Indian Forest Service	32	44,098	1378	1
Engineering Services	557	61,625	110	1
Indian Statistical Service	50	1,370	27	0.54
Geologist	148	3,647	25	1
Combined Medical Services	327	31,374	96	1

Source. Union Public Service Commission, 51st Annual Report 2000–01, table following para. 2.7, p. 12. Note: we have omitted data for less important services.
[a]The Civil Service Exam recruits India's elite federal bureaucracy including the Indian Administrative Service, Indian Foreign Service, Indian Revenue Service, Indian Account, and Audit Service, etc.

often than not, statistical offices are underfunded. We know that over the last two decades virtually all developing countries have undergone major financial and fiscal crises. When fiscally strapped countries have to cut their budgets, what are they likelier to cut: politically sensitive subsidies or support for institutional infrastructure, such as statistics departments? Indeed, when cast in such stark terms, this seems like a rhetorical question. Consider this comment on the state of support for the statistical system of a country whose "statistical agencies were having to make do with antiquated equipment, uncompetitive pay packages, and the elimination of less important (but still valuable) data series . . . . It was apparently easier [for that country] to subsidize [its] mohair industry, which cost more than the additional funding requested by the statistical agencies, than to ensure adequate data" (Swonk 2000). The comment was made of the political support for statistical offices in the United States. What then can we expect of poorer countries?

3.6 *Data Collection Agencies—Consequences*

When we consider the incentives facing data collection agencies as well as the generally weak capabilities of such agencies in terms of human and financial resources, there are several potentially negative consequences: lack of data collection or incomplete collection, unintentional errors, intentional misreporting or manipulation of data, and selection bias in responses. Lack of data collection or incomplete collection can be the result of a lack of resources, but these problems can also result from external pressure, as a way to hide embarrassing information about a state. Unintentional errors in the collection or processing of data are most likely to be the result of human or financial resource problems. Intentional misreporting and manipulation of data, however, are probably a result of external pressure.

Incentives that result in manipulation of data are especially manifest in those cases where the data are both a measure and a target. In pursing the target, the measure—and the data—is invariably contaminated. Hoskin (1996) writes that measures that are targets "precisely and systematically embody a conflation of the 'is' and the 'ought'; for their nature is simultaneously to describe and prescribe . . . measures as targets also prescribe what ought to be.". Consequently, when a measure becomes a target, it often ceases to be the appropriate measure.

This insight largely comes from Charles Goodhart's analysis of Margaret Thatcher's efforts to control inflation in Britain in the late 1970s by targeting the money supply.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Goodhart argued that, although there was a stable link between money supply and inflation, it might not persist if the government were to try to control the money supply. Goodhart's Law states that "as soon as a particular instrument or asset is publicly defined as money in order to impose monetary control, it will cease to be used as money and replaced by substitutes which will enable evasion of that control" (Goodhart 1989). In other words, when the measure (money supply) became a target, it ceased to be a good measure (of inflation), breaking down the relationship between money supply and inflation.

In China, local bureaucracies are often charged with collecting data as well as meeting targets set by their political principals, thereby increasing the likelihood that the data are subject to Goodhart's Law. When Beijing established the objective of 8% annual growth as a "great political responsibility," targeting the measure (GDP growth) vitiated that measure, resulting in the "winds of falsification" that affected the country's statistical reporting system (Rawski 2000). For example, in 1997–1998 the average growth rate reported by all 32 of China's provinces, main cities, and regions was 9.3%, even while the state statistics bureau's GDP growth rate was 7.8%!

China is hardly an exception. Under IMF programs, fiscal deficits are a critical target and therefore are becoming less meaningful as a measure, as governments learn to game the target. In 1999 the new government in Pakistan discovered that the previous regime had fudged budget figures between 1997 and 1999 to meet IMF program targets because budget deficits are a measure of the fiscal health of a country. In the EU, the rules of the Stability and Growth Pact were designed to ensure that countries had sustainable public finances. Any Euro-zone country reporting a deficit above 3% of GDP risks a large fine. However, "since countries collect their own numbers and report them to the EU, given the penalties of transgression, there is a clear incentive to cheat" (*The Economist* 2002), or to use such statistical sleight-of-hand as off-budget transactions, deferring liabilities and so on. The point is that these actions may become more pronounced when there are targets, thereby undermining the validity of the measures.

Some of the data commonly used in political science are in fact such data-skewing targets of governments and data collection agencies. Taking into account the incentives on data quality when data are both a measure and a target gives us insight into the direction of the biases that are likely to occur in such cases. When targets are ceilings (such as fiscal deficits), the data are likely to have downward bias. When targets are floors (such as social-sector indicators), the data are likely to be biased upward.

The quality of data can even be an indicator of the variable under investigation. Given that many governments, especially those in LDCs, suffer from limited capacities and weak institutions, we would a priori expect data-collecting institutions in LDCs also to be weaker. The quality of data produced by such states' statistical institutions might suffer from the same limited institutional capacity as the states themselves. The weak capacity of statistical agencies raises problems of endogeneity. Far too frequently, data are treated as exogenous to the problem being studied: in their work on governance indicators and institutional quality, Kaufmann, Kraay, and Zoido-Lobatón (1999a, 1999b) do not consider that where governance and institutional quality are weak, the quality of data is also likely to be weak—hence affecting their results.

3.7 *IOs/NGOs—Incentives*

IOs and NGOs play an important role in the collection and distribution of data sets across countries. Internal organizational incentives, such as professional norms, are as important

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

for such entities as for state agencies, but IOs and NGOs are also subject to pressure from their several donor states. Although they are unlikely to be pressured to meet targets by governments, they do need the cooperation of states in order to receive state-collected data.

Sometimes the data collection work of IOs is biased toward supporting the concerns of their donor states, as is the case with government debt data. The World Bank's "Global Development Finance" data set (formerly the World Debt Tables) is an exhaustive resource for the external debt of developing countries, but it reflects in part the interests of creditor countries, which exercise greater influence on the institution. By contrast, internal debt data are still much less easily available.[11]

Similarly, there is simply no comparison in the data quality regarding the two principal cross-border traffic flows—capital and labor—the former reflecting the endowments of the capital-rich North and the latter of the labor-rich South. It is therefore hardly surprising that data on international migration (labor) reflect many weaknesses in data quality.[12] Additionally, IOs and NGOs must secure the cooperation of states that supply data. Poor states that produce less data, states that are at war or facing other kinds of devastation (drought, HIV/AIDS, etc.), and closed societies in general are all less likely to cooperate with IOs and NGOs by providing data or allowing them to work inside the country.

3.8 *IOs/NGOs—Capabilities*

Like state data collection agencies, the capabilities of IOs and NGOs are primarily human and financial resources. IOs such as the United Nations (UN), IMF, and World Bank tend to have more resources than NGOs, such as Human Rights Watch or Greenpeace. But there is variation of course across these organizations in terms of both human and financial resources.

3.9 *IOs/NGOs—Consequences*

The chief quality consequence for IOs and NGOs as data actors is a likely lack of data collection on topics not supported by donor states and for poor or inaccessible countries. This can lead to selection bias in responses across countries, as UN development and poverty data show. In 2000, the largest ever gathering of heads of states adopted the UN Millennium Declaration aimed at advancing development and reducing poverty. It soon became apparent, however, that many member countries lacked data on development and poverty, and IOs did not have the capabilities to compensate for this glaring lacuna. A recent UN analysis of the relevant indicators found that "not only are there significant gaps for every indicator, there are also extensive problems in relevance, accuracy, consistency and reliability" (UNDP 2003, 35). The sheer number of countries where this is the case is starkly illustrated in Table 3.

3.10 *Academics—Incentives*

Finally, let us turn inward and look at political scientists as data actors susceptible to the same range of incentives and capabilities as other actors. All sorts of actors and situations have been studied with regard to the role of incentives, but rarely have we taken a critical gaze to the effect of incentives on academic research, particularly with regard to our use

[11]Evidence of this problem can be found in a recent paper by Brown and Hunter (1999), which uses debt service ratio as a variable but ignores internal debt because those data are not as easily available as external debt data.
[12]For a fuller discussion of data on international migration, see United Nations (2004).

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

Table 3 Data gaps in basic human development indicators, 1990–2001

Indicator	*Countries lacking trend data*	*Countries lacking any data*
Children underweight for age	100	22
Net primary enrollment ratio	46	17
Children reaching grade five	96	46
Births attended by skilled health personnel	100	19
Female share of nonagricultural wage employment	51	41
Urban HIV prevalence among pregnant women ages 15–24	100	91
Population with sustainable access to an improved water source	62	18
Population living on less than $1 a day	100	55

Note. A country is defined as having trend data if at least two data points are available, one in each half of the decade, and the two points are at least 3 years apart. Source: UNDP, 2003, Box 2.1.

and construction of data sets. The relevant incentives for academic scholars consist primarily of the following: rewards for publication quantity, rewards for theoretical innovation, rewards (or costs) for data collection and improvement, and support of other academics. This last incentive applies particularly to junior (untenured) scholars who need the support of senior faculty.

It almost goes without saying that scholars at research institutions are under intense pressure to publish their work. Getting tenure, remaining employed, and receiving pay raises at a research institution depend largely on the number and quality of a scholar's publications. Quality of publications matters, but that quality is not judged on the basis of the underlying data quality used in a publication. Instead, publication quality largely depends on the reputation of the journal or publisher and the theoretical contribution of the work, rather than the empirical contribution per se. As long as publication quantity and quality are judged primarily on the basis of outlet reputation or theoretical contribution, there is little incentive to improve data quality.

The incentives for new data collection or improving data quality are unfortunately rather limited. The costs of being attentive to quality in data are not trivial. Data collection and improvement are costly in time, skills, and financial resources. Moreover, the effort required to determine whether comparative data are truly comparative or whether individual elements do represent what they purport to, is substantial, and there is limited credit in tenure or review processes for those considered to be merely data collectors or correctors. The payoffs for data quality improvement are high only if the new and/or improved data set is used in some kind of innovative theoretical analysis. This means that in order to be recognized, those who work to improve data quality still have to do just as much theoretical or analytical work as those who do not bother with minding data quality.

Finally, academia is a community, and as such the support of other scholars constitutes an important incentive in individual work. Scholars' need for support varies according to career stage. Junior (untenured) scholars have more pressure to publish and also are more dependent on community support than senior scholars. Therefore, junior scholars have even less incentive to devote time to improvement of data quality, and junior scholars also have fewer incentives to be critical of existing data sets, especially if criticism would put

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

them at odds with senior faculty. The tenure process might be defended as a response to this incentive problem, in that it eventually gives scholars the freedom both to work longer on improving data and to criticize each other's work. However, junior scholars who have most recently done fieldwork are the likeliest to have fresh empirical knowledge, yet they are the least likely to engage in debates over data quality. The people most qualified are thus the least likely to devote time to data quality improvements.

3.11 *Academics—Capabilities*

Time, research funding, quantitative skills and technology, and existing data sets are the capabilities most in play for scholars. Because of their enormous expense, only a limited number of data set construction proposals will be funded. Fixing existing data sets—a less flashy task than coming up with something new—would be substantially less likely to find financial support. Unfortunately, although scholars may discover errors in existing work, there are not many low-cost options for correcting data errors.

Today scholars can access more off-the-shelf and downloadable data sets than ever before. Such resources afford researchers access to information about many places in the world about which they may not have specific area training or expertise. But the costs of in-depth fieldwork have not similarly declined, meaning that fieldwork remains quite expensive relative to off-the-shelf data sets. Given limited time and funding, freely available data sets can, and often do, substitute for new and/or improved data sets based on detailed fieldwork. And data sourced from reputable institutions (like the IMF, World Bank, OECD, the UN family, Polity, Freedom House, the Minorities at Risk project, or American National Election Studies) are all the more attractive because an institution's reputation gives the data sets a badge of credibility.

Finally, a researcher's skill level affects the type of data and analysis that he or she is capable of. In recent years, exogenous technological trends have led to a steep drop in the price of tools for quantitative analysis, such as better and cheaper software and hardware. These user-friendly advances require minimal statistical and mathematical training. The combination of new technology and greater availability of data sets may be driving down the cost of quantitative analysis. Such trends, though welcome, can also drive down the incentives and opportunities for improving data quality since researchers may be at too great a remove from the nitty-gritty of the data's construction to effectively scrutinize it.

3.12 *Academics—Consequences*

For academics, the worst consequence of our incentive and capability structures is the ongoing recycling of low-quality data and the failure to produce new data of high quality. Obviously, political science research would be more valuable if data quality improved; this would require individual scholars to devote more of their limited time and resources to improving data quality rather than producing more publications from existing flawed data sets. Because the resources, including time and money, that go into a publication are limited, trade-offs must be made. Work devoted to theoretical and model formulation and hypothesis testing using off-the-shelf data has to be weighed against the time it would take to improve the quality of a data set or to better match measurable indicators to concepts and dimensions. In order for researchers to focus on data quality, their incentives and capabilities would have to change: the use of high-quality data in publications ought to be rewarded, or at least it ought to be rewarded more highly than the use of lower quality data.

One problem with the current system of incentives is that the penalties for using low-quality data are small, and the costs of pointing out errors in data usage are high.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

If a researcher devotes his or her time to refuting the findings of a published article by using better data, the chance of publication (or benefit) is relatively high, but so too is the cost, because it takes a lot of time and effort to replicate and/or disprove results. Moreover, it is hardly a disgrace to be challenged empirically by future work; indeed it is a sign of interest in one's research agenda. Thus, the downside (or sanctions) for using low-quality data is rather limited.

In addition, some incentives for low-quality data use seem to be self-reinforcing. The more scholars that use existing flawed data sets, the more likely such data sets are likely to be used by others. In other words, *data are used because they are used*—and the data sets, problematic or not, become acceptable by repetition. Using reputable institutions only shifts the locus of the problem. The reputation of a prestigious data collection organization, such as those cited above, may actually reduce the incentives for scrutinizing the data: should there be any problems with the data, the data-collecting institutions, rather than individual researchers, would bear the brunt of the criticism.

4 Conclusion

Modern political science is data driven. If political scientists and institutional data actors were not trying to explain real outcomes, then data quality might not be so important. But to the extent that we are trying to develop and test theories about outcomes, data are the fundamental basis for our enterprise. We should expect that fundamental changes in the quality of information produced by political scientists, governments, and IOs would have substantial effects on public policy.

Some have asked, *are bad data better than no data?* We reject this either/or choice. "No data" or "bad data" are not the only choices because scholars need not be complacent with the status quo, and improvement of data sets is a continuous task. And thus, the best is not the enemy of the good. There were, are, and always will be shortcomings and limitations in data sets, and the costs of poor data must be traded-off against the opportunity costs of the effort required to improve the data. However, a focus on lowering the costs of data quality and changing the incentives for improving data quality will make higher quality data a likelier norm for the future.

Our conclusion is by no means that quantitative analysis based on large-*N* data sets should be limited or that data sets are inherently or irreconcilably flawed. Indeed, quantitative and statistical research is necessary for testing and improving data as well as testing theories.[13] We have pointed out problems in data quality and studied data actors' incentives and capabilities in order to suggest mechanisms for improvement of data sets, while at the same time discouraging continued use of overly troubled data sets.

In summary, we offer five broad suggestions: (1) encourage the production and dissemination of the growing literature on data quality and methods for improvement; (2) consider better ways to use data sets known to be flawed; (3) consider incentives as an instrument for improving data quality; (4) consider ways to lower the costs of producing high-quality data; and (5) consider institutional solutions to solve certain collective action problems related to data quality.

[13]One example is the discussion that followed the publication in 1996 of the Deininger and Squire dataset on income inequality by the World Bank. When the theorized relationship between economic growth and inequality using these dataset did not hold up, scholars scrutinized the dataset itself, calling into question certain measures. This in turn prompted further refinements of the data, as well as allowing for further testing of the theoretical relationship between inequality and other outcomes.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

There is a certain irony in the fact that methodology is a high-prestige area of political science and that a lot of work is devoted to improving *methods*, but that work on methods does not necessarily translate into improved everyday use of *data*. We believe that greater attention to the existing literature that evaluates data sets and to methodological issues concerning the use of data sets would be a step in the right direction.

As a first step, researchers should *examine* data sets. Researchers can subject data sets to some simple "smell-tests" by asking a number of questions: Who created the data? What incentives and capabilities were they subject to? Were they an independent agency? Were they governed by an external actor with a stake in the data? Subjecting the data to these questions will make the user more aware of possible quality problems with the data.[14] When data sets do have problems in their construction, we can at least be more circumspect about how we use them.

A second, related, recommendation is that we need to consider ways around bad data, that is, ways to better use data sets that are known to be flawed.[15] For example, if a data set for a particular dependent variable is flawed (due to validity issues, coverage, or errors), researchers should look for other observable implications of the argument and test those relationships. Triangulating by way of multiple tests of an argument using different data sets would be better and make a more convincing case than relying on one test using poor data.

In addition, as we pointed out in the section on validity above, selected data need to be appropriate for a given dimension of a concept under study. However, a positive implication of this obligation to match data with appropriate dimensions and concepts is that although a particular data set might be inappropriate for measuring a particular dimension or concept, it may well be appropriate for another dimension or concept. Thus, just because a data set is flawed for one purpose, it is not necessarily of no use for other purposes. The same logic applies to data sets with known biases. Researchers can make use of these biases, if they are acknowledged. If a significant relationship that works against the bias is found, we might be more confident of the results because the bias would move the results toward the null hypothesis. Rather than treating biased data sets as unusable, in the absence of improved data (which we discuss below), researchers should think about how to make use of flaws or biases in ways that might strengthen the validity of results. In other words, because measurement error is so endemic, part of the solution to dealing with measurement error is learning how to deal with it more sensibly. Although we do not have space to review the growing literature on statistical fixes for flawed data sets (many of which have appeared in the pages of this journal), those technical solutions are another obvious place to look for ways to deal with flawed data sets.

A third recommendation is that as a scholarly community, we must pay closer attention to incentives. Rather than treating data quality problems as an unfortunate result of ignorance or incompetence, we need to consider the incentives facing respondents, statistical offices, IOs, and scholars when they produce data. Given the degree to which researchers analyze the effects of incentives, their own supply elasticity of effort with respect to the incentives they themselves face might be presumed to be fairly high.[16] We suggest that the focus should be on ways to change these incentives to improve data quality.

[14]For example, since Becker's seminal article on crime (Becker 1968), researchers using officially reported crime statistics have had to be attentive to a number of quality issues. Errors due to underreporting by victims and underrecording by police may or may not be normally distributed, and an attentive researcher should check to see whether errors are systematically related to explanatory variables. Similarly, rather than relying only on one data source, researchers could compare data from a number of sources and consider the competency and independence of those sources.

[15]We thank one of the anonymous reviewers for bringing this point to our attention.

[16]The issue has been emphasized by Cheibub (1999) and Widner (1999).

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

The academic community as a whole needs to consider ways of lowering the costs of data quality. Increasing transparency and availability of the details of data sets, including coding, is a way to at least enable users to engage the data critically.[17] With more people able to recognize a data set's problems, the costs of improving the data set can be reduced. A few journals now mandate that authors make their data sets available upon request to readers. This is a positive development, but there are only minimal enforcement mechanisms for such rules. If authors fail to provide data or provide it in a form that is not very usable, the burden falls on the reader to pursue action.

If journals, on the other hand, made the data sets available on their Web sites, then it would be less costly for individual researchers to check and hopefully improve the quality of data sets. Additionally, a relatively low-cost error-revelation mechanism such as a "letters to the editor" section could be adopted by journals. *International Security*, for example, already has this in place. The proliferation of such mechanisms would have two effects: they would increase incentives for authors to attend to data quality by increasing the likelihood of being publicly criticized, and they would provide other scholars with important information regarding data errors, thus improving quality in future work with the same data sets.

Institutions also have roles to play in changing incentives. Small-scale institutional changes would include supporting more forums for error discussion and greater transparency. On a larger scale, major research funding agencies such as the NSF or the World Bank and UN need to make data quality a priority. Data quality in large grants could be improved if there were funding specifically earmarked for cleaning up existing or newly collected data sets and making them more widely accessible. Although the NSF does have an archiving requirement, it is not systematically enforced. Rather than the archiving component constituting a separate part of the grant, scholars have to take funds from some other part of their grant to work on fulfilling the archiving task, meaning they have less incentive to do so.

The APSA needs to take a leading role in advocating and perhaps codifying higher data quality norms. APSA as an institution might be able to overcome collective action problems among field and subfield sections, as well as among individual scholars. Given the importance of cross-country data sets, and the considerable scope for improving data comparisons across countries, we believe that debates regarding the merits of area studies versus cross-national large-*N* studies need to shift toward the collaborative possibilities between the two rather than the focus on competition. Joint work between area specialists as well as methodologists can considerably enhance the quality of cross-national data sets. However, there are considerable collective action problems inherent in organizing such efforts. APSA or other umbrella institutions may be able play a leadership role by supporting partnerships between area specialists and methodologists to improve existing data sets.

Finally, and on a more positive note, we wish to draw attention to some promising developments in recent years with regard to changing the incentive structures for researchers in constructing data sets. The Comparative Politics section of APSA, for example, now offers an award for data sets, and the Comparative Politics newsletter reviews new data

[17]The State Department analysis of terrorism provides a textbook case of how transparency of coding rules and availability of data can improve data quality. In April 2004, the State Department issued a report entitled "Patterns of Global Terrorism," claiming terrorist attacks had declined in recent years. Using the State Department's own guidelines which accompanied the report, Alan Krueger and David Laitin reviewed these data and found that "significant" terrorist attacks had actually risen between 2002 and 2003. They published this review of the data in an op-ed piece in the *Washington Post* and in an article in *Foreign Affairs*. In response, the State Department admitted that the report was wrong. For additional analysis of the State Department report, as well as recommendations for improving U.S. government data, see Krueger and Laitin (2004).

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

sets. In addition, a relatively new section of APSA, the Qualitative Methods section, is largely oriented toward taking empirical work, including the content of data sets, more seriously. And there have recently been a rising number of panels at professional meetings devoted to the consideration of the quality of data sets on a range of topics including ethnicity, democracy, and war. There are growing signs that institutional mechanisms for changing scholars' incentives—that is, reducing costs for producing high-quality data, and increasing rewards for using high-quality data—are underway.

There are many more ways that data quality can be improved which we have not had space to discuss here. We have endeavored to outline some problems with data quality and also to develop an explanation for the persistence of this problem, focused in particular on the incentives and capabilities among the data producers and users. Our goal has been to encourage further debate and serious consideration of the quality of political science data.

References

Abdelal, Rawi, Yoshiko Herrera, Alastair I. Johnston, and Rose McDermott. 2006. Identity as a variable. *Perspectives on Politics* 4:695–711.

Achen, Christopher. 1985. Proxy variables and incorrect signs on regression coefficients. *Political Methodology* 11:288–316.

Adcock, Robert, and David Collier. 2001. Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review* 95:529–46.

Aiyar, Swamininathan Anklesaria. 2001. Poverty-stricken statistics. *Economic Times*, September 1, 2001.

Becker, Gary. 1968. Crime and punishment: An economic approach. *Journal of Political Economy* 76:169–217.

Berkowitz, Daniel, Katharina Pistor, and Jean-Francois Richard. 2003. Economic development, legality, and the transplant effect. *European Economic Review* 47:165–95.

Brown, David, and Wendy Hunter. 1999. Democracy and social spending in Latin America. *American Political Science Review* 93:779–90.

Bueno de Mesquita, Bruce. 1981. *The war trap.* New Haven, CT: Yale University Press.

Cederman, Lars-Erik, and Luc Girardin. 2005. Beyond fractionalization: Mapping ethnicity onto nationalist insurgencies. *American Political Science Reviews* 101:173–85.

Chander, Ramesh. 1988. *Strengthening information systems in SSA.* Washington, DC: World Bank.

Chandra, Kanchan, ed. 2001. Symposium: Cumulative findings in the study of ethnic politics 2001. *APSA-CP Newsletter* 12(1):7–25.

Chandra, Kanchan, Rachel Giffelquist, Daniel Metz, Chris Wendt, and Adam Ziegfeld. 2005. A constructivist dataset on ethnicity and institutions. In *Identity as a variable*, ed. Rawi Abdelal, Yoshiko Herrera, Alastair Ian Johnston and Rose McDermott. New York: New York University.

Cheibub, Jose Antonio. 1999. Data optimism in comparative politics: The importance of being earnest. *APSA-CP* 10(2):21–5.

Collier, David, and Robert Adcock. 1999. Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science* 2:537–65.

Coppedge, Michael. 2002. Democracy and dimensions: Comments on Munck and Verkuilen. *Comparative Political Studies* 35:35–9.

Gleditsch, Kristian S., and Michael D. Ward. 1997. Double take: A reexamination of democracy and autocracy in modern polities. *The Journal of Conflict Resolution* 41:361–83.

Goertz, Gary. 2005. *Social science concepts: A user's guide.* Princeton, NJ: Princeton University Press.

Goodhart, Charles. 1989. *Money, information and uncertainty.* 2nd ed. Cambridge, MA: MIT Press.

Hoskin, Keith. 1996. The 'awful idea of accountability': Inscribing people into the measurement of objects. In *Accountability: Power, ethos, and the technologies of managing*, ed. Rolland Munro and Jan Mouritsen, 265–82. London: Thomson International.

Kapur, Devesh, John P. Lewis, and Richard Webb. 1997. *The World Bank: Its first half century.* Washington, DC: Brookings Institution.

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2002. Governance matters III: Governance indicators for 1996–2002. World Bank Policy Research Working paper 3106.

Kaufmann, D., A. Kraay, and P. Zoido-Lobatón. 1999a. Aggregating governance indicators. World Bank Working paper 2195.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press

———. 1999b. Governance matters. World Bank Working paper 2196.
King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* 98: 191–207.
King, Gary, and Jonathan Wand. 2007. Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis* 15:46–66.
Krueger, Alan, and David Laitin. 2004. Misunderestimating terrorism. *Foreign Affairs* 83(5):8–13.
Kynge, James. 1999. China uncovers falsified accounts at state groups. *Financial Times*, December 24, 1999.
Laitin, David, and Daniel Posner. 2001. The implications of constructivism for constructing ethnic fractionalization indices. *APSA-CP Newsletter* 12(1):13–17.
Marshall, Monty G., Ted Robert Gurr, Christian Davenport, and Keith Jaggers. 2002. Polity IV, 1800–1999: Comments on Munck and Verkuilen. *Comparative Political Studies* 35:40–5.
Mishler, William, and Richard Rose. 2001. Political support for incomplete democracies: Realist vs. idealist theories and measures. *International Political Science Review* 22:303–20.
Munck, Gerardo L., and Jay Verkuilen. 2002a. Conceptualizing and measuring democracy: Evaluating alternative indices. *Comparative Political Studies* 35:5–34.
———. 2002b. Generating better data: A response to discussants. *Comparative Political Studies* 35:52–7.
Nagraj, R. 1999. How good are India's industrial statistics? An exploratory note. *Economic and Political Weekly* 34:350–5.
Posner, Daniel N. 2004. Measuring ethnic fractionalization in Africa. *American Journal of Political Science* 48:849–63.
Rawski, Thomas G. 2000. *China by the numbers: How reform affected Chinese economic statistics.* http://www.pitt.edu/~tgrawski/papers2000/REVD00.HTM (accessed July 26, 2005).
Rose, Richard. 2002/2003. Economies in transition: A multidimensional approach to a cross-cultural problem. *East European Constitutional Review* 11/12(4/1):62–70.
Rozanski, J., and A. Yeats. 1994. On the (in)accuracy of economic observations: An assessment of trends in the reliability of international trade statistics. *Journal of Development Economics* 44:103–30.
Slantchev, Branislav L. 2004. How initiators end their wars. *American Journal of Political Science* 48:813–29.
Srinivasan, T. N. 1994. Data base for development analysis: An overview. *Journal of Development Economics* 44:3–27.
Swonk, Diane. 2000. The value of good data. *Financial Times*, September 27, 2000.
The Economist. 2002. Roll over, Enron. August 3, p. 44.
Treier, Shawn, and Simon Jackman. 2006. *Democracy as a latent variable*. http://www.tc.umn.edu/~satreier/DemocracyAsLatentVariable_041906.pdf.
United Nations. 2004. *Current status of the collection of international migration statistics. World Economic and Social Survey.* New York: United Nations, 211–7.
United Nations Development Programme (UNDP). 2003. *Human development report, 2003.* New York: Oxford University Press.
Velkoff, Victoria A., and Jane E. Miller. 1995. Trends and differentials in infant mortality in the Soviet Union, 1970–90: How much is due to misreporting? *Population Studies* 49:241–58.
Wallack, Jessica. 2006. *The highs and lows of revenue estimating: Explaining bias and inaccuracy.* San Diego, CA: University of California. http://irpshome.ucsd.edu/faculty/jwallack/revest_7_2006.pdf.
Ward, Michael D. 2002. Green binders in cyberspace: A modest proposal. *Comparative Political Studies* 35:46–51.
Watson, Reg, and Daniel Pauly. 2001. Systematic distortions in world fisheries catch trends. *Nature* 414:534–6.
White, Halbert. 1994. *Estimation, inference, and specification analysis.* New York: Cambridge University Press.
Widner, Jennifer. 1999. Maintaining our knowledge base. *APSA-CP* 10(2):17–21.
Wilkinson, Steve. 2002. *Memo on developing better indicators of ethnic and non-ethnic identities.* http://www.duke.edu/web/licep/5/wilkinson/wilkinson.pdf (accessed July 26, 2005).
Wong, George Y., and William M. Mason. 1991. Contextually specific effects and other generalizations of the hierarchical linear model for comparative analysis. *Journal of the American Statistical Association* 86: 487–503.
World Bank. 2000. *India: Policies to reduce poverty and accelerate sustainable development.* Report No. 19471-IN. Washington, DC: World Bank.
Yeats, Alexander. 1990. On the accuracy of African observations: Do Sub-Saharan trade statistics mean anything? *World Bank Economic Review* 2:135–56.

https://doi.org/10.1093/pan/mpm007 Published online by Cambridge University Press