Improving Data Quality:
Actors, Incentives and Capabilities

Yoshiko M. Herrera*
and
Devesh Kapur**

July 28, 2005

*Associate Professor
Department of Government
Harvard University
WCFIA, 1033 Massachusetts Ave.
Cambridge MA 02138

**Associate Professor
Department of Government
The University of Texas at Austin
1 University Station A1800
Austin, TX 78712-0119

Acknowledgments:
The authors would like to thank several people for generous comments at many stages in the paper, including Lisa Baldez, Dawn Brancati, Bear Braumoeller, Kanchan Chandra, Timothy Colton, Jorge Dominguez, Errol D'Souza, Richard Grossman, Ana Grzymala-Busse, Michael Hiscox, Andrew Kydd, Susan Pharr, Daniel Posner, Jasjeet Sekhon, Hillel Soifer, Jessica Wallack, and Steve Wilkinson. Of course 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.

Improving Data Quality:
Actors, Incentives and Capabilities

Abstract:
This paper examines the construction and use of datasets 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 dataset 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; states (including bureaucracies and politicians); international organizations; and finally, academic scholars. We conclude by making some suggestions for improving the use and construction of datasets. We present evidence from a variety of contexts but especially from Africa, China, India and Russia.

Key words:
Data quality, incentives, validity, coverage, accuracy, quantitative research, Africa, China, India, Russia

"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

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 P/E 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, i.e. 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? And 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 datasets, 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 datasets 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 dividend into two sections. We first outline some problems with existing dataset 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; and finally, academics. We conclude by making some suggestions for improving the use and construction of datasets.

I. PROBLEMS WITH DATASETS AND WHY THEY MATTER

Problems of data quality are manifest and significant in a wide range of settings, from information collected by international organizations and governments to the datasets 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 datasets. We end this section of the paper by covering some ways to recognize quality problems and a brief discussion of consequences.

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 (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, while 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 under-theorized. 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 datasets. Some authors, notably Munck and Verkuilen, have suggested general standards for assessment of datasets and outlined a framework for evaluation that specifically draws attention to issues of conceptualization, measurement, and aggregation (Munck and Verkuilen 2002a). And, the issue of measurement validity has been addressed by Adcock and Collier in the APSR (Adcock and Collier, 2001). 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. Gleditsch and Ward's analysis 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 dataset that many of us take for granted is important, it is hardly unique.

Another case of troubled concept validity was covered in a recent symposium on identity in the 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 US 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

[1] The datasets 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.
[2] Efforts are underway to address this problem. For example, alternatives to the ELF include: the politically relevant ethnic group dataset (PREG) by (Posner 2005); a constructivist dataset on ethnic voting by (Chandra et al. 2005); attempts to measure identity more generally (Abdelal et al. 2005); and an index of ethno-nationalist mobilization (Cederman and Giradin 2005).

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 U.S. Moreover, the production of precise numbers to code survey responses masks the incomparability that occurs when identical questions are interpreted differently by respondents.[3]

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 sub-state or regional measures of democracy, rule of law, and corruption are not available. Similarly, beyond macro-economic data, we lack information on several important economic indicators. We all recognize that a significant part of production and trade in less-

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

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 unambiguously determined in more than a few African countries, with the margin of error often near 20 percent. 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.

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 under-acknowledged, such bias is no less rife among the population of "experts" whose responses underpin widely used datasets like the Freedom House democracy ratings and Transparency International's corruption index. The generous Freedom House scores towards certain Central American countries in the 1980s may have reflected cold-war, i.e. 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 datasets, and thereby to better understand underlying biases and ultimately *improve* the construction and use of such data. Ostensibly objective datasets 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.

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.

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 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 GDP for countries with non-existent GDP estimates (Kapur et al., 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 1981-86, 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 macro-measures of the economy. On the other hand, the National Sample Survey Organization (NSSO) provides micro-measures of the economy through surveys on consumption, education, and so on. In principle, the consumption data estimated by the macro approach of the CSO and the micro data 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 figures NSSO

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 datasets are on the rise, a trend we strongly encourage. Munck and Verkuilen, for example, have evaluated nine datasets on democracy (Munck and Verkuilen 2002a).[4] Some analyses have been cautionary. Assessing the latest, fourth edition of the Polity series, Treier and Jackman 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" (Treier and Jackman 2003). Others have been more forceful in their criticism. In assessing the Bretton Woods institutions, T.N. Srinivasan stated bluntly: "publications of international agencies, such as the Human Development Report [of the UNDP] and World Development Indicators of the World Bank, give a misleading, if not altogether false, impression of the reliability,

[4] (Munck and Vekuilen 2002a) was followed by three discussion pieces as well as a response by the authors: see (Coppedge 2002); (Marshall et al. 2002); (Ward 2002); and (Munck and Verkuilen 2002b). 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 et al. 1999a, 199b, 2002). On rule of law, see (Berkowitz et al. 2003). And on ethnicity, see (Laitin and Posner 2001); (Wilkinson 2002); and (Abdelal et al. 2005).

comprehensiveness of coverage, comparability and recency of the data, and fail to warn the unwary users of the serious deficiencies in the data" (Srinivasan 1994: 4).

Consequences

Problems of low data quality, i.e. 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 datasets: 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.[6]

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 towards zero. In other words, if one is

[5] For a more general discussion of measurement bias see (White, 1994).
[6] Similarly, correlation among correctly specified (i.e. error-free) independent variables will lead to larger coefficient variances.

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. As long as there is no correlation among independent variables, it is possible to correct for even biased measurement error in one variable using a range of statistical techniques such as robust estimators.[7]

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 non-experimental 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 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

[7] There are statistical models which specifically take measurement error into account, such as LISREL models and a variety of robust estimators such as the hyperbolic tangent (tanh) estimator, but these are seldom used. For more on robust estimators see (Mebane and Sekhon, 2004) and (Wand et al., 2001).

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.[8]

II. 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 datasets.[9] And a variety of scholars have individually endeavored to improve upon existing datasets[10] or to suggest novel indicators and measures.[11] These painstaking efforts at evaluation and corrections have so far received too little attention. The uncritical use of problematic datasets, without regard to these attempts at improvement, continues relatively unabated. Despite well-known problems, high-profile datasets like the Polity series, retain, in the words of Treier and Jackman, "near-canonical status" (Treier and Jackman 2003, 43). All of which leads to a big question: *why do these problems with dataset 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 datasets are an important part of the explanation as well: the incentives

---

[8] 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.

[9] See for example the preceding discussion, as well as (Treier and Jackman 2003) on adjustments to the Polity IV series. For attempts to address contextually specific effects across contexts, see (Wong and Mason 1991); (King et al. 2004); and (King and Wand 2004).

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

[11] 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 2002/03) on measurement of the informal economy in transitional regimes.

and capabilities of actors and institutions in the data supply chain have significant yet under-acknowledged consequences for data quality.

[Figure 1 somewhere here]

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 international organizations, which have emerged as critical repositories of comparable cross-national datasets. Academics receive and share data with international organizations, but sometimes also receive data directly from either state statistical offices or private data collection firms. Below we discuss each of these data actors in terms of incentives, capabilities, and consequences, summarized in Table 1.

[Table 1 somewhere here]

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 force in countries where answering the questionnaire is mandated by law.

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 de-licensing 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 non-existent 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 under-invoicing of exports and over-invoicing of imports. And the spate of corporate accounting scandals in the US 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 datasets 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 under-report their catches, and consequently, most countries can be presumed to under-report 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 reported production increases. Statistics can thus be fishy in different ways depending on the different incentives for reporting across multiple countries.

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 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.

Respondents — Consequences

The incentives and capabilities of respondents can result in non-responses, 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, 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.

Data Collection Agencies — Incentives

Data collection agencies include state statistical offices as well as private firms and non-governmental organizations 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 international organizations 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 non-democratic 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 towards serving government goals. This problem of government pressure continues in many countries, especially non-democratic 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 (NBS) functionaries, making it difficult for statisticians to act independently of the Party's wishes.

Even in democracies, state statistics may be subject to political pressure. In the U.S., 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, sub-national governments may have incentives to misreport or manipulate data submitted to federal or national governments in order to maximize transfers from the federal government.[12] 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: e.g., 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.

International organizations 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, while EU members have a similar incentive to meet the Maastricht criteria.

Data Collection Agencies — Capabilities

The capabilities of data collection agencies primarily consist of human capital and financial resources from governments, international organizations, 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 haven't gone to work for international organizations) 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 links between

[12] On incentives for revenue forecasts among U.S. states, see (Jessica Wallack 2004).

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?

[Table 2 somewhere here]

In addition to human capital, data collection agencies, and especially state statistical offices compete for financial resources from governments, IOs, and researchers. More often than not, statistical offices are under-funded. 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 hidden state 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 U.S. What then can we expect of poorer countries?

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 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" (Hoskin 1996). 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. 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 1997-98 the average growth rate reported by all 32 of China's provinces, main cities and regions was 9.3 percent, even while the state statistics bureau's GDP growth rate was 7.8 per cent!

China is hardly an exception. Under IMF programs, fiscal deficits are a critical target, and therefore became 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 percent 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 upwards.

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 et al. (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.

IOs/NGOs — Incentives

International organizations (IOs) and non-governmental organizations (NGOs) play an important role in the collection and distribution of datasets across countries. Internal

organizational incentives, such as professional norms, are as important for such entities as for state agencies, but IOs and NGOs are also subject to pressure from their several donor states. While 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 towards supporting the concerns of their donor states, as is the case with government debt data. The World Bank's "Global Development Finance" dataset (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.[13]

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) reflects many weaknesses in data quality.[14] 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.

IOs/NGOs — Capabilities

Like state data collection agencies, the capabilities of IOs and NGOs are primarily human and financial resources. IOs such as the UN, IMF, and World Bank tend to have more resources

[13] 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.

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.

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 international organizations 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.

[Table 3 somewhere here]

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 and construction of datasets. 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.

---

[14] For a fuller discussion of data on international migration, see (United Nations 2004).

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. As long as publication quantity and quality are judged 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 dataset 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 datasets, especially if criticism would put 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.

Academics — Capabilities

Time, research funding, quantitative skills and technology, and existing datasets are the capabilities most in play for scholars. Because of their enormous expense, only a limited number of dataset construction proposals will be funded. Fixing existing datasets — 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 datasets 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 datasets. Given limited time and funding, freely available datasets can, and often do, substitute for new and/or improved datasets 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 NES) are all the more attractive because an institution's reputation gives the datasets 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 datasets 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.

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 datasets. 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 dataset 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 meaningfully 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. 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's hardly a disgrace to be challenged empirically by future work; indeed it's 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 datasets, the more likely such datasets are likely to be used by others. In other words, *data are used because they are used* — and the datasets, 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 of the data, the data-collecting institutions, rather than individual researchers, would bear the brunt of the criticism.

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 international organizations 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 datasets 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 datasets, 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 datasets should be limited or that datasets are inherently or irreconcilably flawed. Indeed, quantitative and statistical research is necessary for testing and improving data as well as testing theories.[15] We have pointed out problems in data quality and studied data actors' incentives and capabilities in order to suggest mechanisms for improvement of datasets, while at the same time discouraging continued use of overly troubled datasets.

In summary, we offer four broad suggestions: 1) Encourage the production and dissemination of the growing literature on data quality and methods for improvement; 2) Consider incentives as an instrument for improving data quality; 3) Consider ways to lower the costs of producing high-quality data; and 4) Consider institutional solutions to solve certain collective action problems related to data quality.

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 doesn't necessarily translate into improved everyday use of *data*. We believe that greater attention to the existing literature that evaluates datasets, and to methodological issues concerning the use of datasets would be a step in the right direction.

---

[15] One example is the discussion that followed the publication in 1996 of the Deininger and Squire dataset on income inequality. 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

As a first step, researchers should *examine* datasets. Researchers can subject datasets 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.[16] When datasets do have problems in their construction, we can at least be more circumspect about how we use them.

A second thing to do is to pay closer attention to incentives. Rather than treating data quality problems as an unfortunate result of ignorance or incompetence, consider the incentives facing respondents, statistical offices, international organizations, 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.[17] We suggest that the focus should be on ways to change these incentives to improve data quality.

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 datasets, including coding, is a way to at least enable users to engage the data critically.[18] With more people able to recognize a

---

allowing for further testing of the theoretical relationship between inequality and other outcomes. For the dataset, see http://www.worldbank.org/research/growth/dddeisqu.htm.

[16] 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 under-reporting by victims and under-recording 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.

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

[18] The recent 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

dataset's problems, the costs of improving the dataset can be reduced. A few journals now mandate that authors make their datasets 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 datasets available on their websites, then it would be less costly for individual researchers to check and hopefully improve the quality of datasets. 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 datasets.

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 datasets and making them more widely accessible. Although the NSF does have an archiving requirement, it is not

---

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).

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 American Political Science Association (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 sub-field 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 datasets.

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 datasets. The Comparative Politics section of APSA, for example, now offers an award for datasets, and the Comparative Politics newsletter reviews new datasets. In addition, a relatively new section of APSA, the Qualitative Methods section, is largely oriented towards taking empirical work, including the content of datasets, more seriously. And there have recently been a rising number of panels at professional meetings devoted to the consideration of the quality of datasets on a range of topics including ethnicity, democracy, and war. There are

growing signs that institutional mechanisms for changing scholars' incentives — i.e. 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, eds. 2005. *Identity as a Variable*. Unpublished manuscript, July 2005.

Achen, Christopher. 1985. "Proxy Variables and Incorrect Signs on Regression Coefficients." *Political Methodology* 11:3-4, 288-316.

Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95:3, 529-546.

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." Forthcoming in the *European Economic Review* 47:1, 165-195.

Brown, David and Wendy Hunter. 1999. "Democracy and Social Spending in Latin America." *American Political Science Review* 93:4, 779-790.

Bueno de Mesquita, Bruce. 1981. *The War Trap*. New Haven, Conn: Yale University Press.

Cederman, Lars-Erik and Luc Girardin. 2005. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." Unpublished manuscript, 27 February 2005.

Chander, Ramesh. 1988. *Strengthening Information Systems in SSA*. Washington, D.C.: World Bank.

Chandra, Kanchan, Rachel Giffelquist, Daniel Metz, Chris Wendt, and Adam Ziegfeld. 2005. "A Constructivist Dataset on Ethnicity and Institutions." Unpublished manuscript, 2005.

Cheibub, Jose Antonio. 1999. "Data Optimism in Comparative Politics: The Importance of Being Earnest." *APSA-CP* 10:2, 21-25.

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* 35:1, 35-39.

The Economist. 2002. "Roll Over, Enron." August 3, 2002, p. 44.

Gleditsch, Kristian S. and Michael D. Ward. 1997. "Double Take: A Reexamination of Democracy and Autocracy in Modern Polities." *The Journal of Conflict Resolution* 41:3, 361-383.

Goertz, Gary. 2005 *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press. Forthcoming.

Goodhart, Charles. 1989. *Money, Information and Uncertainty*. 2nd ed. Cambridge, Mass.: 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. London: Thomson International, 265-282.

Kapur, Devesh, John P. Lewis, and Richard Webb. 1997. *The World Bank: Its First Half Century*. Washington, D.C.: 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.
Kaufmann, D., A. Kraay, and P. Zoido-Lobatón. 1999b. "Governance Matters." World Bank Working Paper 2196.
King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95:1, 49-69.
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* 98:1, 191-207.
King, Gary and Jonathan Wand. 2004. "Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes." Unpublished manuscript, 14 November 2004.
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*, Dec. 24, 1999.
Laitin, David and Daniel Posner. 2001. "The Implications of Constructivism for Constructing Ethnic Fractionalization Indices." *APSA-CP*, (Winter), 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* 35:1, 40-45.
Mebane, Walter, Jr. and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." American Journal of Political Science, 48 (April): 391-410.
Mishler, William and Richard Rose. 2001. "Political Support for Incomplete Democracies: Realist vs. Idealist Theories and Measures." International Political Science Review 22:4, 303–320.
Munck, Gerardo L. and Jay Verkuilen. 2002a. "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices." *Comparative Political Studies* 35:1, 5-34.
Munck, Gerardo L. and Jay Verkuilen. 2002b. "Generating Better Data: A Response to Discussants." *Comparative Political Studies* 35:1, 52-57.
Nagraj, R. 1999. "How Good Are India's Industrial Statistics? An Exploratory Note." *Economic and Political Weekly* 34:6, 350-355.
Posner, Daniel. 2005. "PREG Values for African and Asian Countries." Forthcoming at http://www.polisci.ucla.edu/faculty/posner/data.htm.
Rawski, Thomas G. 2000. "China by the Numbers: How Reform Affected Chinese Economic Statistics." http://www.pitt.edu/~tgrawski/papers2000/REVD00.HTM. Last accessed: 26 July 2005.
Rose, Richard. 2002/2003. "Economies in Transition: A Multidimensional Approach to a Cross-Cultural Problem." *East European Constitutional Review* 11:4/12:1, 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* 44:1, 103-130.

Slantchev, Branislav L. 2004. "How Initiators End Their Wars." *American Journal of Political Science* 48:4, 813-29.
Srinivasan, T.N. 1994. "Data Base for Development Analysis: An Overview." *Journal of Development Economics* 44:1, 3-27.
Swonk, Diane. 2000. "The Value of Good Data." *Financial Times*, September 27, 2000.
"Symposium: Findings in the Study of Ethnic Politics." 2001. *APSA-CP Newsletter* 12:1, 7-25.
Treier, Shawn and Simon Jackman. 2003. "Democracy as a Latent Variable." 2003 Annual Meeting of the Society for Political Methodology, University of Minnesota, Minneapolis, July 17-19.
United Nations. 2004. "Current Status of the Collection of International Migration Statistics," *World Economic and Social Survey*. New York: United Nations, 211-217.
United Nations Development Programme. 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* 49:2, 241-258.
Wallack, Jessica. 2004. "The Highs and Lows of Revenue Estimating." Unpublished manuscript, 16 June 2004.
Wand, Jonathan et al. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95:4, 793-810.
Ward, Michael D. 2002. "Green Binders in Cyberspace: A Modest Proposal." *Comparative Political Studies* 35:1, 46-51.
Watson, Reg and Daniel Pauly. 2001. "Systematic Distortions in World Fisheries Catch Trends." *Nature* 414, 534-536.
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. Last Accessed, 26 July 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* 86:414, 487-503.
World Bank. 2000. "India: Policies to Reduce Poverty and Accelerate Sustainable Development." Report No. 19471-IN. Washington, D.C.: World Bank.
Yeats, Alexander. 1990. "On the Accuracy of African Observations: Do Sub-Saharan Trade Statistics Mean Anything?" *World Bank Economic Review* 2:4, 135-156.

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, 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, 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
International and Non-Governmental Organizations	• 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 datasets • skills and technology for quantitative analysis	• lack of new datasets • continued use of low-quality data sets • misuse of data that do not match dimensions or concepts

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.

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 non-agricultural 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 three years apart. Source: UNDP, 2003, Box 2.1.

Figure 1. Supply-chain of data production