

NEW STRATEGIES FOR THE COMPARATIVE ANALYSIS OF LATIN AMERICAN POLITICS*

Philippe C. Schmitter, University of Chicago†

... We now live in an era in which it is scarcely worthwhile to lie without statistics.

Raymond Bauer

SEVERAL YEARS AGO, STUDENTS OF LATIN AMERICAN POLITICS DISCOVERED with some alarm that the subdiscipline of comparative politics had not only been ignoring their scholarly efforts, but the area altogether. At that time the principal focus of discontent was conceptualization. Classification systems, typologies, checklists, functions and isolated concepts about the politics of transition were being derived and applied without reference to and relevance for Latin America.¹ While the terminological estrangement has by no means ended, some reconciliation has subsequently occurred. Recent theoretical works make occasional references to the area.² It has become essential for all readers or collections of essays on political development to contain at least one (often the same) article on Latin American politics.³ Conversely, new research in Latin America has been increasingly sensitive to approaches prevalent in the general comparative politics literature.⁴

Meanwhile, that literature has continued to expand at an exponential rate. Its axis has shifted, however. In particular, scholars have been placing greater emphasis on operational strategies—upon passing directly to problems of discrete measurement, analysis of fit and causal inference. In many instances conceptual disputes have been thrust aside or postponed in the drive for quantification and index construction.

One result of this new emphasis on aggregate data has been the publication of the *World Handbook of Political and Social Indicators* and the *Cross Polity Survey*⁵ as well as the establishment of a growing number of data banks at various universities.⁶ These are universalist in scope and, hence, count, score

* Prepared for delivery at the Latin American Studies Association meeting, New York City, November 7–9, 1968.

† I would like to acknowledge the support of the Social Sciences Research Council of the University of Chicago for computational expenses and the Center for the Comparative Study of Political Development for secretarial assistance. A special vote of thanks I offer to Mrs. Evelyn Hortik for sticking it out during a "long, hot summer."

or categorize information on the twenty Latin American republics along with all others. Individual authors have constructed new multivariate indicators, coded variables of their own, introduced error-compensating measures and manipulated this growing body of information in a variety of ways. The purpose of this work has been to describe ranges of variance, to create new typologies, to test for association and to infer causal relations cross-nationally. These authors have also stimulated and themselves contributed to a growing body of methodological criticism about the potentialities and limitations of this mode of analysis.⁷

Aggregate data analysis—the attempt to explain patterns of similarity and difference through the statistical examination of information about areal or collective units—is, then, a growing field of inquiry and one with important implications for the study of Latin American politics. Whereas formerly the problem with functional and ideal type approaches was their ignorance and avoidance of Latin America, this approach wholeheartedly incorporates the region, either as part of a universal sample or, increasingly, as a distinct subsample. Like it or not, the Latin Americanist will be finding his stock assumptions and pet hypotheses being examined and challenged by this form of analysis.

METHODOLOGICAL PROBLEMS OF DATA COLLECTION

The advantages of aggregate data can be resumed succinctly: (1) ready, public availability; (2) inexpensiveness of compilation; (3) freedom from sampling error (if they consist of total enumerations); (4) coverage of extended time spans; (5) facility with which they can be broken down into smaller areal units or built up into larger ones; (6) great statistical flexibility.⁸ The disadvantages deserve more extensive treatment. But before examining such problems as availability, comparability, sensitivity, reliability and validity within the Latin American context, I would like first to make a few comments about the utility of universally-oriented books and data banks for our specialty.

When one considers the ambitious scope and pioneering nature of these compilations, they come off rather well. For example, the Yale *Handbook* breaks new ground in assessing data reliability and in exploring analytical techniques. The Banks and Textor *Survey* makes important innovations in the field of classification.

Yet existing broad-scale compilations such as these pose two special problems of relevance for the area specialist. While interval-type data in the *World Handbook of Political and Social Indicators* is presented in superior tabular form and is available for machine use, a quick comparison with such sources as the United Nations' *Demographic Yearbook*, ECLA's *Statistical Bulletin*

for Latin America, the OAS' *América en Cifras* or the UCLA-produced *Statistical Abstract of Latin America* will demonstrate its limited utility. Recognizing that the number of indicators for which comparable data can be gathered for some 130 countries is quite restricted, the Yale authors still are conspicuously silent on the concepts which their indicators are to serve and there are many, especially those related to political system characteristics and policy outcomes, which cannot be operationalized by any one or any combination of their seventy-five variables. But for those of us who are students of Latin American politics, these are the concepts which interest us the most. Why should we admit such limitations by crude or non-existent indicators, when they are often available in much greater variety and detail for our sub-sample of twenty units?

Take, for example, such indicators of policy outcomes as the total amount and distribution of governmental expenditures and revenue. The *Handbook* offers three tabulations of gross revenues. In the first two, Brazil is the *only* Latin American country to appear; in the third, data is offered on Venezuela, Peru, Bolivia and Argentina. Two distributions of expenditures are presented: (1) social security and public enterprise; (2) defense. Only in the latter example are more than five Latin American countries represented. In my data bank I have some twenty-five variables measuring these "performances" in much greater detail (and often for several time periods) with a case base varying from seventeen to twenty.

Another limitation of the Yale *Handbook* is its exclusive focus on synchronic or cross-sectional data. The time period covered consists principally of the 1950s, although some information comes from 1920 and some as recently as 1961. Most of the hypotheses of interest to me involve process relationships with time-lagged or time-led variables. Thanks again to greater availability of aggregate data in Latin America, we can utilize such diachronic information. In the data bank I am currently using most variables can be scored for both 1950 and 1960. Although the case base would undoubtedly shrink, time-series could be compiled on such variables as total population, economic growth, external trade, government revenues, capital flows, party systems, electoral participation, or even budgetary expenditures for a longer period of time. Of course, the general censuses listed in Table I should not be considered by any means to contain comparable information over the time periods indicated—even for a single country—but when coupled with reports of other governments, private and public leading agencies, international organizations, national councils, planning boards and electoral commissions, the range is impressive and should not be left unexploited.

The Banks and Textor *Survey* contains *categorized* data, some nominal, e.g. "Constitutional Status": (A) Constitutional, (B) Authoritarian, (C) Totalitarian; and some ordinal, e.g., "Stability of Party System": (A) Stable,

TABLE I
General Censuses in Latin America: 1850-1966

	Prior to 1900	1900-1919	1920-1939	1940-1949	1950-1959	1960's
Argentina	1869, 1895	1914	—	—	—	1960
Bolivia	—	1900	—	1947	—	—
Brazil	1872, 1890	1900	1920	1940	1950	1960
Chile	1864-5, 1875	1907	1920, 1930	1940	1952	1960
	1885, 1895	—	—	—	—	—
Colombia	1851, 1864, 1870	1905, 1912	1938	—	1951	1964
	—	1918	—	—	—	—
Costa Rica	1864, 1892	—	1927	—	1950	1963
Cuba	1899	1907, 1919	1931	1943	1953	—
Dominican Republic	—	—	1920, 1935	—	1950	1960
Ecuador	—	—	—	—	1950	1962
El Salvador	—	1901	1930	—	1950	1961
Guatemala	1880, 1893	—	1921, 1930	1940	1950	1964
Haiti	—	—	—	—	1950	—
Honduras	1887	1901, 1905	1926, 1930	1940	1950	1961
	—	1910, 1916	1935	—	—	—
Mexico	1885	1900, 1910	1921, 1930	1940	1950	1960
Nicaragua	—	—	1920	1940	1950	1963
Panama	—	1911	1920, 1930	1940	1950	1960
Paraguay	—	1917	—	—	1950, 1956	1962
	—	—	—	—	1958	—
Peru	1876	—	1933	1940	—	1961
Uruguay	1860	1900, 1908	—	—	—	1963
Venezuela	1873, 1881, 1893	1910	1920, 1926	1941	1950	1961
	—	—	1936	—	—	—

Source: "Census Nacionales de Países de América Latina. Generales y de Población, Instituto Torcuato di Tella, 2nd ed., 1966. "Situación Demográfica: Estado y Movimiento de la Población," *América en Cifras-1967* (OAS, Instituto Interamericano de Estadística, 1968).

(B) Moderately Stable, (C) Unstable. The authors proudly announce that the sources of these categories are "concepts considered analytically important by leading scholars in the field of comparative politics."⁹ However, these are often precisely the sort of concepts which Latin Americanists have found to be divorced from the political and social reality of their area. They are, at worst, ethnocentric derivations from Anglo-American political experience; at best, sensitive attempts to order observations about African and Asian polities.

Three examples will illustrate the problem. Banks and Textor offer a nominally coded variable: "Ideological Orientation." The categories are doctrinal, developmental, situational, conventional, traditional. Three Latin American countries rate "developmental": Dominican Republic, Venezuela, Bolivia; two are "situational": El Salvador and Haiti; seven score "conventional": Chile, Colombia, Costa Rica, Ecuador, Honduras, Mexico and Uruguay; none are "doctrinal." All the rest are uncoded—either on grounds of ambiguity or lack of information. Beyond the obvious critique that their categories violate the elementary canons of any classification system, i.e. they are neither jointly exhaustive nor mutually exclusive, they result in some rather strange groupings and the "genocide by categorization" of a substantial number of Latin American countries.

The insensitivity of many of their variables to ranges of variance within the area is even better illustrated by the one labeled "System Style." No fewer than 17 Latin American cases are filed under the same rubric: "non-mobilizational." Independent of the accuracy of the scoring, it is quite clear that by aiming at discriminations within a universalist sample, the authors fail to provide sensitive enough indicators for the purpose of area specialists.

The accuracy of their scoring, even on crude ordinal variables, may certainly be questioned. My own research on Brazil and currently on Argentina, Chile and Mexico led me to examine closely their variable: "Interest Articulation by Associational Groups."

TABLE II

Rankings on "Interest Articulation by Associational Groups"

Significant (A)	Argentina, Brazil
Moderate (B)	Chile, Mexico, Uruguay, Venezuela
Limited (C)	Bolivia, Colombia, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Peru
Negligible (D)	Cuba, Haiti, Paraguay

Although the sensitivity has improved and the distribution has been "normalized," the attributed rank-orderings do not correspond either to my preliminary research findings or, even, to existing descriptions of interest group politics in

the area, such as Robert Alexander's *Labor Relations in Argentina, Brazil and Chile*. I am sure that those who take the trouble to read the fine print of the "Raw Characteristics Code Sheet" at the back of the volume will be similarly surprised.

Granting the fact that these categorized variables do fill one of the voids so apparent in the Yale *Handbook* by attempting to order information on political system characteristics, they scarcely contribute to meaningful analysis in the Latin American context. Worse, many are erroneous and seem to be based on impressionistic guesses. While more experienced observers might wish to correct some of these gaps, many would balk at attempting this because of the absence of descriptive information.

Peter Snow provides an interesting example of some of the inherent difficulties. Using twenty dichotomized Banks and Textor variables, he constructed a scalogram of Latin American politics. The coefficient of reproducibility is a highly respectable .95. While he notes that the resultant rank-ordering "approximates that which a specialist in Latin American politics might devise intuitively," he does not seem to consider the fact that the scale works out so well precisely because the coders acted on what one might call "patterned intuition." In the face of inadequate information, they scaled their evaluations accordingly. Hence it was assumed that if El Salvador had a less than "fully effective legislature," it would also be low on the interest articulation by associational groups. Significantly, the country for which the most descriptive material is available, Mexico, has the most scalogram errors; predictably, Cuba emerges with a low rank-ordering.¹⁰

Useful then, as existing compilations are—especially for placing Latin America in its global context—we must grapple with the methodological problems of collecting relevant, sensitive, reliable and valid data about the area, on a systematic and comparative basis, if we are to concede an active role to aggregate analysis in explaining variance in political structures and policy outcomes. We need data over a sufficient number of cases which will allow us either to enumerate actors or collectivities in analogous circumstances (e.g. percent urban, size of armed forces, number of deaths by domestic group violence per 1,000,000 population) or to group them into isomorphic categories (e.g. type of party system, degree of rationality of recruitment to public bureaucratic positions, extent of diffusion of nationalist ideology). Ideally, this comparable data should be available for an extended time span, at least for 1950 and 1960. Detailed, analogous information about a pair of countries is useful for case study analysis and the generation of research hypotheses, but for anything approaching statistically significant covariance, we must have data which reflects the full range of variation within our area sample. Latin America is something of an exception to the rule of thumb that data availability and

reliability increase with the level of the country's development. Two of the most difficulty countries to get data on are Cuba and, until recently, Uruguay.¹¹ These two relatively developed cases, perhaps, "balance out" those of Haiti and Bolivia, neither of which has conducted a census since 1950.

National census data form the bulk of the information coded and tabulated. As Table I indicates, these have been relatively frequent and are becoming increasingly regular. When supplemented by the numerous specialized censuses of industry, agriculture and housing as well as such governmental enumerations as trade statistics, cost-of-living indicators, production indices, estimates of annual economic growth, budgetary accounts, and reports of electoral commissions, the result is an impressive and rapidly increasing flow of information covering most, if not all, the twenty cases.

What is not so readily apparent is that it is becoming more and more standardized. Thanks to the Organization of American States, especially the Interamerican Statistical Institute (IASI),¹² the Economic Commission for Latin American (ECLA),¹³ the United Nations and its specialized agencies,¹⁴ demographic, economic, social, even political variables are being collected and compiled in a more analogous manner. Organs associated with the Alliance for Progress have also begun to produce important original material,¹⁵ and, of course, the "private" effort of the *Statistical Abstract for Latin America* has been making a great variety of information readily available for years. The *Election Factbook* series is an important supplementary source.

Aggregate data collection is not, however, limited to things being officially enumerated by national or international agencies. On the contrary, the most fertile field for innovation lies in the search for and construction of new, "unobtrusive measures."¹⁶

"Event-scoring" is one unobtrusive technique with a wide potentiality for quantifying theoretically relevant conditions inaccurately, inconsistently or incomparably collected by official agencies. A concept is defined, e.g. administrative stability; parameters are specified, e.g. occupants of ministerial posts and executives in public enterprises of the central government during a specified period, e.g. 1950–1960; and a scoring formula is devised. Points are allotted on the following basis: one point for each year in office; two points for each year in office beyond the term of the president who made the appointment, if of the same party; and three points for each year in office beyond the original presidential term, if incumbent party was not returned to office, divided by the total number of observations. Finally, a source of information such as the *Political Handbook of the World*, the *New York Times*, or *Facts on File* is located. The result is a composite quantitative indicator, often of the equal interval type, which can be analyzed statistically, as well as used for more concise and accurate descriptions.

A number of these event-scored comparative variables are already available. Perhaps the best known is Phillip Cutwright's measure of democratic political development.¹⁷ This has been used by a number of other authors and has become a more-or-less standard indicator, both for testing hypotheses about political development and as a proposition against which to test alternative indicators of the same concept. Political instability and violence, however, seem to have been the preferred focus of "event-scorers." The Yale *Handbook* codes the average tenure of chief executives from 1945 to 1961 and the number of deaths due to domestic group violence per million population from 1950 to 1962.¹⁸ Using a more detailed coding system,¹⁹ Rudolph J. Rummel and Raymond Tanter have collected data by types of violent political conflict for 1955–1957 and 1958–1960, respectively. Ivo and Rosalind Feierabend have devised a somewhat cruder system, with scores ranging from zero for an orderly election to six for civil war, and calculated an aggregate indicator of instability for 1955–1961.²⁰ The total number of incidences of "internal war" have been recorded by Harry Eckstein.²¹ Due to discrepancies in the period covered, scoring methods, data sources, error-estimating procedures and transformation techniques, the indicators are by no means identical for the same country, although some are highly correlated with others. All cover the Latin American universe and should be an integral part of any comprehensive area data collection.

Increasingly, however, event-scoring is likely to focus on Latin America and to use scoring systems especially designed to quantify concepts and test hypotheses suggested by the existing literature on the area. Robert Putnam's index of military intervention is without doubt the first of numerous such attempts, first to delineate carefully and self-consciously the empirical boundaries and referents of a "Latin American" concept, and second to reduce it to a statistically manipulable symbol.²² Recently, a graduate student at the University of Chicago has attempted to operationalize the empirical propositions implicit in the work of Che Guevara and Regis Debray, by deriving discrete indicators for number of guerrilla incidents, protractedness and intensity of violence, then aggregating these into a total Magnitude of Guerrilla Warfare (MGW) index and testing it for various hypothesized patterns of association.²³

The potentiality for expanding the factual basis available for comparative analysis in Latin America by event-scoring techniques seems limited only by the ingenuity and patience of researchers. Of unquantified concepts and postulated, but unconfirmed, relationships there is an unlimited supply.

A word of caution, however. Event-scoring often produces a speciously "hard" looking datum. At one extreme it comes close to index construction by t-scoring,²⁴ or by straight arithmetic calculation²⁵ of raw census data. At the other extreme it approaches the purely evaluative-subjective, "informed panel"

method used, for example by Russell Fitzgibbon in his quinquennial questionnaire.²⁶ Between the questionable accuracy of the former and the rampant ethnocentrism of the latter lies a large and unexploited "hard-soft" area in which systematic observation and trained, personal evaluation interpenetrate. In my opinion, it contains the greatest opportunities for expanding the empirical basis of regional comparative analysis and for making it relevant to existing and emerging theories of political change.

Even this does not exhaust the scope of potentially available comparative data. Content analysis, for example, could tell us a great deal about the frequency with which nationalist symbols and policy slogans appear in different Latin American polities. The Yale *Handbook's* inclusion of McClelland's Achievement Motivation Index is a rare instance of the utilization of such a data collection technique and should be imitated for a variety of other motivational and ideological conditions.²⁷

The easiest technique, of course, is classification. Banks and Textor have demonstrated, albeit somewhat unsatisfactorily, that the coding of countries according to ordinal or nominal categories permits the coverage and statistical manipulation of a wide variety of relevant concepts for which "hard" data are not available. The cost in time and effort is minimal compared to more exacting and laborious event-scoring.

Area specialists, in particular, should welcome this opportunity. As Kalman Silvert has slyly observed, "a hobby of Latin Americanists, their hands forced by the variety of their area of study, is to develop typologies within which to squeeze all twenty republics."²⁸ The literature abounds in these *ad hoc* ratings and impressionistic labels. Those few which meet at least the minimal methodological requisites should be utilized beyond the immediate, essentially descriptive ends for which they were devised. Their most obvious role is in distinguishing different types of total political systems—global data in Lazarsfeld's terminology—but they can also be important in specifying structural and sub-systemic relationships within the political system: party types and degree of competitiveness, interest group strategies and strengths, civil-military relations, electoral arrangements, etc.

The data bank I have been using has some sixty categorized variables. Twenty of these I simply borrowed from the Banks and Textor volume, although many of them should be reconceptualized and/or rescored for Latin American use along the lines suggested above. Most of the remaining ones were culled from the area studies literature (e.g. typologies of military influence by Lieuwen, Wycoff and Needler; a national-status ranking by Silvert; a political development scale and a typology of revolutionary experience by Poblete-Segundo; and degrees of developmental viability by Jaguaribe) or simply translated from formal legal provisions (e.g. constitutional restrictions

on the military, legal status of the Church, electoral representative system, literacy requirements, nature of court appointments, and degree of electoral decentralization). A typology of party systems and of corporatism in interest group relations has been devised specifically for the data bank. This data bank also has been designed to facilitate the eventual incorporation of other existing and future sets of categories. Obviously, this only begins to tap the enormous potential sources of this type of data. With the burgeoning configurative case-study type literature on Latin American party organization, recruitment and strategies, for example, it should be increasingly possible to code this material into comparable, exclusive and exhaustive, categories. All this would be greatly facilitated if the monographists would order their original observations by certain standardized indicators.²⁹

Inconsistencies persist, of course, in both availability and comparability. Some can be removed or mitigated by imaginative statistical procedures³⁰ and this, doubtless, will be an important field for methodologists and political accountants in the future. Nevertheless, the student of Latin America can hardly complain too strongly. His lot is far more enviable than that of other international area specialists, with the possible exception of those working on post-war Western Europe.

Related to the problems of availability and comparability in handling aggregate data is reliability. More than just anthropologists will, undoubtedly, sympathize with the plight of Shepard Forman who saw the *latifundio* of the Brazilian coastal village he was studying evaporate mysteriously into small parcels for the benefit of the 1965 Land Tenure Census. In a recent article, an eminent Latin Americanist has hinted darkly that two unidentified countries have been known to doctor their statistics to favor their international image.³¹ If these two were the only cases of such purposive distortion, we should consider ourselves quite fortunate.

It is generally conceded that aggregate data is "notoriously inaccurate." What is not so often conceded is that it does not have to be so reliable for many, if not for most, of the purposes to which it has been put. Johann Galtung has pointed out this methodological paradox: "data has to be more exact if the purpose is simple estimation or just accurate measurement than if the purpose is analytical."³² If one's objective is testing of analytical propositions about the behavior of collectivities, substantial error margins can be tolerated—especially under the *ceteribus paribus* assumption that errors are randomly distributed.

Even ignoring this caveat, students of Latin America are not in such a disadvantaged position. The authors of the Yale *Handbook* have explored in detail certain questions of reliability. Taking the generous estimate they suggest for the data on total population as an example, the average error margin

for the twenty Latin American republics is 6.2%. The corresponding figures for 19 countries in the North Atlantic-Western European area is 1.7%, but for 20 Asian countries it is 6.8% and for 29 African nations 12.2%.³³

More important is the fact that accuracy of data is increasing rapidly. This is in large part a side-payment from the drive toward planning and policy-analysis, imposed originally in most cases by the Alliance for Progress. Governments are increasingly concerned with basing their projections upon reliable data and with evaluating the consequences of their policies by accurate indicators. Granted, not all of this gets published and circulated publicly, but some of it dribbles down, especially under the pressures of international and interamerican agencies. Regional integration, whether real or merely contemplated, is causing member governments to pay closer attention to the comparative performance of neighbors. In Central America, for example, the Common Market has triggered a veritable revolution in the availability and accuracy of data—in this case, closely sponsored and supervised by new regional institutions. Parenthetically, the still sporadic but growing utilization of the OAS as a guarantor of electoral integrity indirectly contributes, along with other forces, to an improvement in data gathering in this traditionally sensitive field.

Another "attractive" feature of the region to the political statistician is fortuitous. The institutional overlap between ECLA and the UN agencies, on the one hand, and the OAS and associated Alliance for Progress organs, on the other, often places him in the enviable, but admittedly confusing position of having to choose between competing sets of different, but equally "hard" data. The same ambiguity is frequently reflected at the national level in disparities between census bureaus and planning agencies.

The most serious change leveled at Latin American data is that its errors are purposively and not randomly distributed.

Bruce Russett, on the basis of his extensive data-gathering experience, concludes that "in general, it may be stated that there is a clear tendency for the quality and availability of data to rise with the level of economic development of a country."³⁴ Given the range of variance in Latin America this could mean systematic distortion, especially if the poorer units were consistently "optimistic" in their estimates. Earlier, I observed that inconsistency and inaccuracy—while perhaps greater in the lesser developed and more politically restrictive countries, such as Haiti, the Dominican Republic and Bolivia—are by no means absent from such developed societies as Uruguay and Cuba. In fact, many of the less developed have been particularly receptive to the technical assistance offered by interamerican, and in the case of Central America, regional agencies. Their data may be more accurate than those of some of their more autonomous and prosperous brethren. In short, while one cannot rule

out the possible existence of systematic distortion in Latin American census data, it does seem less critical than for the world as a whole.

Event-scored data poses other reliability problems. Rudolph Rummel has made an imaginative attempt to assess their probable effect. He constructed a scale of censorship and an index of international status (based on the nature of diplomatic contact with the USA) to see whether either the suppression of news at its source or the low saliency of the country would systematically distort coverage by the *New York Times* and other standardized news sources. He found no significant correlations with his event-scored data on violence, suggesting the absence of purposive distortion.³⁵

Errors in raw data cannot usually be eliminated, but they can be compensated for. Hence, a great deal of effort must be expended in detailed examination of error margins. The experience of the IASI, ECLA and various UN agencies will be an indispensable guide. Erwin Scheuch has suggested that specious "hardness" can be removed by following this rule of thumb: "the intervals for grouped data (or when grouping them) and the thresholds of sensitivity for a measure should be greater than the possible errors in data. . . . If there is any reason to suspect error, one should sacrifice information and settle for a rougher measure."³⁶ Ted Gurr has recently proposed an even more elaborate system of error compensation and has recalculated such key variables as the magnitude and change in governmental expenditure levels, indices of group discrimination and separatism, trade union membership, non-agricultural employment, and military participation ratios.³⁷

The uneven reliability of Latin American aggregate data, particularly census enumerations, is a problem which should make students wary, but it is not an insurmountable obstacle. At the usual level of generalization and with the appropriate techniques of analysis, this approach to comparison is tolerant of rather wide error margins. Unreliability, then, is no excuse for impeding the construction of new indicators and the testing of propositions about Latin American politics.

A problem associated with reliability is that of quality. This in turn raises somewhat related points. Is the Latin American data "representative"? Do the ranges of variance encountered in Latin America correspond to those of the political universe as a whole? Is the Latin American data technically well distributed? Is it excessively skewed or peaked, is it unimodal, making some forms of quantitative analysis difficult to interpret?

The answer to the question of variance is, with qualifications, yes. For the sort of question which aggregate data analysis is best equipped to answer, e.g. propositions about the "ecological" impact of variables such as per capita income, social structure, economic productivity, and land tenure, Latin America exhibits ranges of variance more or less equivalent to that of a "world sample."

There is, however, a certain underrepresentation of extreme cases (which is, statistically speaking, not a bad thing). At the one end, the scores approach those of the moderately developed Western European nations; at the other, they fall short of the most impoverished and structurally undifferentiated cases of Africa, but in between, they are representative, at least in terms of independent variables. Where they are quite noticeably skewed is in the dependent and intervening variables concerning political structures and public policies. The long formal political independence and strong Western cultural tradition of the Latin American states makes them unrepresentative, biased toward indicators of political development based on that political and cultural tradition. In short, for some propositions, the analysis of the Latin American case can legitimately claim to contribute to an understanding of the societal universe as a whole; for many others, especially those with political dependent variables, students of the region are perforce limited to testing inference within their restricted subset of nations.

To the question regarding the technical distributional characteristics of the data, no single answer can be given. Of course, for nominally and ordinally categorized information the distribution depends on the classification system devised. As already indicated, *supra*, universalist, schemes such as Banks' and Textor's, are deficient in this regard. For basic census-type enumerations the distribution is often acceptable. As the following table indicates, the standard indicators of urbanization, industrialization, tertiarization, economic growth, literacy, armed forces participation ratio, public expenditures and source of government revenue, are neither too skewed nor too peaked. However, variables such as the percentage of *Amerindios* and, especially, deaths by domestic group violence are very irregularly distributed. Clearly, these should either be "normalized" mathematically³⁸ or, better, transformed into a rank ordering or other ordinal scale before utilization. Again, some further methodical methodological work is needed *prior* to the stage of testing for patterns of association.

The most serious challenge one can throw at aggregate data is irrelevance, that the things they measure are not valid indicators of the concepts they purport to represent. So, for example, when political participation is operationalized by enumerating the proportions of those who vote in general elections, objections are raised about the significance of the electoral process, about the importance of a wide variety of less public forms of representation, and, in general, about the manifest "democratic bias" of that indicator.

We must simply admit that for a large number of variables there exist only proximate indicators.

Analysts of aggregate data are often aware of tenuous validity. Unfortunately, they limit their concern to a few cautionary remarks and fail to employ what seems to me an obvious mitigating strategy, that of operationalizing

TABLE III
Parametric Characteristics of Selected Variables: Latin America ca 1960

Label	Mean	Standard Deviation	Range		Skewness
			Maximum	Minimum	
Urbanization	41.96	17.40	81.8	12.4	.646
Per Capita Income	360.10	169.32	799.	122.	.954
Agriculture as a % of GNP	29.01	11.25	50.4	7.2	.048
Occupation in Manufacturing	12.57	5.09	25.2	2.7	.505
Middlesector Population	15.18	8.99	35.9	3.0	.750
Literacy	63.15	18.93	91.4	20.0	.435
Electorate	25.48	9.54	45.1	13.6	.388
Armed Forces per 1000	.36	.19	.69	.10	.387
Public Expenditures as % of GNP	22.30	9.01	46.7	7.3	.809
Taxes on Exports/Imports as % of Govt Revenue	40.27	18.60	79.6	8.9	.066
Amer. Indian Population	11.35	19.92	63.0	0.0	1.70
Per Capita US Private Investment	59.53	101.26	436.64	3.85	2.87
Newspaper Circulation per cap.	78.50	58.57	260.	11.	1.55
Trade Union Members per cap.	5.78	6.41	22.5	0.1	1.68
Public Employees per 1000	30.22	78.35	82.5	7.1	1.50
Govt Budget in \$ per cap.	49.42	47.00	183.70	6.50	1.89
Deaths by Domestic Group Violence	230.02	647.81	2900.0	0.3	3.77

Source: P. C. Schmitter, "Latin American Aggregate Data Bank," University of Chicago, 1968.

their conceptual schemes with alternative indicators and presenting *all* the evidence to the reader. At times one suspects that authors have indeed run alternative indicators, but conveniently reported only those which fit best. This procedure is deceptive, to put it politely. "Political statisticians" have an obligation to report both inconclusive as well as conclusive indicators, insignificant as well as significant correlations. Such "non-findings" are also part of the process of cumulative theory-building.

One source of our difficulties is that governments do not always collect information on the sorts of variables most relevant to political analysis. Or, if they do, they often do not make it public.³⁹ Perhaps, in the long run, domestic pressures for more comprehensive social and economic indicators will also spillover into the overtly political sphere. However, in the United States, where policy-makers have a supposed penchant for facts many of these indicators are deficient.⁴⁰ It is rather unrealistic to expect Latin American countries to proceed even faster.

One way out is to devote more effort to event-scoring and categorization. These data-gathering techniques are specifically intended to provide *valid* in-

formation. They are chosen and ordered with a precisely delimited concept in view, normally one for which reliable, interval-type, information is not available. Many of the challenges of irrelevance and triviality can be parried by shifting to this form of data.

Phillip Cutwright has pointed out another operational strategy. Where no single datum is valid or reliable for measuring a complex variable, such as economic development or mass communications exposure, he constructs a multivariate indicator by standard scoring a set of interrelated, but significantly independent ones (e.g. energy consumption, steel consumption, income in US dollars and number of motor vehicles per capita) and sums them. Presumably, the composite variable is not only a more valid representative of the empirical complexity of what is being measured, but is also more reliable in that inaccuracies in a single variable, such as those in per capita income figures due to over or undervalued exchange rates, are mitigated or compensated for by others. In other words, a regime may distort its image some of the time on some variables, but not all of the time on all variables.

Another possible recourse is the use of factor analysis and the corresponding factor scores for each unit as indicators of complex political and social relationships or as guides to more sensitive taxonomic systems.

The compiler of a data bank has an understandable tendency to include all available, comparable information because it is there. Even when categorized information is being handled, there is a temptation, which Banks and Textor acknowledge, to code and punch everything currently fashionable in comparative politics. Nevertheless, the "political accountant" must continuously remind himself that the ultimate test of a set of aggregate bits of information is validity. Does it represent or indicate some concept or type which is theoretically relevant to explaining the political behavior of the unit?

Comparative analysis does *not* involve the simultaneous examination of "similar" events, proportions or relations. It supposes a prior ordering of these acts—which obviously differ on a variety of dimensions—into analogous classes or categories. The proportion living in cities of over ten thousand population is not simply juxtaposed to the proportion casting votes in the last election. The former is taken to indicate a concept, urbanization, whose pattern of variance is compared with that of another concept, political (or, more modestly, electoral) participation. Only by specifying the indicated concept can we understand the qualities that those living in cities and voting are supposed to be sharing in Argentina and Haiti and the social logic of their interrelationship. And, of course, only then can we assess their validity.

The 650 variables in my Latin American Aggregate Data Bank are divided into three broad conceptual categories: I. Ecological Conditions (250); II. Political System Characteristics (241); III. Policy Outcomes (159). These

are in turn divided into sub-classifications: I. Demographic, ethnic, linguistic, religious (14); urbanization (13), size (5), growth rate (7), internal economic (27), external economic (23), social structural (26), social developmental (42), communications (21), assorted ecological (47), multivariate indicators (25); II. General political characteristics (21), participation (26), military (65), multivariate potential influence (20), competitiveness (37), bureaucratic (31), formal-constitutional (10), executive-legislative (31); III. Violence-instability (29), expenditures-payoffs (20), taxes-investment patterns (16), repression (4), reform (14), international exchanges (59). In the absence of an accepted operational model of the political process, any such attempt at comprehensive classification is bound to be tentative. The filling-in of existing gaps and the rectification of inaccurate and insensitive variables are not likely to advance very rapidly—unless they are put to analytical use.

METHODOLOGICAL PROBLEMS OF DATA ANALYSIS

The pitfalls of analyzing aggregate data are even more substantial than those of collecting it. Fortunately, they have been extensively debated. My understanding of this literature leads me to be cautiously optimistic about the possible payoff for this approach. While it will never succeed alone in predicting, much less explaining, all the variation in outcomes which interests students of Latin American comparative politics, it can be very helpful, even indispensable, in establishing the parameters within which idiosyncratic variables, random occurrences and creative individual choice have an influence. Despite its gloss of technological sophistication, there is nothing intrinsically dehumanizing about this style of analysis. Used circumspectly, one of its principal functions is precisely to delimit the ranges of alternatives open to political actors. Only systematic, qualitative investigation of specific cases can help elucidate the moral and material motives of actors choosing between these alternatives.

Nevertheless, there is a considerable amount of unreflective enthusiasm for the approach and subsequent danger that practitioners of it, in their complete absorption in statistical technique, will either devote detailed and mindless attention to trivial subjects or advance grandiose inferences upon spurious findings. The two major methodological issues concerning this strategy of inquiry are related to these pitfalls: the question of inference about individual behavior and the question of inference about causality.

This is no place to review the extensive discussion concerning the "ecological fallacy."⁴¹ Simply stated, a number of methodologists, beginning with Robinson's classical article, have warned against interpreting relationships observed between aggregate or collective properties as a basis for inferring a like relationship between these same variables at the level of individuals. Hence, a strong ecological correlation between an indicator of the size of a country's

middle class and its voting population (+.728 for 1950 according to the Bank) does not necessarily imply that middle-class individuals are more likely to have voted—no more than an even stronger correlation with per capita telephone ownership (+.915) implies that they own almost all the phones, or worse, that they almost all own phones.⁴² Earlier work on electoral behavior in the United States rather consistently committed this error by inferring such individual cell frequencies from marginal frequencies established by collective properties. As a result of this serious critique—and the rapid diffusion of survey research—aggregate data analysis declined in this country.⁴³

It did not decline in Europe, however, especially in France. Scheuch points out that the root cause of the divergence was and still is not exclusively methodological:

Although it is obvious, it is seldom realized that a preference for either individual or aggregate measurements is related to some well-established inclinations for differing strategies in theory construction. Macro-theories are easier to connect with aggregate data, while micro-theories and reductionistic explanations of systems are apparently more easily related to individual measurements. Both preference for different strategies in theory construction and the availability of either individual measurements or aggregate data often appear to be connected with the actual degree of centralization and formalization of the social systems and societies in which and with which the researcher deals. In government and political sociology, preference for macro-theories and aggregate data is specifically connected to the smaller relevance of individual citizens for political processes, and to the greater relevance of planning done and of decisions made on the level of aggregates. This characterization fits many Continental (and Latin American—PCS) societies, just as the reverse is more characteristic of the United States as a polity.⁴⁴

The current resurgence of interest in aggregate analysis is based on an acceptance of the limitations imposed by the methodological critique and a recognition of the desirability of adopting a macro-theoretical perspective. Correlations between the collective properties of areal units are not merely an inferior alternative to individual correlations. On the contrary, “(they) may be of interest in themselves; the kinds of questions that can be answered by a study of ecological correlations are sometimes of direct concern to social scientists.”⁴⁵ Such questions about social and political relations at the territorial or whole system level are of obvious (if not exclusive) concern to any student of Latin American politics.

Actually, I would argue that these students are more commonly guilty of the inverse fallacy—“the individualistic fallacy.” Numerous texts and monographs on the area infer from the systematic (or worse, impressionistic) observation of the behavior of individuals the characteristics of whole systems.⁴⁶ Thus we are treated to the “fact” that Costa Rica is democratic because its citi-

zens are freedom-loving, that Mexicans are obsessed with *machismo* and, therefore, have and deserve an authoritarian polity. As Scheuch points out, "what is ignored is that one may have a democratic system with few 'democratic' personalities and various types of authoritarian systems with high percentages of democratic personalities. Democracy is the term of a political system and a political system is obviously not just the aggregate of the individuals comprising it."⁴⁷

Both these common methodological procedures are related to a disparity between the units of observation and the units of inference. In the former case, the unit of inference is small; in the latter, it is larger. Aggregate data analysis then is neither inherently inferior nor superior to that based on individual enumerations. Both are logically restricted to operating at different levels of association and inference.

Yet a third fallacy has recently been uncovered—"the universalistic fallacy." This special case of the original ecological fallacy should be of particular interest to area specialists. It is committed when one infers from correlations between all units (e.g. the universe of the world's polities), associations of the same variables between units within particular subsets or regions. Again within cell variance (e.g. relations specific to a geographic or cultural context) is being inferred from marginal frequencies across all polities. Hence, if one finds a significant association between, say, the size of the industrial working force and the percent voting for radical left parties at the global level, this is no assurance it is also valid for the Latin American subset. To the extent that area specialists tend to be convinced that certain contextual variables (e.g. the area's pre-Renaissance, Catholic, European cultural heritage, its isolated geographic position, its dependent neo-colonial economies and its premature urban development) make a unique contribution to understanding political structures and policy outcomes, they should be interested in exploring this potential fallacy. As its "discoverer," Hayward Alker Jr. points out, "regional specialists . . . should make continual distinctions between local events and worldwide patterns. . . . There is need to use and develop more adequate formal models of the various ways in which particular regional contexts influence configurations."⁴⁸ There is, then, a real danger that contextual patterns characteristic of Latin America will be overlooked or subsumed in the drive toward analysis of universalistic samples. Regional data banks paralleling and amplifying universalistic ones can contribute to avoiding this fallacy. However, even within the Latin American region we must be sensitive to possible analyses of covariance across such sub-regional samples as Central America, the Caribbean, the Andean regions, and the Cuenca de la Plata.

One might expect that having, hopefully, avoided the pitfalls and ac-

cepted the limitations discussed above, the researcher should now be ready to plunge into the data itself. The remaining problems seem essentially "technical," e.g. the choice between the instruments of statistical manipulation appropriate to the type and quality of the data. For much of contemporary political analysis this would be the case. Its objectives are (1) to describe empirically existing ranges of variance; (2) to discover significant correlations between variables (or better, to reject null hypotheses that such ranges vary independently of each other); (3) to predict the probability that change in one variable will be associated with change of a determined magnitude in other variables. In short, a major objective of aggregate data analysis is to identify what varies together, especially among ecological conditions, political structures and policy outcomes.

This search for probabilistic association rather carefully avoids the problem of causality. An elaborate and semi-precious vocabulary has evolved which eschews expressions such as "causes," "creates," "determines," "invariably follows from" in favor of "accompanies," "is usually associated with," "reciprocally influences," "are stochastically covariant." Desirable as this caution may be and respectful as we may be of the scepticism of positivists, functionalists and behaviorists, we have to admit it leaves our curiosity unsatisfied. *Why* are the phenomena related and how?—is the inevitable question and its leads us into the controversial area of causal inference.

Again the methodological literature is enormous and, this time, much less conclusive.⁴⁹ I will attempt to summarize my understanding of the issues and their particular relevance for comparative political analysis.

(1) The key element which distinguishes causality from association or conjunction is the quality "producing." In the latter case, the variables predict or estimate each other equally well. To use Blalock's example, knowing the weather in Chicago may be of considerable help in predicting the weather in Milwaukee or vice versa. It does not, however, imply that one's weather is producing—causing—the other. A causal relation implies assymetry, a temporal lag and/or a directional component through which one variable effects a change in another but not vice versa. Even in the case of "feedback" or reciprocal interaction, some temporal or directional assymetry occurs before the role of producing agent is reversed.

(2) The Humean postulate of "constant conjunction" is not likely to be satisfied in the scrutiny of most social and political relationships. *X and only X is necessary and sufficient to produce a change in Y* implies an isolated and closed system of explanation not often encountered in empirical reality. Social causality proceeds more modestly by allowing for residual error (that is to say, the influence of conditions not explicitly included in the explanatory system) and by aiming at a probabilistic estimate of the change in the *mean* value of Y

for different values of X. As Blalock observes, these error terms can be interpreted in a variety of ways—as inaccuracies of measurement, “chance” or random processes, or outside variables not brought into the explanatory scheme.⁵⁰

(3) Causal inferences are theoretical, i.e. based on a stipulated set of conceptualized conditions, and hypothetical, i.e. working assumptions never definitively verified and constantly subject to falsification. Practitioners readily admit that this involves certain very strong assumptions about unspecified, unincorporated variables (namely that their impact is negligible, random or invariate) and, hence, that alternative explanatory schemes or models may be equally valid. “In other words, there is nothing absolute about any particular model, nor is it true that if two models make use of different variables, either one or the other must in some sense be “wrong”.”⁵¹

A variety of techniques for inferring causality have been proposed. All, however, utilize the same basic strategy. A zero-order association between two variables, no matter how significantly predictive or how convincingly time-lagged,⁵² is “doubted” as necessary and sufficient. One of the variables is labelled “dependent” or caused, the other is considered hypothetically as “independent” or the causal agent. An error term is introduced. Since the problem is non-experimental and values cannot be manipulated deliberately, control variables are introduced either by successive 2x2 crosstabulations, breakdowns into subgroupings and analyses of covariance, or by partial correlational techniques. The object is to test for spurious [i.e. due to the effect of another independent variable(s)] or masked [i.e. due to the effect of another intervening variable(s)] relationships. After the introduction of control variables or the testing of alternative models, the best statistical fit provides the grounds for inferring causality.

These grounds are still far from absolutely convincing. Perhaps the most important contribution is that in the process of elaborating the relationship in such a multivariate manner, the researcher discards as inconsistent or insignificant a vast number of other theoretical models, some of which may have been enshrined in the conventional wisdom of his field. Substantial as this negative pay-off may be, can we expect more from causal inference?

Johann Galtung has observed that

at least three elements of a non-statistical nature will usually enter in any discussion of causality: the time-relation between the variables, the theory needed to “see a mechanism working” between the variables and the subjective feeling that the relation has been tested for a *sufficient number of relevant variables* in a sufficient number of combinations. This additional kind of insight required is not better or poorer, only different; statistics alone is not sufficient.⁵³

After statistical analysis has established the basic relationship, Latin

Americanists may be in a privileged position to provide those “supra-statistical” insights. The time perspective is relatively good—with both quantitative and qualitative data extending over one hundred years of formal political independence. The accumulation of a large descriptive monographic literature often permits us to observe mechanisms at work. Hopefully, aggregate data analysis will help systematize the collection of this detailed information in the future by pointing out clusters of independent and dependent variables and, equally important, by investigating deviant cases. Lastly, the very rationale underlying Latin American area studies assumes that a substantial number of variables, especially cultural, ideological, historical, even political structural ones, can be considered not to vary significantly across the twenty units.

CONCLUSION

Recognizing that the collection of various forms of aggregate data and their statistical manipulation are not going to resolve all our dilemmas single-handedly, let us summarize the benefits to be gained by using this data.

(1) Aggregate data analysis will encourage—if not force—students to make explicit the variables they postulate as related to a specified set of outcomes.

(2) It will require them to make explicit in the form of a formula or working hypothesis exactly how independent variables affect the dependent ones.

(3) It will make us think comparatively, by rooting generalizations in the total range of variance within the area, rather than by limiting attention to individual, often self-serving cases or, worse, to isolated anecdotal examples.

(4) Explicitness and comparability will spur the examination and probable falsification of many customarily accepted truths of the field and shift its focus from description to causal analysis. It will also serve to pinpoint deviant cases and the impact of idiosyncratic and random variables.

(5) Aggregate data analysis will emphasize the importance of making the distinction between concepts and their empirical indicators and will encourage more systematic and imaginative data collection where such indicators of key conceptual variables are missing or inadequate. Personally, I hope (but do not expect) that it will lead to less emphasis on such easily quantifiable items as elections and incidences of violence and more on patterns of public policy and their impact.

(6) The major analytical payoff of all this tedious effort, I submit, is not likely to be highly determinate, closed systems, which explain all the variance in political outcomes, but open, probabilistic approximations which establish the parameters of political choice imposed and tolerated by different levels of

development, types of participation, degrees of institutional complexity and past policies. With this matrix of loosely determinate interdependencies, we will be in a better position to interpret and evaluate the actions, motives and consequences involved in specific decisions. In short, we will be better able to see them in their ecological and political structural context and not purely from some vantage point we ourselves have selected.

(7) Eventually, this style of analysis should cause us to examine critically the intellectual basis of area specialization itself. Latin Americanists have never been very explicit about this, beyond the obvious appeal to convenience and convention.

Is there a discrete, measurable "contextual effect" which contributes uniquely to explain Latin American ranges of variance? If so, we should be more specific about its historical, geographic, cultural, ideational components. If not, we should simply pack our bags and move upstairs to the comparative politics suite.

Of course, independent of whether scholars with training and experience in the area take advantage of it, this quantitative-manipulative approach to comparison will be used increasingly. My special plea is for Latin Americanists to "get in on the action" in a concerted and systematic way so that their special competence is enlisted to enhance both the quality of the data (to expand its scope, to improve its sensitivity and to assess its reliability) and the quality of the analysis (to make it focus on issues relevant to an understanding of the area and not simply those that are generally fashionable in the discipline or with certain US public officials).

The most obvious way to promote this is to enlist the skills of social scientists active in Latin America in the collective elaboration of an area Data Bank.

My own preliminary experience strongly convinces me that such an effort must be collaborative and multi-disciplinary. Anyone can throw a few variables together and call them an aggregate data bank. This may even suffice for the limited research problems he has in mind, but the creation of a flexible and comprehensive compilation of utility to a wide variety of researchers demands skills and a breadth of knowledge beyond the grasp of most individuals. It is not simply a function of limited time and patience, but of the need for an encyclopedic awareness of the availability of data and of the probable margins of error in different sources, a wide grasp of theory in different fields to guide the selection of concepts and operationalization of indicators, and a number of intersubjective checks on the reliability of evaluative information. In addition it requires a variety of statistical and mathematical skills to compile, categorize, scale, standard-score, weigh and otherwise transform the raw data, and a knowledge of data-processing and retrieval techniques to make the Bank more manipulable and accessible. Lastly, a quotient of imagination is essential to see

gaps and devise ingenious indicators for crucially relevant variables. I have no doubt that these talents are collectively available among those working on Latin America from this country and within the area itself. Their combination could go a long way to providing a working instrument, a set of social, economic and political accounts, which would put students of Latin America in a privileged position within the expanding field of cross-national social science research.

Such a twenty-nation bank, while it may be a prerequisite for systematic exploitation of aggregate data analysis, is not an end in itself. It opens up new strategies in several directions. One macro-strategy would be to incorporate the twenty some US, British, French and Dutch ex-colonies and dependencies in the Caribbean. Another might be to go outside the area for countries which exhibit certain parametric similarities, e.g., Spain, Portugal, Ireland, Philippines, Greece, Turkey, Tunisia, Lebanon, Malta; even Belgium, France and Italy. A bit more daring might be analyses of covariance with the countries of Eastern Europe and the Balkans in the pre-war period. I am particularly intrigued with a "paired" strategy whereby for each Latin American unit one would assign the country outside the region most like it, in such general dimensions as size, level of development, social mobilization, and class and ethnic stratification. Then we might really see if there is anything in this "Latin American-ness" presumption of ours.

Micro-strategies are already being exploited. This involves the breaking down of national units into their constituent political or territorial units: states, provinces, municipalities or cantons, the duplication of as many variables as possible from the regional sample and the replication of cross-national analyses performed on the Bank as a whole.

A modest first step in the compilation of the basic twenty-unit bank would be the designation of a center (the International Data Library and Reference Service in Berkeley and FLACSO in Santiago seem the logical choices) to which area scholars would agree to send raw aggregated data and/or derived indices. Also helpful would be the transmission of error estimates for existing data. The centers, in turn, would agree to incorporate the corrections and new data and make the Bank available at moderate cost to all interested scholars.

A more ambitious step forward would be the convening of a symposium where researchers and authorities actively engaged in primary data collection (for example, national census bureaus, the UN, ECLA, ILPES, the OAS, and IASI) could discuss the quality and availability of existing data. LASA might sponsor regular panels or working sessions at which Latin American scholars from all the social science disciplines could exchange critiques of accomplished work and thoughts on future priorities. Hopefully, the end product of this could be—in addition to a machine-manipulable bank—a volume similar to the Yale *Handbook* for twenty units.

In essence, I suggest that, in keeping with the political culture of the region that preoccupies us, we meet this latest methodological challenge by practicing prudent cooptation rather than intransigent opposition.

NOTES

1. C.F. John D. Martz, "The Place of Latin America in the Study of Comparative Politics," *Journal of Politics*, 28, 1 (February 1966), pp. 57–80; Merle Kling, "The State of Research on Latin America: Political Science" in C. Wagley (ed.), *Social Science Research on Latin America* (New York: Columbia University Press, 1964), pp. 168–207.
2. E.g. S. N. Eisenstadt, *Modernization: Protest and Change* (Englewood Cliffs: Prentice-Hall, 1966), esp. pp. 83–97; David E. Apter, *The Politics of Modernization* (Chicago: University of Chicago Press, 1965); C. E. Black, *The Dynamics of Modernization* (New York: Harper Torchbooks, 1966); Irving Louis Horowitz, *Three Worlds of Development* (New York: Oxford University Press, 1966); Dankwart Rustow, *A World of Nations* (Washington: The Brookings Institution, 1967); C. W. Anderson et al., *Issues of Political Development* (Englewood Cliffs: Prentice-Hall, 1967); Andrew M. Scott et al., *Simulation and National Development* (New York: Wiley, 1966).
3. E.g. John H. Kautsky (ed.), *Political Change in Underdeveloped Countries* (New York: Wiley, 1962); R. Macridis and B. Brown (eds.), *Comparative Politics: Notes and Readings* (Homewood, Ill.: Dorsey Press, 1961)—succeeding editions have increased their Latin American content—; Harvey B. Keshchull (ed.), *Politics in Transitional Societies* (New York: Appleton-Century-Crofts, 1968); J. L. Finkle and R. W. Gable (eds.), *Political Development and Social Change* (New York: Wiley, 1966); Wilson C. McWilliams (ed.), *Garrisons and Government* (San Francisco: Chandler, 1967); Karl W. Deutsch and William J. Foltz (eds.), *Nation-Building* (New York: Atherton, 1963); Louis Hartz (ed.), *The Founding of New Societies* (New York: Harcourt, Brace and World, 1964).
4. However, of three recent introductory texts: Alexander Edelmann, *Latin American Government and Politics* (Homewood, Ill.: Dorsey Press, 1965); B. G. Burnett and K. F. Johnson (eds.), *Political Forces in Latin America* (Belmont, California: Wadsworth, 1968); Martin C. Needler (ed.), *Political Systems of Latin America* (Princeton: Van Nostrand, 1964), only the latter shows even a verbal acquaintance with them.
5. (New Haven: Yale University Press, 1964) and (Cambridge: M.I.T. Press, 1963). The former is a collection of ratio and interval data on some 75 variables, mostly from national censuses and United Nations sources. The latter contains 57 categorized pieces of nominally or ordinally coded information, gathered by the authors and, reportedly, checked by academic area specialists in the Boston region. Somewhat analogous handbooks which focus more specifically on economic development variables are Norton Ginsberg, *Atlas of Economic Development* (Chicago: University of Chicago Press, 1961) and Irma Adelman and Cynthia Taft Morris, *Society, Politics and Economic Development* (Baltimore: The John Hopkins Press, 1967). Again, the former uses metric and the latter categorized data.
6. The most important of these is the Dimensionality of Nations Project at Northwestern University. For a description of it, see Rudolph J. Rummel, "The Dimensionality of Nations Project" in R. Merritt and S. Rokkan (eds.), *Comparing Nations* (New Haven: Yale University Press, 1966), pp. 109–130. Professors Ivo and Rosalind Feierabend at San Diego State College have also put together an important compilation of data on violence.
7. In this regard, see especially the articles in R. Merritt and S. Rokkan (eds.), *ibid*; the ex-

- cellent review of the Yale *Handbook* and the *Cross-Polity Survey* by Michael Hass, "Aggregate Analysis," *World Politics*, XIV, 1 (October 1966), pp. 106–121 and Ralph H. Retzlaff, "The Use of Aggregate Data in Comparative Political Analysis," *Journal of Politics*, 27 (November 1967), pp. 797–817.
8. C.f. Ralph Retzlaff, *op. cit.*, pp. 799–800.
 9. *Op. cit.*, p. 1.
 10. "A Scalogram Analysis of Political Development," *The American Behavioral Scientist*, IX, 7 (March 1966), pp. 33–36.
 11. Since the census of 1961, this gap in the information has slowly been filled. See, for example, A. Solari, *Uruguay en Cifras* (Montevideo: Universidad de la Republica, 1966). The scarcity on Cuba has been somewhat alleviated by Dudley Seers (ed.), *Cuba: The Economic and Social Revolution* (Chapel Hill: University of North Carolina Press, 1964).
 12. The major source is *America en Cifras* (Washington, D.C.: Unión Panamericana, Instituto Interamericano de Estadística, 1960–1968). IASI's *Boletín Estadístico* is a valuable interim source and regularly publishes cost-of-living information. Special compilations are also available, e.g., *Characteristics of the Demographic Structure of the American Countries* (1964); *La Situación de la Vivienda en América* (1962); *La Estructura Agropecuaria de la Naciones Americanas* (1957); *Estudio Social de América Latina* (1961, 1962, 1964). Also the Pan American Health Organization's *Health Conditions in the Americas* (1966).
 13. The two principal sources are the yearly *Economic Survey of Latin America* (1949–) and the quarterly *Statistical Bulletin for Latin America* (1964–). *The Economic Bulletin for Latin America* (1956–) often carries articles with important original data. ECLA special studies are another indispensable source, e.g. *The Process of Industrial Development in Latin America* (1966); *External Financing in Latin America* (1965); *The Economic Development of Latin America in the Post-War Period* (1964).
 14. For example UN, *Demographic Yearbook*, *World Economic Survey* (1948–), *Yearbook of International Trade Statistics* (1950–); *World Energy Supplies* (1929–1950, 1951–); IMF, *International Financial Statistics* (1948–); *Balance of Payments Yearbook* (1940–); IBRD, *Annual Report* (1946–); FAO, *Production Yearbook* (1958–); ILO, *Yearbook of Labour Statistics*; UNESCO, *Current School Enrollment Statistics* (1948–). Also UN, *Compendium of Social Statistics: 1963. Growth of World Industry 1938–1961*.
 15. The *Annual Reports of the Social Progress Trust Fund* (1961–) are good, if somewhat incomplete, sources of information. The latest (1967) has detailed important material on higher education. IADB special studies are very useful, e.g. the recent *El Desarrollo Agrícola de América Latina en Próxima Década* (1967). Alliance for Progress, CIAP documents and country summaries are also full of information, but difficult to obtain.
 16. Cf. Eugene J. Webb *et al.*, *Unobtrusive Measures: Non Reactive Research in the Social Sciences* (Chicago: Rand McNally, 1966).
 17. "National Political Development: Measurement and Analysis," *American Sociological Review*, Vol. 28, no. 2 (April 1963), pp. 253–264. Cutwright's weighting procedures have changed. See also his "Political Structure, Economic Development and National Social Security Programs," *American Journal of Sociology* (March 1965), pp. 537–550. In the latter article, he devises a scale of experience with social security programs with information on this crucial policy outcome for 76 countries including those of Latin America.
 18. *Op. cit.*, pp. 97–104.

Latin American Research Review

19. "The Dimensions of Conflict within and between Nations," *General Systems Yearbook*, 8 (1963), pp. 1–50; "Dimensions of Conflict Behavior within and between Nations, 1958–1960," *Journal of Conflict Resolution*, X, 1 (March 1966), pp. 41–64.
20. "Aggressive Behavior within Polities, 1948–1962: A Cross-National Study, *Journal of Conflict Resolution*, X, 3 (September 1966), pp. 249–271.
21. *Internal War: The Problem of Anticipation*, a report submitted to the Research Group in Psychology and the Social Sciences, Smithsonian Institution, Washington, D.C., 1962. For a ranking which differentiates only the "unequivocal" incidences of internal war, see Merle Kling, "Violence and Politics in Latin America," *The Sociological Review*, Monograph no. 11 (February 1967), p. 123. For more detailed data on violence, see Ted Gurr with Charles Ruttenberg, "The Conditions of Civil Violence. First Tests of a Causal Model" (Princeton: Center for International Studies, *Research Monograph* No. 28, April 1967).
22. "Toward Explaining Military Intervention in Latin American Politics," *World Politics*, XX, 1 (October 1967), pp. 83–110.
23. James Preminger, "Guerilla Movements in Latin America: A Comparative and Theoretical Analysis," MA paper, University of Chicago, August 1968.
24. Cf. Cutwright's indices of economic development and communications development in the above-mentioned articles, footnote 17.
25. Cf. Peter Heinz, "Paradigma del desarrollo latinoamericano"; Ruben D. Katzman, "Dependency and Absorption of Social Tensions in Underdeveloped Countries," Survey Research Center, University of California (Berkeley), mimeo, 1968; Alao S. Passos, "Developmental Tension and Political Instability: Testing some Hypotheses concerning Latin America," *Journal of Peace Research*, No. 1 (1963), pp. 70–85.
26. For the most recent version, see "Measuring Democratic Change in Latin America," *Journal of Politics*, 29 (February 1967), pp. 129–166.
27. *Op. cit.*, pp. 192–195. Also David C. McClelland, *The Achieving Society* (Princeton: Van Nostrand, 1961).
28. "A Proposed Framework for Latin American Politics," in J. D. Martz (ed.), *The Dynamics of Change in Latin American Politics* (Englewood Cliffs: Prentice Hall, 1965), p. 9.
29. For an excellent suggested coding scheme, see Kenneth Janda, *Information Retrieval* (Indianapolis: Bobbs-Merrill, 1968), pp. 118–164. See also Sidney Verba, "Some Dilemmas in Comparative Research," *World Politics*, XX, 1 (October 1967), pp. 111–127 for a discussion of the encounter between configurative and comparative analysis and a plea for a "disciplined configurative approach"; and Simon Schwartzman and Manuel Mora y Araujo, "The Images of International Stratification in Latin America," *Journal of Peace Research*, 3 (1966), pp. 225–243.
30. Cf. the work of the International Population and Urban Research Center at the University of California (Berkeley) in the fields of birth and mortality rates. O. A. Collver, *Birth Rates in Latin America: New Estimates of Historical Trends and Fluctuations* (1966) and E. E. Arriaga, *New Life Tables for Latin American Populations in the Nineteenth and Twentieth Centuries* (1968).
31. Sometime their enthusiasm trips them up. Paraguay, for example, reported more children in secondary school in 1960 than there were people in the relevant age category in the 1962 general census.
32. "Some Aspects of Comparative Research," *Polls*, II, 3 (Spring 1967), p. 3.

33. *Op. cit.*, pp. 15–21.
34. *World Handbook of Political and Social Indicators*, *op. cit.*, p. 3.
35. "The Dimensions of Conflict Behavior Within and Between Nations," *op. cit.*
36. "Cross-National Comparisons using Aggregate Data: Some Substantive and Methodological Problems" in R. Merritt and S. Rokkan (eds.), *op. cit.*, p. 140.
37. Ted Gurr, "New Error-Compensated Measures for Comparing Nations: Some Correlates of Civil Violence," Princeton University, Center of International Studies, *Research Monograph*, No. 25, May 1966.
38. For example, see "Toward a Theory of Political Instability in Latin America," *Journal of Peace Research*, 3 (1967), pp. 209–227. For an intelligent and practical discussion of the consequences of such transformations, see H. Alker *et al.*, *World Handbook of Political and Social Indicators*, *op. cit.*, pp. 311–313.
39. An interesting case in point is the total number of government employees. Fragmentary (and wildly disparate) data are available, but for most Latin American countries this seems to be a state secret. For an independent estimate in 1955, see "Government Income and Expenditure: 1947–1954," ECLA *Economic Survey of Latin America—1955*, pp. 111–176. Also see *America en Cifras—1963*, Vol. VI, "Situación Política y Administrativa" and the *Social Progress Trust Fund Report* (1965) for incomplete and contradictory, but more recent, figures.
40. Cf. Albert D. Biderman, "Social Indicators and Goals," in Raymond A. Bauer (ed.), *Social Indicators* (Cambridge: MIT Press, 1966), pp. 68–153.
41. For the original discussion of this methodological fallacy, see W. S. Robinson, "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, XV (1950), pp. 351–357. An excellent restatement which incorporates the comments of Menzel, Goodman, Arrow, Duncan and Davis is Raymond Boudon's "Propriétés individuelles et propriétés collectives: un problème d'analyse écologique," *Revue Française de Sociologie*, IV, 3 (July–Sept. 1963), pp. 275–299.
42. Robinson (*op. cit.*) also demonstrates that the strength of association varies directly with the size of the unit over which the same data is aggregated. The greater the amount of internal (within cell) variance that is squashed the higher the association. In other words, correlations between data on whole political systems, e.g. Brazil, are likely to be artificially inflated. See also E. Scheuch, *op. cit.*, pp. 149–150.
43. Cf. Austin Ranney, "The Utility and Limitations of Aggregate Data in the Study of Electoral Behavior," in *Essays in the Behavioral Study of Politics* (Urbana: University of Illinois Press, 1962), pp. 9–102.
44. *Op. cit.*, p. 133.
45. Leo A. Goodman, "Some Alternatives to Ecological Correlation," *American Journal of Sociology*, LXIV, 6 (May 1959), p. 611; also H. Menzel, "Comment on Robinson's Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, XV (October 1950), p. 674."
46. For one recent example see D. P. Bwy, "Political Instability in Latin America: The Cross-Cultural Test of a Causal Model," *Latin American Research Review*, III:2 (1968), pp. 17–66.
47. *Op. cit.*, p. 159.

Latin American Research Review

48. *World Handbook of Political and Social Indicators*, *op. cit.*, p. 323. See also his *Mathematics and Politics* (New York: Macmillan, 1965). Evidence of the contextual effect of different regions is sometimes disconcerting to analysts of aggregate data, if encouraging to area specialists. For example, Raymond Tanter and Manus Midlarsky, "A Theory of Revolution," *Journal of Peace Research* (September 1967), discovered that while change in GNP per capita and primary school enrollment ratios "worked" in the predicted directions for Asia and the Middle East, they worked in the opposite direction for Latin America. They "save" their initial hypotheses by drawing in parametric conditions, e.g. the improbable observation that "Asian countries. . . may not experience the same degree of class rigidity as the Latin American nations."
49. The essential works are Herbert A. Simon, *Models of Man* (New York: Wiley, 1957); Hubert M. Blalock Jr., *Causal Inferences in Nonexperimental Research* (Chapel Hill: University of North Carolina Press, 1961); *Idem.*, "Theory Building and Causal Inferences," in H. Blalock Jr. and A. Blalock (eds.), *Methodology in Social Research* (New York: McGraw Hill, 1968); Paul Lazarsfeld, "Evidence and Inference in Social Research," *Daedalus*, 87, 4 (Fall 1968); *Idem.*, "The Interpretation of Statistical Relations as a Research Problem," in P. Lazarsfeld and M. Rosenberg (eds.), *The Language of Social Research* (New York: Free Press, 1955), pp. 115–125. Hayward Alker Jr. has shown the way for political scientists. See his "Causal Inference in Political Analysis" in Joseph Berhd (ed.), *Mathematical Applications in Political Science*, 2nd Series, (Dallas: S.M.U. Press, 1966) and *Mathematics and Politics*, *op. cit.* For a clear, concise philosophical treatment, see Mario Bunge, *Causality* (Cambridge: Harvard University Press, 1959).
50. Hubert Blalock Jr., *Causal Inference in Nonexperimental Research*, *op. cit.*
51. *Ibid.*, p. 15.
52. Cf. *Ibid.*, and Herbert Simon, *Models of Man*, *op. cit.*, pp. 12–13.
53. *Theory and Methods of Social Research* (London: George Allen and Unwin, 1967), pp. 383–384.