Article:
Boyd, BK, Gove, S and Solarino, AM orcid.org/0000-0001-5030-7375 (2017) Methodological Rigor of Corporate Governance Studies: A Review and Recommendations for Future Studies. Corporate Governance, 25 (6). pp. 384-396. ISSN 0964-8410

https://doi.org/10.1111/corg.12208

METHODOLOGICAL RIGOR OF CORPORATE GOVERNANCE STUDIES: A REVIEW AND RECOMMENDATIONS FOR FUTURE STUDIES

Brian K. Boyd
City University of Hong Kong
bkboyd@cityu.edu.hk

Steve Gove
University of Vermont
Burlington, Vermont USA
steven.gove@uvm.edu

Angelo Solarino
Leeds University Business School
a.m.solarino@leeds.ac.uk

ABSTRACT

Manuscript Type: Article/ Review

Research Question/Issue: This study presents the results of a longitudinal content analysis of governance articles published in both CGIR and other leading management outlets. Our aim with this review is to identify problematic areas, as well as opportunities to improve the methodological rigor of governance studies.

Research Findings/Insights: We examine key design aspects of corporate governance articles. We identify and compare trends (1) over time and (2) across publishing outlets. Specifically examined are the geographic scope of samples, design characteristics (i.e., single versus multi-year data, the use of archival, survey, experimental, and qualitative designs, data sources, and reliance on concurrent versus longitudinal data), survey response rates and the use of tests for non-response bias, the statistical analysis techniques used, reporting of results (number and proportion of hypotheses supported, non-significant, and counter significant, use of statistical power analysis, reporting of descriptive statistics and correlations, use and significance of control variables, and endogeneity controls), and the sophistication of the models tests (i.e., mediation and moderation).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/corg.12208

Theoretical/Academic Implications: Based on the results of the review, recommendations for academic researchers are offered in the following areas: (1) designing theoretical tests which both incorporate and go beyond main effects, (2) evaluating a research design's analytical parameters on an a priori basis to maximize reliability, validity, and the potential for publication, (3) incorporating assessment of sensitivity and robustness checks into the analysis, and (4) comprehensive reporting of results including post hoc power analysis, complete and expanded correlation matrix(es), reporting and assessing counter results, and avoiding HARKing.

Keywords: corporate governance, boards of directors, methodology, research design, statistical analysis

INTRODUCTION

This year, Corporate Governance: An International Review (CGIR) publishes its 25th volume. This silver anniversary offers a unique opportunity to assess the state of the corporate governance field. While previous reviews have assessed the state of the corporate governance research from a theoretical perspective (e.g. Daily, Dalton, & Cannella 2003; Filatotchev & Boyd, 2009; Hambrick, Werder, & Zajac 2008) assessing methodology provides further insight into how the field has matured over time. Academic fields and disciplines advance in two interrelated ways: the nature and sophistication of the questions being addressed, and the techniques employed to explore those questions (Kuhn, 1967) – i.e., theory and methods. Consequently, a content analysis of corporate governance methodology is a critical complement to prior reviews of theory.

Indeed, the research topics and methods within a field are not fully independent, but rather are recursive: Addressing new questions may require the development of new methods and measures. Consequently, the availability of analytic tools often constrains the types of questions that can be asked. For example, if a researcher's statistical arsenal includes only linear regression, then contingency effects such as moderation and mediation may not be considered when framing new theory. Similarly, imprecise measures may cause attenuation, leading researchers to mistakenly conclude that two variables are unrelated. Consequently,

seeking to address new questions often requires new methodologies (Hoskisson, Hitt, Wan & Yiu, 1999). Once utilized, such new methods are commonly employed to reassess prior questions, due to their potential for providing new insights, either by challenging or extending prior findings (Boyd, Gove & Hitt, 2005B). It is through this recursive cycling of questions and methods that a field advances. To aid this progression, we present the results of a longitudinal content analysis of governance studies and conclude with a series of methodological recommendations and suggestions for authors.

METHODOLOGICAL NORMS OF PRIOR GOVERNANCE WORK

How robust is the methodology of empirical governance research? To address this question, we identified and examined a pool of journal articles drawn from CGIR, the leading field-specific outlet. Using quantitative content analysis, we compare methodological practices of CGIR articles against comparable governance articles published in a broader pool of management-oriented journals. Similar to prior work (Crook, Shook, Morris & Madden, 2010), we examine multiple aspects of the methodology employed by studies, including geographic scope of the sample, basic design characteristics, the analytical tools employed, aspects relating to the validity and generalizability of surveys, the presentation of empirical results, and the sophistication of the empirical model employed. Within each aspect we summarize findings for studies published in CGIR and those published in the comparison pool of journals. We then compare and contrast the approaches, and offer suggestions for future practice.

The Samples

CGIR Sample. An ideal sample provides the longitudinal span necessary to detect trends in methodological approaches over an extended period of time (Duriau, Reger & Pfarrer, 2007; Lovejoy, Watson, Lacy & Riffe, 2014). We first created a convenience sample

of articles published in CGIR during the years 1995, 2000, 2005, 2010, and 2015. A twenty-year period appears sufficient to address changes in the methodology used in corporate governance studies as it is consistent with that used by similar review pieces in the literature (Boyd, Gove & Hitt, 2005A; Duriau et al., 2007; Hitt, Gimeno & Hoskisson, 1998) and covers virtually the full life of CGIR from its 1993 inception to today. Assessment of early CGIR volumes posed two problems: access to the articles and an editorial focus on practitioner as opposed to academic audiences. With just two or three empirical articles per issue, early volumes included far fewer research articles than in more recent years. As such, we concluded that it was both impractical and less productive to content analyze early volume of CGIR. An expansion of academically oriented articles began with CGIR's fifth volume in 1997. Many, however, were conceptual and thus ill-suited for this review of methods. To obtain a sufficient sample of articles, we therefore combined articles published in 1997 and 2000 (12 articles; 2 qualitative), and compared them to those published in 2005 (22 articles; 2 qualitative), 2010 (29 articles), and 2015 (27, 1 of which was qualitative). In total, we coded 90 articles.

Comparison Pool. As a complement to our analysis of CGIR articles, we also analyzed governance papers in a comparison pool of high quality, but more general management outlets. This approach allows us to examine whether methodological practices of articles differ between the governance-specific outlet and more general outlets, and how possible differences might manifest over time. We used the same time windows (i.e., 1995, 2000, 2005, 2010, and 2015) to better observe any differences between journal pools. We then identified empirical governance articles published during these years in Academy of Management Journal, Journal of International Business Studies, and Strategic Management Journal. These journals have been publishing high quality work over the duration of the period examined (MacMillan, 1991), including studies in the governance area.

Next, we identified a random subset of articles, stratifying by journal and year (e.g., if there were more governance articles published in one window than another, we selected proportional subsets of articles). The same process was used to stratify publication outlets. Our final comparison article pool was comprised of 78 articles, of which six used qualitative designs absent any empirical method. Many of our content coding items are based solely on the remaining 72 quantitative papers.

DESIGN CHARACTERISTICS OF STUDIES

In assessing the design characteristics of studies we examined four aspects: the type and number of data sources (i.e., archival, survey, experimental, or qualitative), the use of single-year versus longitudinal designs, the use of concurrent versus lagged variables, and the geographical scope of the sample. **Table 1, Panels A** reports design characteristics for studies published in CGIR.

Data source. There has been an increased reliance on archival data: archival data's usage surpassed 80 percent by 2015 up from about half in 1997/2000. The use of archival data has replaced primary survey data in studies. Once used in a third of studies, survey data's usage declined to just six percent of studies by 2015. A small yet relatively consistent set of studies, 11 percent across the panel, utilize a variety of qualitative data. The sourcing of data is also shifting. In 1997/2000 the split of studies using single versus multiple data sources was nearly fifty-fifty. The use of multiple sources of data has since increased steadily to 88 percent in 2015. Studies relying on a single data source are now uncommon.

Longitudinal design. Paralleling the diffusion of archival data, there has been a clear steady shift from single to multi-year data designs. All studies published in CGIR in 1997/2000 used data from a single year; by 2015 multi-year and panel designs constituted 54 percent of the publications. As one would expect given this shift, the average effective

sample size within studies, N for single-year samples and firm-year N for panel designs, has increased more than tenfold, averaging 4,140 for articles published in 2015.

Lagged design. CGIR studies are increasingly using lagged IV/DV designs. Lagged designs, not utilized by studies in the early period, are now used by the more than half, with usage more than doubling from 2010 to 2015 (23 versus 58 percent of studies, respectively).

Geographical scope. We distinguished U.S.–only samples, non-U.S. single country samples, and multi-national samples. Articles in CGIR predominately use samples drawn from outside of the U.S. Among studies published in 1997/2000, one in four relied exclusively on data from U.S. firms; declining to just one study in eight by 2015. Two-thirds of studies draw from single-country, non-U.S. samples. Single nation studies, while the most commonly observed, are declining in usage too, largely replaced by multi-nation studies (30 percent of all articles in 2015). Nearly one quarter of articles published in CGIR incorporate multi-national samples with an average of eighteen nations included, but with high variability. We found studies ranging from a comparison of just two national governance contexts to one which used a global sample of firms drawn from 83 nations. Caution is warranted as not all papers detail the total number or the specific countries included, many multi-national samples did include U.S. firms, and the number of firms from each national context was frequently not presented. In such instances, U.S. firms may actually be the majority of cases in what otherwise appears to be a largely non-U.S. sample.

Table 2, Panel B reports key design characteristics for corporate governance studies published in the comparison pool outside of CGIR.

Data Sources. Approximately seventy percent of articles relied on archival data sources, while over twenty percent relied on survey data. A small number of articles utilized qualitative data for their analyses. From 2000 onwards, the proportion of analyses based on survey versus archival data was relatively flat. Regarding data sources, three quarters of

articles utilized multiple data sources, with the remaining studies drew on single data sources. The reliance on single data sources has declined substantially in the past decade from almost 40% to about 25%.

Longitudinal design. Nearly two thirds of articles relied on cross-sectional designs, with the remaining third of articles using a panel approach. The use of panel data has grown substantially in the three most recent time windows, reaching 58% in 2015. The mean sample size was 8,446 observations, and varied substantially across all five observation windows. Anecdotally, variability in sample sizes appears to correspond to the proportion of multi-year studies published within a given year.

Lagged design. About two thirds of studies used a cross sectional approach, with concurrent data for independent and dependent variables, with the remaining articles incorporating time lags between the collection of predictor and outcome variables. There has been a small, upward trend in the use of lagged designs.

Geographical scope. In 1995, three quarters of the samples were based on U.S. firms. This proportion has declined in subsequent years; while the U.S. is still the most widely studied region, the majority of more recent samples are based on non-U.S. data. Also, while there is a great deal of variation across individual time windows, non-U.S. samples are fairly evenly split between single and multi-nation approaches. As cautioned with the CGIR articles, a multi-nation sample may include the U.S.; many did. For multi-nation samples, the number of nations studied varied widely: In some years, the typical article was based on two or three countries, while 2010 saw publication of articles with extensive geographic coverage.

Comparison of Design

The CGIR and comparison studies are similar in many design aspects, but several important differences do exist. The use of single year of data has been the predominant approach; averaging 63 percent and 64 percent in the two samples. However, reliance on this

has declined in both samples and is now used by less than half of those examined. Researchers increasingly rely on a variety of forms of multi-year data. This shift partly explains the rapidly increasing sample size among studies in both journal pools, due in part to the multiplicative nature of the shift toward firm-year sample sizes reported in panel designs. While this trend is noted in both samples, there exists a continuing disparity with samples used by studies appearing in CGIR which are consistently smaller than those published elsewhere.

More recent articles appearing in CGIR and those in the comparison pool are nearly equally likely to utilize an archival approach. While survey-based work has been the basis for a minority of articles in both samples, usage outside CGIR has remained relatively stable yet has decreasingly steadily within the journal. Historically, half of studies in CGIR rely on primary survey data. More recently, the use of the method within CGIR has declined to less than half of that in the broader pool of journals, just six percent of articles in 2015. Work based on qualitative methods is present, but usage erratic across years in both samples. Work based on experimental methods is virtually absent in the literature examined.

Both samples show a clear uptrend in the use of multiple data sources. This may be due to a combination of factors, including declining usage of the survey approach and a shift toward research questions which cannot be answered by the data available within an individual database. Within CGIR, the use of lagged variable designs in studies was virtually absent, but now constitutes the majority and is generally consistent with that appearing in the comparison journals. Additionally, the usage of lagged designs in CGIR is more than 50% higher in 2015 than in other journals.

Several differences in the geographic scope of the studies in CGIR in comparison to those in the broader pool of journals are notable. First, articles in both pools rely primarily on single-nation studies. However, which individual country is examined differs dramatically.

The proportion of studies using U.S. as the sole geographic context in CGIR were just one fourth of that appearing in the broader pool of studies. In contrast, while two-thirds of the single nation CGIR articles focused on a non-U.S. context, just one fourth of studies outside of CGIR did so. In both sample pools the reliance on single-context studies shows a generally declining trend, with multi-national contexts increasing modestly. Within this shared trend differences do exist. The average multi-national sample in CGIR consists of nearly double the number of countries, 18 countries versus 10, than the broader pool of journals.

ASPECTS OF SURVEY METHOD

A declining, but sizable portion of CGIR studies rely on survey data. We assessed two design components related to the validity and generalizability of studies using survey methodology: response rate and tests for non-respondent bias. **Table 2, Panel A** presents results from studies in CGIR. The response rates reported are steadily declining, falling from forty-three percent in 1997/2000 to just eleven percent in 2015. In 1997/2000, two-thirds of studies using the survey method presented no analysis of non-respondents. In 2015, half of articles used one test. While this trend is positive, caution is warranted as fully half of the 2015 articles report no analysis of non-respondents.

As noted previously, approximately twenty percent of studies in the comparison pool were based on survey data. **Table 2, Panel B** reports statistics on response rates. In general, response rates were consistently high given the norms for executive surveys (Seidel & Westphal, 2004): Across all years, the typical response rate reported was fifty-nine percent. Two thirds of articles tested the generalizability of their samples with some test of non-response bias. There was little consistency in specific tests used: Some articles compared data for early and late responders, whereas others compared descriptive archival data between respondents and non-respondents. More recent survey based articles consistently included tests for response bias.

Comparison of Survey Aspects

While the survey approach was once the basis for one third of studies in CGIR, it was not conducted with a high level of rigor in comparison to studies published in the comparison pool. Along the key aspects of response rate and assessments for non-response bias, survey-based studies in CGIR lag behind those published in peer journals. The average response rate of thirty-six percent is well below the fifty-nine percent reported in other journals. Whereas the response rate reported outside CGIR is stable, within the journal it has declined steadily, to just eleven percent in 2015. Perhaps more problematic is less usage of tests to assess reliability and generalizability by these studies. Outside of CGIR, all of the studies using survey methods report at least one test of non-response bias whereas only half of the more recent studies within the journal do so.

ANALYTIC TOOLS EMPLOYED

Our assessment of analytical tools employed focused on the use of a variety of statistical methodologies ranging from univariate correlations and tests of mean differences to multivariate and multilevel designs. The analytic techniques used within CGIR studies are presented in **Table 3, Panel A**. Most studies used a combination of analytical tools. By far the most common technique was regression analysis (66%), though usage varied widely by year. Two clear uptrends involve specialized regression techniques, specifically logit / probit, as well as alternative methodologies. For example, an increasing number of studies incorporate logit or probit models, with their usage common in the earliest studies examined, falling dramatically in 2005, then increasingly steadily to thirty percent of studies in 2015.

Table 3, Panel B reports on the analytic tools used in the comparison pool. Virtually all of the studies in our article pool were based on regression analysis (81%), analysis of variance (13%), or specialized regression counterparts (i.e., logit or probit models – 17%). None of the articles in our subset used network analysis – which sees occasional use to study

board interlocks – and only a small proportion of articles were based on structural equation modeling (4%). The main observable trends in the use of analytic tools over time were (a) greater likelihood of reporting descriptive statistics in more recent studies, and (b) lesser reliance on comparisons of means via t-tests over time.

Comparison of Analytical Tools

The analytical approaches employed – and those not employed – in corporate governance studies across the two samples are quite similar. In both samples, regression and specialized regression techniques, such as logit and probit models, have consistently been the primary analytical tools utilized. Compared to other methods topics, the choice of analytic tools seems quite similar for both the CGIR and comparison pools.

RESULTS AND THE PRESENTATION OF RESULTS

Turning now to the results section of articles, we have assessed the support rate of hypotheses, the presence of power analysis, reporting of descriptive statistics, the inclusion and significance of control variables, the assessment of endogeneity, and the assessment of construct measurement. **Table 4, Panel A** presents the findings from the assessment of reporting practices in CGIR.

Number of Hypotheses. The typical study is largely unchanged over the period examined in regard to the number of hypotheses, and the number and proportion of hypotheses found significant, non-significant, and significant counter to expectations, and the proportion of studies within which all hypotheses were found to be significant. A typical study contains five to six hypotheses, two to three of which are typically significant, with at least one found significant in a direction counter to that expected. A relatively stable one in five CGIR studies report all hypotheses significant and in the expected direction.

Statistical power. The assessment of statistical power is virtually entirely absent from the studies. Across the period examined, just one percent of studies mentioned statistical

power analysis and none presented a power calculation. A small, positive trend exists as the sole mention of power within articles was among those most recently published. A positive trend was identified in the reporting of essential descriptive statistics.

Descriptive statistics. While nearly eighty percent of studies published in 1997/2000 presented no correlation matrix, this declined to just thirty percent in 2015. A substantial increase was identified in the inclusion of partial correlation matrixes, up from zero to twenty-six percent. The information commonly excluded included dependent variables, mean and standard deviation values, control variables, and in some instances even independent variables. Among the most recent articles, approximately half include complete correlation matrixes, up from just one quarter in 1997/2000. While not specifically coded, it was noted that recent studies are more likely to include moderators among the variables presented in the correlation matrix. This was a positive trend and allows more comprehensive evaluation of the relationship among variables.

Control variables. The number of control variables included in CGIR studies has steadily and dramatically increased, from an average of under one in 1997/2000 to more than eight in 2015. Of concern is that among control variables only, a relatively stable forty percent are found to be statistically significant.

Endogeneity. The most pronounced trend among the reporting of results is the increased attention to potential endogeneity within studies. Among articles in 1997/2000, just nine percent presented any discussion of or control for potential endogeneity. In 2015, a majority, fifty-nine percent of studies, utilized some approach to controlling for endogeneity.

Construct measurement. Overall, the majority of variables across studies were measured with single indicators, which precludes the possibility of assessing reliability. Overall, more than half of studies measured all scales with a single indicator, although this

proportion has declined substantially over time. Among studies with multi-item measures, it has been the norm to not report reliability statistics.

Table 4, Panel B present statistics on the reporting practices within Results sections for studies in the comparison pool.

Number of Hypotheses. Across all time periods, a typical article has roughly six hypotheses with slightly more than half reported as significant at the $p < .05$ level. Studies more often reported non-significant results as opposed to results significant yet counter to directional expectations. Across all time windows, roughly three in ten articles reported consistent support for all hypotheses.

Statistical power. Despite a substantial number of null findings, none of the articles we examined raised the issue of statistical power – either at a conceptual level or through statistical power analyses.

Descriptive statistics. In the comparison pool, papers were much more likely to include correlation matrices: This was reported in over eighty percent of articles. A substantial number of articles reported only partial versus full matrices, however. Additionally, studies published in the most recent ten years are far more likely to report complete correlation matrices for study variables.

Control variables. The typical article included seven to eight control variables, of which slightly less than half were significant. Over time, the number of control variables used in a given article has grown substantially, from 3-4 in 1997/2000, to more than 12 in 2015. The proportion of statistically significant controls is consistently under fifty percent.

Construct measurement. The norm in the majority of studies is to measure all constructs in a given article with single indicators. One fifth of articles reported having at least some items measured with multiple indicators, but with no reliability data reported. Only a small minority of articles (sixteen percent) included reliability data in their results.

There were no clear trends or changes in measurement practices over the five time windows we observed.

Endogeneity. Endogeneity tests are reported in only fifteen percent of studies, although norms have changed dramatically in recent years: in 2010, about twenty percent of studies reported endogeneity tests, while in 2015 this was reported by more than forty percent of articles.

Comparison of Results

The reporting of results between CGIR articles and the comparison pool are remarkably similar, but modest differences do exist. The number of hypothesized relationships is quite similar, although the comparison pool reports one additional hypothesized relationship significant as opposed to non-significant relative to CGIR. Studies in both pools are nearly equally likely, on average, to report relationships which are significant but counter to the hypothesized direction. The proportion of studies reporting all hypothesized relationships significant was generally stable within CGIR, and both higher among studies in the comparison pool and increasingly common.

The presentation of correlation matrixes differed across the samples. In the comparison pool there is an increasing trend toward the presentation of complete correlation matrices, with three quarters of studies published in 2015 meeting this criteria. Within CGIR, under half present complete matrices and thirty percent of articles present none at all.

Trends in the inclusion and significance of control variables also differ. On average, both samples average six to seven control variables with approximately half reported as non-significant. Within both samples the number of control variables included is increasing, yet the proportion which are non-significant continues to hover just under half. Similarly, both samples demonstrate the recent trend for the inclusion of mechanisms to control for endogeneity with usage being more common in recent years in article published in CGIR.

Finally, there are differences across the two article pools regarding reliability. On one hand, CGIR articles were more likely to utilize multi-item constructs in their research design. However, CGIR articles were also less likely to report any reliability information, relative to the comparison pool.

MODEL SOPHISTICATION

Lastly, we assessed the sophistication of the models used in the studies (**Table 5, Panel A**). We examined the most sophisticated models and analysis presented in each article, coding if it consisted solely of main effects (least sophisticated), moderation or mediation, or a combination of both moderation and mediation (most sophisticated). Across the panel, the most common approach to testing hypotheses was the main effect (sixty percent of articles). However, the use of these models has declined steadily and substantially, from eighty-eight percent of articles in 1997/2000 to thirty-eight in 2015. There appears to be a trend toward increasingly sophisticated models among CGIR articles. The largest growth was is the in the incorporation of moderation which, by 2015, was used in the majority of articles. Moderation appeared in two forms: subgroup analysis and statistical interaction terms. Mediation was less common within studies, but has grown to four percent of articles. The most sophisticated models, those incorporating both moderation and mediation, were not present among CGIR articles in 2010 or prior, but were incorporated in four percent of articles in 2015. Collectively, these suggest a clear trend toward the incorporation of increasingly sophisticated models within studies.

In our comparison pool, as shown in **Table 5, Panel B**, only one quarter of articles focused exclusively on main effects. This proportion is relatively stable for most of our time windows, but dropped substantially for 2015. In contrast, moderation appeared frequently among article in the comparison pool. While not detailed in the Table, moderation effects were far more commonly tested using interaction terms versus subgroup analyses. In

comparison, mediation models were used in less than ten percent of studies. There was no clear trend toward a growing use of mediation in more recent years. Finally, there were no examples of mediated moderation in any of our sample articles.

Comparison of Empirical Model Sophistication

The sophistication of models in CGIR articles lags behind that of studies appearing in our comparison pool. Model sophistication tends to increase as a field advances, with moderated and mediated models providing incremental advancements to the body of knowledge beyond existing explanations. There are several possible explanations for studies in CGIR placing greater emphasis on main effects. One explanation would be that articles in the journal focus on testing boundary conditions in additional national contexts where a sufficiently high degree of value added can be obtained from main effects alone to warrant publication. This would be consistent with the higher number of national contexts included in the multi-national samples published in the journal. However, the trend among studies in both samples is toward increase utilization of more sophisticated designs. More than half of the most recent studies in CGIR and the comparison pool go beyond testing main effects. In contrast, all of the more recent studies within the comparison pool go beyond tests of main effects. Collectively, this trend suggests an increasing emphasis in the field for providing refinement of existing models as opposed to tests of application within new and novel contexts.

A ROADMAP FOR AUTHORS

Roughly twenty years ago, the first author published a review of international governance studies (Boyd, Carroll & Howard, 1996). At that time, international governance was a nascent field: Work spanned a diverse set of disciplines and journals, and there was little continuity across articles. Work focused on a handful of nations – primarily the UK, Germany, and Japan. Other European nations were largely excluded, and emerging

economies received no attention at all. More critically, the studies themselves were relatively unsophisticated: The bulk of papers confined themselves to reporting descriptive information about governance practices in a single nation. The more advanced articles might compare governance norms across regions. Substantially less than ten percent of these papers developed hypotheses to explain governance practices. From Kuhn's (1967) paradigm development framework, the nature of early research is not surprising: By definition, work in a new area is exploratory, and the development of robust theories, frameworks, and the means to test them an iterative process. Rather than criticizing early governance studies, the point of this comparison is to illustrate just how far international governance studies have progressed in a relatively short span of time. In order to foster this development, we offer a set of methodological recommendations – for both authors and reviewers – when designing and evaluating studies. **Figure 1** offers an overview of the steps discussed in the following paragraphs.

Step 1: Design Theory

As reviewers and editors, one observation we have is that many governance submissions are framed around a research gap or omission. This means testing a relationship between x and y in a region that has not been studied, even if x.y has been widely tested – perhaps with consistent findings as well – elsewhere. The trend of examining a broad set of countries is most notable for studies in CGIR. Few would argue against testing effects across national contexts. However, the reasons for assessing a x.y relationship in a another national context or the inclusion of a larger numbers of countries in a sample should be approached cautiously. If comparative/multi-national studies are important for tests of boundary conditions, what number of countries and which specific countries should be included to allow for valid and statistically powerful tests? On what basis should countries be compared? What factors drive the selection of a given region? We found little justification as to the

number of countries examined and support for the inclusion of specific countries in the samples. The main rationale for data collection is often that of a convenience sample or "more is better." A richer approach is to ask how a specific region might extend or challenge an existing theory, or, alternately, how a theoretical proposition would make one region more desirable than another.

Scholars should therefore approach the problem from a theoretical perspective, recognizing that there are indicators of a need for better theoretical explanations. For instance, a substantial number of studies report statistically significant results counter to the expected direction. This represents a dilemma for the field as it suggests either an insufficient basis for hypotheses, or the use of hypotheses that are deliberately undertheorized to facilitate ex post explanation. As the development of hypotheses is built upon the previously published theoretical perspectives and empirical examinations, these counter relationships suggest flaws in the nomological logic (Cronbach & Meehl, 1955). These may be attributable to a range of factors including unidentified boundary conditions, flawed samples, weak study design, among a host of other factors. For example, one tactic commonly employed in CGIR articles was the presentation of directionally competing hypotheses in lieu of exploratory relationships. These are commonly presented in the form of two hypotheses where Hypothesis 1A posits a positive relationship between x_1 and y_1 and Hypothesis 1B posits a negative relationship x_1 and y_1. This virtually guarantees the identification of a statistically significant relationship, yet also guarantees a non-significant one. In instances where directionally cannot be confidently proposed, perhaps the best approach is the presentation of the competing theoretical logics along with an exploratory hypothesis. Such an approach more accurately reflects the exploratory nature and avoids potential of HARKing (e.g., Bosco, Aguinis, Field, Pierce & Dalton, 2016). One effective technique is to approach a phenomenon using multiple theoretical lenses. Authors can explore develop a framework to explore the

similarities and differences of two perspectives: Theories could be competing or complementary in their effect, for example, leading to possible suppression or synergistic effects. Many studies on the interaction of CEOs and their boards of directors use this multi-theoretic approach, and provided unique insights regarding governance (Boyd, Haynes & Zona, 2011).

A second commonly observed gap is to test x.y with an incremental refinement in the measurement in either variable. Such papers are difficult to publish, as there are a near-infinite set of gaps, and many are not necessarily novel or insightful. An alternate approach is to ask, how can methodology be used to frame theoretically interesting questions?

There are several methodologies that can be used. Network analysis, for example, is rarely seen in governance studies. Yet, the value of relational data – e.g., a map of pairwise connections among all members of a system – offers the potential for insights beyond what current methods can yield. As an example, many scholars will use a simple count of interlocks to measure a firm's centrality in an intercorporate network. However network analysis algorithms can offer a range of overlapping yet distinct metrics of centrality, all more nuanced than a simple count of ties. Meta-analysis represents another opportunity for theory development. Historically, the utility of meta-analysis was primarily in calculating a composite effect size across a number of studies, including some tests for possible moderation. More recently, governance scholars have been combining meta-analysis with structural equation modeling (MASEM), allowing the development of competing models (e.g., Bergh, Aguinis, Heavey , Ketchen, Boyd, Su, Lau, & Joo, 2016) and complex mediated models (e.g., Essen, Oosterhout & Carney, 2012), while drawing on extensive samples.

Step 2: Research Design

Once a problem has been identified as a theoretical contribution, the scholar must select analytical parameters sufficient to test hypothesized relationships. For sake of space we

will not discuss how to choose the right methodology, but instead we will focus on key elements that have been highlighted as concerns in our content analysis.

Construct measurement. The analytic parameters of the study help to ensure correspondence between what the authors are proposing and what they actually test. This problem is especially prominent regarding constructs. Many macro constructs are measured with single indicators (Boyd, Gove & Hitt, 2005A; Boyd, Bergh, Ireland & Ketchen, 2013), a practice mirrored in the results of our content analysis. However, many governance constructs are quite nuanced. As such, more rigorous approaches to the measurement of governance constructs are desirable (e.g., Boyd, 1994; Dalton & Aguinis, 2013; Gove & Junkunc, 2013). Regardless of the type of data used (i.e., archival versus survey), it is critical to ensure that constructs are comparable across regions. For instance, a common practice in cross-cultural studies is back translation of survey measures after the initial transformation to the language of study (e.g., McGorry, 2000). Anecdotally, few studies in our content analysis reported such a component to their analysis. Even with archival measures, it is imperative to ensure than a seemingly objective measure is capturing the same phenomenon across regions. As a very simple example of the latter, consider a study which examines the performance effects of board independence. National differences in accounting standards will lead to different x.y relationships that are an artifact of measurement, versus a substantive difference between the variables themselves.

Control variables. Control variables play a critical role in shaping the interpretation of the analysis, by ruling out possible alternative explanation. A lack of significance by a control variable may be due to poor measurement, the inclusion of inappropriate and/or confounding controls, previously unidentified boundary conditions, and an alternative theory's supercedence over that included as a control among numerous other factors. As control variables represent known significant relationships, their lack of significance within a

model should call into question the validity of the results. In our sample we found a substantial number of non-significant controls. This is a concern as the inclusion of control variables which are non-significant presents an "illusion of statistical control" and the erroneous perception that explanations extend beyond those previously reported in the literature (Carlson & Wu, 2012). Bernerth and Aguinis (2016) offer best practice recommendations on how to choose and justify the inclusion of a control variable.

Power analysis. Assessment of statistical power is essentially absent from corporate governance studies. This is especially problematic given the small effect sizes often found in governance studies and related research (e.g., Bergh, et al., 2016; Boyd & Solarino, 2016).While the overall trend toward increased sample sizes helps to ensure sufficient power, this may be offset by the increasing number of control variables and, depending on how they are included within models, the increased number of hypotheses. Unfortunately, the validity of either perspective cannot be assessed absent an actual calculation of power, ideally on an a priori basis. The inclusion of a statistical control for power has important theoretical implications, especially in presence of statistically non-significant findings. Indeed, it will allow the reader to distinguish between hypotheses that need to be replicated in other contexts from hypotheses that need to be replicated with larger samples.

Endogeneity. There is a clear trend for incorporating controls for endogeneity in corporate governance studies, with those appearing in CGIR more likely to do so than studies in the comparison pool, but still below the level of governance studies from the finance discipline (Boyd & Solarino, 2016). Among the approaches and remedies utilized were lagged designs, two-stage, and instrumental variable approaches. The field will find benefit in the continued usage of controls for endogeneity as well as an attention to the quality and effectiveness of alternative approaches.

Step 3: Robustness of the Analysis

Assuming the authors have correctly performed their analysis, they should be prepared to be challenged on every methodological decision they make, and also prepare a contingency plan to respond to such challenges. We recommend that authors conduct both pre- and post-assessments for supporting the robustness of their findings. For example, what if the reviewers disagree about the generalizability of the researchers' sample, or operationalization of key constructs? With the design, will the researcher have the ability to make adjustments post hoc to address such concerns? As a starting point, authors should assess the likely magnitude of effects they plan to study (Cohen, 1987), and how much the effects can be attenuated by measurement error. These details guide the selection of a minimal sample size needed to detect effects.

Similarly, the researchers should have already run their own tests to demonstrate the generalizability of their theoretical contribution beyond the sample, or the effects of substituting alternate variables. For example, governance researchers publishing in CGIR using the survey method should seek to improve both the response rate achieved and be vigilant in assessing bias due to non-respondents. The average response rate of thirty-six percent among CGIR articles using the survey approach across the period is consistent with the mean of thirty-four percent reported for mail surveys involving top managers (Cycyota & Harrison, 2006). However, the response rates for CGIR in 2010 and 2015 are twenty-nine percent and eleven percent, respectively; well below the Cycyota and Harrison (2006) norm, and declining rapidly. While it has long been recognized that the challenges of study managerial elites, including board members, can be daunting (Norburn, 1989) it is essential if a survey or other primary data approach is undertaken. Low response rates increase the likelihood of nonresponse bias (Heckman, 1979). As response rates decline, analysis of non-respondent bias becomes increasingly important. At present response rates, however, this

analysis appears to be significantly underutilized given the combination of high rate of usage outside of CGIR and a lower response rate. Conducting and presenting analysis of non-respondent bias should be a mandatory aspect of publishing using the method. Multiple tests are desirable to assess respondent bias, and the most commonly used approaches – e.g., comparison of early versus late respondents, or comparison to archival metrics – are less effective diagnostic tools (Rogelberg & Stanton, 2007). Consequently, this topic deserves strong attention by both authors and reviewers. Many authors take advantage of peer review after drafting their initial manuscript. In many cases, it can be beneficial to solicit feedback in the design stage. While not specific to governance studies, Aguinis and Vandenberg (2014) provide a number of recommendations for preventive steps to take in the design stage of a project.

We advise the authors to develop a developed a list of supplementary analyses that reviewers might ask, such as the use of alternate measures. Other common requests might include handling of outliers and missing data, transformations to address non-normal distributions, and so on. Ideally, authors should know the answers to many of these possible questions prior to submission. A preventive tactic is to include the most salient of these items as supplementary analyses, perhaps labeled as "robustness checks". It is important to note that the purpose of these analyses are not data dredging – or HARKing (or hypothesizing after results are known (Bosco, et al, 2016)) – rather the purpose is to assess how sensitive hypotheses tests are to variations in design parameters: To the extent that hypothesis tests are unaffected by a range of methodological adjustments, they will be perceived as being more robust, and hence more meaningful.

Step 4: Reporting Results

The fourth step concerns reporting results. Correlation matrices provide valuable diagnostic information to readers, and are a critical input for subsequent meta-analyses.

Consequently, authors should provide full matrices, including means and standard deviations, and all control variables. Anticipating the next generation of meta-analytic tools, interaction terms should also be included. Additionally, authors should avoid reporting non-significant correlations as "NS" (Schmidt & Hunter, 2015). If subsamples are used, these matrices should be reported as well. Reviewers should avoid suggesting removal of correlation data in the interest of parsimony.

Authors should also strive to report their findings as fully as possible. In many cases, we observed control variables whose specific findings were omitted to simplify a table or figure. Similarly, power analyses should be reported in the case of non-significant results, particularly for survey-based or cross-sectional datasets, which often have smaller sample sizes. A final item with regard to reporting results concerns HARKing. In this scenario, authors rewrite hypotheses to make them consistent with actual findings. This practice is often driven as much by reviewers as authors, and contaminates theory. In practice, a disproportionate number of supported hypotheses, or an absence of counter findings is a proxy for HARKing behavior (O'Boyle, Banks, & Gonzalez-Mule, 2017). Authors, reviewers, and editors should seek to avoid eliminating non-significant hypotheses on a post hoc basis.

CONCLUSION

In the opening to this study, we described the recursive nature of scientific progress: New questions beget new methods, and new methods beget new questions. In the spirit of advancing corporate governance's next generation of research measurement and methods, we have presented a review of the last 25 years of research methodology, published within and outside of CGIR. Our comparison across time and publication outlets indicates that international governance studies have advanced rapidly in terms of methodology. However, we also identified a number of substantial methodological concerns, which have the potential

to muddy both theory development and the creation of normative guidelines for corporate boards. We hope that our roadmap will stimulate the development of more rigorous, and insightful, studies.

REFERENCES

Aguinis, H., & Vandenberg, R.J. (2014). An ounce of prevention is worth a pound of cure: Improving research quality before data collection. **Annual Review of Organizational Psychology and Organizational Behavior**, 1: 569-595.

Bergh, D.D., Aguinis, H., Heavey, C., Ketchen, D.J., Boyd, B.K., Su, P., Lau, P, & Joo, H. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical demonstration via the strategic leadership – performance relationship. **Strategic Management Journal**, 37: 477-497.

Bernerth, J. B., & Aguinis, H. (2016). A critical review and best practice recommendations for control variable usage. **Personnel Psychology**, 69(1), 229-283.

Bosco, F.A., Aguinis, H., Field, J.G., Pierce, C.A., & Dalton, D.R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. **Personnel Psychology**, 69(3): 709-750.

Boyd, B.K. (1994). Board control and CEO compensation. **Strategic Management Journal**, 15: 335-344.

Boyd, B.K., Bergh, D.D., Ireland, D., & Ketchen, D.J. (2013). Construct measurement in strategic management research. **Organizational Research Method**, 16: 3-14.

Boyd, B.K., Carroll, W.O. & Howard, M. (1996). International governance research: A review and research agenda. In S.B. Prasad & B.K. Boyd (Eds.) **Advances in International Comparative Management**. Greenwich: JAI Press: 191-215.

Boyd, B.K., Finkelstein, S., & Gove, S. (2005). How advanced is the strategy paradigm? The role of particularism and universalism in shaping research outcomes. **Strategic Management Journal**, 26: 841-854**.**

Boyd, B.K., Gove, S., & Hitt, M.A. (2005A). Construct measurement in strategic management research: Illusion or reality? **Strategic Management Journal**, 26: 239-257**.**

Boyd, B.K., Gove, S., & Hitt, M.A. (2005B). Consequences of measurement problems in strategy research: The case of Amihud and Lev. **Strategic Management Journal**, 26(4): 367-375.

Boyd, B.K., Haynes, K.T., & Zona, F. (2011). Dimensions of CEO – board relations. **Journal of Management Studies**, 48: 1892-1923.

Boyd, B.K., Haynes, K.T., Hitt, M.A., Bergh, D.D., & Ketchen, D.J. (2012). Contingency models in strategic management: Use, disuse, or misuse? **Journal of Management**, 38: 278-313.

Boyd, B. K., & Solarino, A. M. (2016). Ownership of Corporations A Review, Synthesis, and Research Agenda. **Journal of Management**, 42(6): 1282-1314.

Bradbury, M.E., & Uno, M. (1993). Analysing Japanese Financial Statements: Practical adjustment procedures for intercorporate share ownership. **Corporate Governance: An International Review**, 1(2): 76-82.

Brown, P.M. (1993). Independent advice on directors' remuneration – Experience from the UK. **Corporate Governance: An International Review**, 1(3): 120-127.

Cadbury, S.A. (1993). Thoughts on corporate governance. **Corporate Governance: An International Review**, 1(1): 5-10.

Carlson, K.D., & Wu, J. (2012). The illusion of statistical control: Control variable practice in management research. **Organizational Research Methods**, 15(3): 413-435.

Cohen, J. (1987). **Statistical power analysis for the behavioral sciences** (2nd edition). Erlbaum: Hillsdale, NJ.

Cooper, B.J., Chow, L., & Wei, T.Y. (1993). The separation of ownership and management in China: A review of the business contract and share capital systems. **Corporate Governance: An International Review**, 1(3): 128-137.

Courtis, J.K. (1993). Recent developments and frustrations in the corporate reporting process. **Corporate Governance: An International Review**, 1(1): 18-25.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. **Psychological Bulletin,** 52: 281-302.

Crook, T.R., Shook, C.L., Morris, M.L., & Madden, T.M. (2010). Are we there yet?: An assessment of research design and construct measurement practices in entrepreneurship research. **Organizational Research Methods**, 13(1): 192.

Cycyota, C.S., & Harrison, D.A. (2006). What (not) to expect when surveying executives: A meta-analysis of top manager response rates and techniques over time. **Organizational Research Methods**, 9(2): 133-160.

Daily, C. M., Dalton, D. R., & Cannella, A. A. (2003). Corporate governance: Decades of dialogue and data. **Academy of Management Review**, 28(3), 371-382.

Dalton, D.R., & Aguinis, H.A. (2013). Measurement malaise in strategic management studies: The case of corporate governance. **Organizational Research Methods**, 16: 88-99.

Daniels, R.J., & Waitzer, E.J. (1993). Corporate governance: A Canadian perspective. **Corporate Governance: An International Review**, 1(2): 66-75.

Duriau, V.J., Reger, R.K., & Pfarrer, M.D. (2007). A content analysis of the content analysis literature in organization studies. **Organizational Research Methods**, 10(1): 5-34.

Essen, M.V., Oosterhout, J.H.V., & Carney, J. (2012). Corporate boards and the performance of Asian firms: A meta-analysis. **Asia Pacific Journal of Management**, 29: 873-905.

Filatotchev, I., & Boyd, B.K (2009) Taking stock of corporate governance while looking to the future. **Corporate Governance: An International Review**, 17: 257-265.

Gove, S., & Junkunc, M. (2013). Dummy constructs? Binomial categorical variables as representations of constructs: CEO duality through time. **Organizational Research Methods**, 16(1): 100-126.

Hambrick, D. C., Werder, A. V., & Zajac, E. J. (2008). New directions in corporate governance research. **Organization Science**, 19(3): 381-385.

Heckman, J.J. (1979). Sample selection as a specification error. **Econometrica**, 47: 153-161.

Hitt, M.A., Gimeno, J., & Hoskisson, R.E. (1998). Current and future research methods in strategic management. **Organizational Research Methods**, 1(1): 6-44.

Hoskisson, R.E., Hitt, M.A., Wan, W.P., & Yiu, D. (1999). Theory and research in strategic management: Swings of a pendulum. **Journal of Management**, 25(3): 417-456.

Kuhn, T.S. (1967). **The structure of scientific revolutions**. Chicago: University of Chicago Press.

Lovejoy, J., Watson, B.R., Lacy, S., & Riffe, D. (2014). Assessing the reporting of reliability in published content analyses: 1985-2010. **Communication Methods and Measures**, 8(3): 207.

MacMillan, I.C. (1991). The emerging forum for business policy scholars. **Strategic Management Journal**, 12(2): 161-165.

McGorry, S. Y. (2000). Measurement in a cross-cultural environment: survey translation issues. **Qualitative Market Research**, 3(2): 74-81.

Norburn, D. (1989). The chief executive: A breed apart. **Strategic Management Journal**, 10: 1-15.

O'Boyle, E.H., Banks, G.C., & Gonzalez-Mule, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. **Journal of Management.** 43: 376–399.

Rogelberg, S.G., & Stanton, J.M. (2007). Understanding and dealing with organizational survey nonresponse. **Organizational Research Methods**, 10: 195-209.

Schmidt, F.L., & Hunter, J.E. (2015). **Methods of meta-analysis: Correcting error and bias in research findings.** Thousand Oaks: Sage.

Seidel, M.-D. L., & Westphal, J. D. (2004). Research impact: How seemingly innocuous social cues in a CEO survey can lead to change in board of director network ties. **Strategic Organization**, 2(3): 227-270.

Wood, R.E., Goodman, J.S., Beckmann, N., & Cook, A. (2008). Mediation testing in management research: A review and proposal. **Organizational Research Methods**, 11: 270-295.

TABLE 1: DESIGN CHARACTERISTICS

PANEL A: DESIGN CHARACTERISTICS IN CGIR ARTICLES

Item	1997/2000	2005	2010	2015	Overall
Sample					
Single year	100%	79%	58%	46%	63%
Panel	0%	21%	42%	54%	37%
Overall N	294.2	606.8	2,317.7	4,140.0	2,255.3
Design					
Archival	54%	58%	88%	81%	74%
Survey	31%	23%	9%	6%	15%
Experimental	0%	0%	3%	0%	1%
Qualitative	15%	19%	0%	13%	11%
Data Sources					
Single	44%	33%	13%	12%	20%
Multiple	56%	67%	87%	88%	80%
Data Timing					
Concurrent IV/DV	100%	84%	77%	42%	71%
Lagged IV/DV	0%	16%	23%	58%	29%
Geographical Scope					
U.S. Sample	25%	5%	13%	15%	13%
Non-U.S., single-nation	75%	68%	67%	56%	65%
Multi-nation	0%	27%	20%	30%	22%
If Multi-nation, N (mean per study)		12.0	26.7	16.5	18.4

TABLE 1 (continued)

PANEL B: DESIGN CHARACTERISTICS IN COMPARISON POOL ARTICLES

Item	1995	2000	2005	2010	2015	Overall
Sample						
Single year	80%	82%	50%	61%	42%	64%
Panel	20%	18%	50%	39%	58%	36%
Overall N	610.1	1294.5	534.5	7063.4	32171.4	8,446
Design						
Archival	64%	73%	63%	82%	67%	71%
Survey	32%	27%	13%	18%	20%	22%
Experimental	0%	0%	0%	0%	0%	0%
Qualitative	5%	0%	38%	0%	13%	7%
Data Sources						
Single	38%	36%	33%	14%	17%	25%
Multiple	63%	64%	67%	86%	83%	75%
Data Timing						
Concurrent IV/DV	71%	82%	33%	61%	64%	64%
Lagged IV/DV	29%	18%	67%	39%	36%	36%
Geographical Scope						
U.S. Sample	76%	27%	43%	39%	43%	47%
Non-U.S., single-nation	6%	36%	43%	18%	43%	25%
Multi-nation	18%	36%	14%	43%	14%	29%
If Multi-nation, N	5.3	2.8	2.0	19.5	3.0	10.6

(mean per study)

TABLE 2: SURVEY RESPONSE RATES AND TESTS

PANEL A: SURVEY RESPONSE RATES AND TESTS IN CGIR ARTICLES

Item	1997/2000	2005	2010	2015	Overall
Response rate	43%	45%	29%	11%	36%
Non-response bias					
No Tests	71%	71%	50%	50%	67%
One test	29%	29%	0%	50%	28%
Multiple tests	0%	0%	50%	0%	6%

PANEL B: SURVEY RESPONSE RATES AND TESTS IN COMPARISON POOL ARTICLES

Item	1995	2000	2005	2010	2015	Overall
Response rate	43%	59%	54%	74%	57%	59%
Non-response bias						
No Tests	67%	33%	100%	0%	0%	33%
One test	33%	67%	0%	100%	100%	67%
Multiple tests	0%	0%	0%	0%	0%	0%

TABLE 3: STATISTICAL ANALYSES

PANEL A: STATISTICAL ANALYSES IN CGIR ARTICLES

Item	1997/2000	2005	2010	2015	Overall
Mean t-tests	0%	14%	24%	22%	18%
Factor analysis	8%	9%	0%	0%	3%
Correlation analysis	0%	14%	0%	4%	4%
Regression analysis	0%	41%	100%	78%	66%
Discriminant analysis	0%	0%	0%	0%	0%
Cluster analysis	0%	0%	0%	0%	0%
SEM	0%	0%	3%	0%	1%
ANOVA/ANCOVA	8%	0%	3%	0%	2%
MANOVA/MANCOVA	0%	0%	0%	0%	0%
Logit/probit models	25%	5%	17%	30%	19%
Network analysis	0%	0%	0%	0%	0%
Other	0%	9%	10%	22%	12%

TABLE 3 (continued)

PANEL B: STATISTICAL ANALYSES IN COMPARISON POOL ARTICLES

Item	1995	2000	2005	2010	2015	Overall
Mean t-tests	25%	27%	17%	14%	0%	17%
Factor analysis	19%	0%	0%	0%	17%	7%
Correlation analysis	38%	0%	0%	7%	0%	12%
Regression analysis	75%	91%	100%	82%	75%	87%
Discriminant analysis	0%	0%	0%	0%	0%	0%
Cluster analysis	0%	0%	0%	0%	0%	0%
SEM	0%	0%	0%	11%	0%	4%
ANOVA/ANCOVA	13%	18%	17%	0%	8%	9%
MANOVA/MANCOVA	13%	9%	0%	0%	0%	4%
Logit/probit models	13%	18%	0%	14%	33%	17%
Network analysis	0%	0%	0%	0%	0%	0%
Other	0%	0%	0%	0%	0%	0%

TABLE 4: RESULTS & REPORTING PRACTICES

PANEL A: REPORTING PRACTICES IN CGIR ARTICLES

Item	1997/2000	2005	2010	2015	Overall
Findings					
# of hypotheses supported	3.0	1.5	2.4	2.9	2.4
# of hypotheses non-sig	3.0	2.6	1.9	2.7	2.5
# of hypotheses contrary	2.0	1.0	1.2	1.8	1.5
all hypotheses supported	8%	18%	24%	19%	19%
Statistical power					
Not mentioned	100%	100%	100%	96%	99%
Mentioned, no analyses				4%	1%
Analyses presented					
Correlation matrices					
Not reported	78%	63%	23%	30%	40%
Partial	0%	21%	23%	26%	21%
Full	22%	16%	53%	44%	39%
Control Variables					
N_used	0.5	3.5	6.8	8.1	6.0
N_significant	0.2	1.8	2.4	3.6	2.6
Endogeneity tests					
None	91%	100%	63%	41%	68%
Any	9%	0%	37%	59%	32%
Reliability					
All items single	90%	40%	0%	25%	55%
Some multi, none reported	10%	20%	67%	50%	27%

Some multi, multiple reported	0%	40%	33%	25%	18%

TABLE 4 (continued)

PANEL B: REPORTING PRACTICES IN COMPARISON POOL ARTICLES

Item	1995	2000	2005	2010	2015	Overall
Findings						
# of hypotheses supported	4.4	3.0	3.0	2.8	2.8	3.2
# of hypotheses non-sig	2.3	1.3	1.0	1.4	1.5	1.6
# of hypotheses contrary	2	1.6	1.0	1.0	0.0	1.2
all hypotheses supported	25%	9%	17%	36%	42%	29%
Statistical power						
Not mentioned	100%	100%	100%	100%	100%	100%
Mentioned, no analyses						
Analyses presented						
Correlation matrices						
Not reported	38%	18%	0%	18%	9%	19%
Partial	19%	36%	17%	7%	18%	17%
Full	44%	45%	83%	75%	73%	64%
Control Variables						
N_used	3.8	2.6	5.0	8.9	12.4	7.5
N_significant	1	1.3	2.8	4	4.9	3.1
Endogeneity tests						
None	100%	100%	100%	79%	58%	85%
Any	0%	0%	0%	21%	42%	15%
Reliability						
All items single	63%	82%	83%	50%	58%	62%
Some multi, none reported	13%	0%	17%	36%	25%	22%
Some multi, multiple reported	25%	18%	0%	14%	17%	16%

TABLE 5: MODEL SOPHISTICATION

PANEL A: MODEL SOPHISTICATION IN CGIR ARTICLES

Item	1997/2000	2005	2010	2015	Overall
Main effects	88%	78%	60%	38%	60%
Moderation	13%	22%	33%	54%	35%
Mediation	0%	0%	7%	4%	4%
Both Moderation & Mediation	0%	0%	0%	4%	1%

PANEL B: MODEL SOPHISTICATION IN COMPARISON POOL ARTICLES

Item	1995	2000	2005	2010	2015	Overall
Main effects	25%	27%	33%	29%	0%	23%
Moderation	69%	73%	67%	54%	100%	68%
Mediation	6%	0%	0%	18%	0%	8%
Both Moderation & Mediation	0%	0%	0%	0%	0%	0%

FIGURE 1: A ROADMAP FOR THE DESIGN OF GOVERNANCE STUDIES