

NEW YORK UNIVERSITY
JOURNAL OF LAW & BUSINESS

VOLUME 14

SUMMER 2018

NUMBER 3

EVALUATING THE EFFECTIVENESS OF
CORPORATE COMPLIANCE PROGRAMS:
ESTABLISHING A MODEL FOR
PROSECUTORS, COURTS, AND FIRMS

EUGENE SOLTES*

When prosecutors, courts, and regulators make charging and sentencing decisions, they must evaluate whether firms have effective compliance programs. Such evaluations are difficult because of the challenges associated with measuring effectiveness. Notably, these obstacles are not limited to evaluation by external parties, as firm managers also struggle to assess the returns associated with investments in compliance. To address these limitations and to facilitate a more consistent assessment, this article provides a framework to rigorously evaluate the effectiveness of compliance programs. I present empirical models that use data collected from compliance initiatives—including whistleblower hotlines, training exercises, and disciplinary procedures—to evaluate whether the initiatives are fulfilling their desired objectives. By being able to more rigorously evaluate compliance programs with data and empirical analysis, prosecutors and courts will be able to more readily discern “window-dressing” and firm managers will be able to better understand the efficacy of their investments in compliance.

I. HISTORY OF CORPORATE COMPLIANCE	966
II. CONTEMPORARY COMPLIANCE AND MEASUREMENT	970
III. MEASURING COMPLIANCE EFFECTIVENESS	974

* Jakurski Family Associate Professor of Business Administration, Harvard Business School. I would like to thank Hui Chen, Chris Costa, Jesse Fried, Todd Haugh, William Laufer, Roman Weil, the Fraud Section in the Criminal Division of the Department of Justice, and corporate governance seminar participants at the Harvard Law School for their thoughtful feedback on the ideas contained in this piece and earlier drafts. I also deeply appreciate the research assistance and perspective provided by Julian Sarafian while developing this article.

IV. OBJECTIVES OF CORPORATE COMPLIANCE PROGRAMS	976
V. COMPONENTS OF A COMPLIANCE PROGRAM	979
A. <i>Prevention</i>	980
B. <i>Detection</i>	982
C. <i>Aligning Corporate Activities with Regulation</i> ...	983
VI. MEASURING COMPLIANCE PROGRAMS.....	984
A. <i>Measurement Bias: Differentiating “Paper” Initiatives from Substantive Initiatives</i>	986
B. <i>Empirically Evaluating Compliance Effectiveness</i> .	992
1. <i>Whistleblower Hotlines</i>	993
2. <i>Training</i>	995
3. <i>Individual Accountability</i>	998
4. <i>Analytic Detection Systems</i>	999
VII. HOW “EFFECTIVE” IS EFFECTIVE?	1000
VIII. THE CONCERN OF UNDERINVESTMENT	1004
IX. CHANGING COMPLIANCE MEASUREMENT.....	1006
CONCLUSION	1010

I.

HISTORY OF CORPORATE COMPLIANCE

Over the past 25 years, the size and complexity of corporate compliance programs have increased dramatically. As budgets for supporting programs and the number of people involved with compliance processes have grown, compliance has transformed into its own division within many firms, akin to marketing, operations, accounting, and finance divisions.¹ Today, the average multinational spends in excess of \$3.5 million a year supporting their compliance program.²

1. The 2010 Dodd–Frank Act was the most recent piece of regulation to substantially increase the size of compliance teams by specifically requiring investment advisors to designate a chief compliance officer who oversees the implementation of compliance processes to prevent violation of the Investment Advisory Act. See 17 C.F.R. § 275.206 (2018). For an example of this greater compliance emphasis, see Abha Bhattarai & Catherine Ho, *Four years into Dodd-Frank, local banks say this is the year they’ll feel the most impact*, WASH. POST (Feb. 7, 2014), https://www.washingtonpost.com/business/capitalbusiness/four-years-into-dodd-frank-local-banks-say-this-is-the-year-theyll-feel-the-most-impact/2014/02/07/12c7ca48-877e-11e3-a5bd-844629433ba3_story.html?utm_term=.0f9976af3d41.

2. There is significant variation across firms with some firms in more heavily regulated industries like financial services and energy spending many

The genesis for modern compliance programs began in the 1970s and 1980s following a series of corruption scandals that included numerous defense firms defrauding the United States government.³ In response, defense contractors banded together and agreed to self-regulate by adopting a series of internal policies to prevent and report misconduct. The firms found this self-regulation appealing in that it provided them more control over the process and mitigated the disruptive effects of external regulation of operations.⁴ To regulators this self-policing was also attractive since it eased their burden of detecting potential violations while purportedly reducing the amount of misconduct occurring within firms.

In 1991, buoyed by the perceived benefits of self-regulation and the perception that firms could themselves be victims of the actions of rogue employees, the United States Sentencing Commission amended the Federal Sentencing Guidelines to provide incentives for companies to create compliance programs. Specifically, firms that had an “effective compliance program” under the guidelines could incur substantially reduced fines. In some instances, fines could be reduced by as much as ninety-five percent if the firm had an effective program in place at the time of the misconduct.⁵ A series of memos followed from senior officials in the Department of Justice (“DOJ”) to prosecutors further emphasized the poten-

times more. PONEMON INSTITUTE LLC, *THE TRUE COST OF COMPLIANCE: A BENCHMARK STUDY OF MULTINATIONAL ORGANIZATIONS 2* (2011).

3. For a history of the development of the modern compliance industry see David Hess, *Ethical Infrastructures and Evidence-Based Corporate Compliance and Ethics Programs: Policy Implications from the Empirical Evidence*, 12 N.Y.U. J.L. & Bus. 317 (2016).

4. The Defense Industry Initiative played a formative role in both creating the earliest formal compliance systems and fostering an environment of self-regulation. See Kimberly D. Krawiec, *Cosmetic Compliance and the Failure of Negotiated Governance*, 81 WASH. U. L. REV. 487 (2003).

5. ETHICS RESOURCE CENTER, *THE FEDERAL SENTENCING GUIDELINES FOR ORGANIZATIONS AT TWENTY YEARS: A CALL TO ACTION FOR MORE EFFECTIVE PROMOTION AND RECOGNITION OF EFFECTIVE COMPLIANCE AND ETHICS PROGRAMS 22* (2012). See also Krawiec *supra* note 4, at 499. Frank Bowman provides an example of a firm with more than 5,000 employees whose employees cause a \$10 million loss going from a fine range of \$20–40 million to \$4–8 million with an effective compliance program. See Frank O. Bowman, III, *Drifting Down the Dnieper with Price Potemkin: Some Skeptical Reflections about the Place of Compliance Programs in Federal Criminal Sentencing*, 39 WAKE FOREST L. REV. 671, 678 (2004).

tial benefits of firms having effective programs.⁶ When deciding to charge a firm for illicit conduct, the memos urged prosecutors to take into account a firm's effort to prevent the misconduct. If the firm had an effective compliance program, prosecutors had discretion over the decision whether or not to even formally charge the firm with wrongdoing.⁷ Between the new sentencing guidelines and the revised guidance for prosecutors, the potential benefits associated with having an "effective compliance program" became considerable.⁸

In response, the compliance industry blossomed. Companies emerged to provide consulting, software, and training materials to support the creation of compliance departments.

6. As will be further described, the DOJ is not the only regulatory body with an interest in effective corporate compliance. Yet, as noted by Mariam Baer, "Although numerous other agencies assist in regulation compliance, the DOJ, by dint of its power to bring criminal charges, is one of the most powerful—and therefore most prominent—institutions with the authority to declare a corporation's compliance program effective or deficient." Mariam Baer, *Governing Corporate Compliance*, 50 B.C. L. REV. 1, 24 (2009).

7. Whether and how much firms ought to receive credit for negotiated cooperation and making amends continues to be debated. See William Laufer & Alan Strudler, *Corporate Crime and Making Amends*, 44 AM. CRIM. L. REV. 1307, 1315 (2007) (arguing that prosecution ought to be determined by gravity of offense); Philip Wellner, *Effective Compliance Programs and Corporate Criminal Prosecutions*, 27 CARDOZO L. REV. 497, 498 (2005) (arguing that sentencing downgrades do not serve to improve deterrence).

8. Compliance program effectiveness is reaching other areas of regulation and litigation. In particular, the Environmental Protection Agency ("EPA"), Department of Health and Human Services ("HHS"), and Securities and Exchange Commission ("SEC") consider a firm's compliance program in determining the level of civil fines. These regulatory bodies' guidelines bear significant similarity to those in the Organizational Sentencing Guidelines. Increasingly, a firm's internal compliance structure is also used as part of the defense and as a remedy for harassment claims filed under the Title VII of the Civil Rights Act of 1991. See Krawiec, *supra* note 4, at 506. Globally, compliance programs are used to support compliance with a variety of consumer and regulatory law. See Christine Parker & Vibeke Lehmann Nielsen, *Corporate Compliance Systems: Could They Make Any Difference?*, 41 ADMIN. & SOC'Y (2009). The SEC's consideration of compliance programs is described in Securities and Exchange Commission, Exchange Act Release No. 44969, Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934 and Commission Statement on the Relationship of Cooperation to Agency Enforcement Decisions (2001). The evaluation of compliance programs is also discussed in the U.S. Attorneys' Manual, Offices of the United States Attorneys, Title 9, Section 28.800: Corporate Compliance Programs (2017).

Firms created formal codes of conduct, offered employees the opportunity to report suspicious activity on anonymous whistleblower hotlines, and mandated training programs to support these new programs. The transnational reach of regulation even compelled firms, which did banking services in the United States but were headquartered *outside* the United States, to create compliance programs.⁹ Other countries also created their own legislation—based on the U.S. model—imposing additional liability for failing to create an effective compliance system. Italy, a country that trails only the United States in the number of enforcement actions taken against firms, created Legislative Decree No. 231/2001 (“Law 231”) which stated that in order for a firm to avoid being charged, it must prove it implemented an effective compliance system in advance of the alleged offenses.¹⁰

Today, estimates of the aggregate direct costs to support compliance programs are routinely in the hundreds of billions of dollars. Yet, the indirect costs would push these estimates even higher due to the many hours of employee time consumed on training and other compliance related activities.¹¹

9. Even sending e-mails that cross or reside on servers in the United States is sufficient for the United States to assert territorial jurisdiction. *See* Complaint, U.S. v. Magyar Telekom, PLC, No. 1:11CR00597 (E.D. Va. 2011).

10. The use of “Law 231” was tested for the first time in November 2009 when the Court of Milan did not charge Impregilo for a crime associated with the insider trading of its CEO and Chairman because the firm was viewed as having an effective Law 231 compliance program. *See* BRUNO COVA, FRANCESCA PETRONIO, VIVIANA MARA & MARILENA HYERACI, PAUL HASTINGS, PROTECTING COMPANIES IN A CHALLENGING ENVIRONMENT: COMPLIANCE PROGRAMS UNDER ITALIAN LAW – THE FIRST NINE YEARS (2010), <https://www.paulhastings.com/docs/default-source/PDFs/1521.pdf>. Although the law has been infrequently utilized in Italian courts, a study by the Italian Internal Auditor Association in 2006 found that 91% of sample firms had developed or were in the process of developing a Law 231 compliance program suggesting that firms nonetheless responded to the perceived benefits/costs associated with noncompliance of the law. In Germany, the Institut Der Wirtschaftsprüfer (IDW) created Assurance Standard 980 in March 2011 that described the expectations for auditors in assessing a firm’s compliance program. *See* IDW ASSURANCE STANDARD 980: PRINCIPLES FOR THE PROPER PERFORMANCE OF REASONABLE ASSURANCE ENGAGEMENTS RELATING TO COMPLIANCE MANAGEMENT SYSTEMS (2011).

11. Citigroup alone reported in 2015 that it had 26,000 compliance staff who were paid an average of \$60,000 a year. Laura Noonan, *Banks face pushback over surging compliance and regulatory costs*, FIN. TIMES (May 28, 2015), <https://www.ft.com/content/e1323e18-0478-11e5-95ad-00144feabdc0>.

II.

CONTEMPORARY COMPLIANCE AND MEASUREMENT

Despite the extraordinary resources and time expended on compliance and the ramifications associated with having an effective or ineffective program, there continues to be little agreement amongst firms and legal bodies (i.e., prosecutors, courts, and regulators) over what constitutes an effective compliance program.¹² To prosecutors and courts, this limitation is especially acute, since firms should not be credited with having an “effective” program if the program merely gives off the appearance of effectiveness, but lacks substance (i.e., a “paper” program). Legal and regulatory bodies often have to make effectiveness assessments despite this ambiguity in the criteria for assessing programs. For example, according to the data files of the United States Sentencing Commission, between 2002 and 2016, 530 firms were ordered to create compliance programs as part of their sentencing.¹³ As these cases only represent the small minority of instances where criminal sanctions were imposed on organizations (and excludes cases

12. See, e.g., U.S. DEP’T OF HEALTH AND HUM. SER., AN INTEGRATED APPROACH TO CORPORATE COMPLIANCE: A RESOURCE FOR HEALTH CARE BOARDS OF DIRECTORS (2004) (describing a sample survey for firms to use in determining whether a program is effective, but failing to describe what constitutes adequate or effective responses). Scholars have proposed management theories to improve efficiency in compliance programs, but have not proposed models to test the effectiveness and quality of the programs. See, e.g., Richard S. Gruner, *Lean Law Compliance: Confronting and Overcoming Legal Uncertainty in Business Enterprises and Other Complex Organizations*, 11 N.Y.U. J.L. & BUS. 247 (2014) (describing a management theory to more efficiently manage compliance systems, but omitting any evaluative factors to measure the efficacy of such programs), and Maurice E. Stucke, *In Search of Effective Ethics & Compliance Programs*, 39 J. CORP. L. 770, 802 (2014) (discussing the inability of firms to determine the efficacy of their compliance programs).

13. The statistic was computed using the United States Sentencing Commission organization/corporate data files under the variable “ORDERCE.” As noted in the Variable Code for Organizational Cases, ORDERCE “indicates (as part of the sentence or plea) if the court ordered an ethics program, compliance program, or any systemic effort or program to prevent or detect violations. This includes requiring the offender organization to implement a compliance or ethics program, hire an ethics officer or compliance monitor, sponsor a seminar, lecture or training regarding compliance with regulations or submit a report or reports to the government regarding the same.” U.S. SENTENCING COMM’N, VARIABLE CODE FOR ORGANIZATIONAL CASES 29 (2017).

where charges were not filed, or where other regulatory bodies considered civil sanctions), evaluation of programs is not an uncommon legal issue facing courts, prosecutors, and regulators.¹⁴

The inability to assess the effectiveness of compliance programs does not simply result from a lack of consensus, but from a basic lack of appropriate data and a means to evaluate programs. According to one recent survey of compliance officers, only 70% of firms even attempt to measure the impact of their compliance programs. Among firms who try to measure their programs' effectiveness, only half are reasonably confident that they are measuring it appropriately. If managers cannot even evaluate the efficacy of their own programs, external parties such as prosecutors and regulators with even less information face an especially difficult challenge.¹⁵

In 2015, Andrew Weissmann, Chief of the DOJ's Fraud Section, sought to address this problem by hiring Hui Chen as the department's first "corporate compliance expert." Chen's mandate was to assist prosecutors in evaluating compliance programs, monitor changes in programs, and create standards to evaluate programs.¹⁶ Under Chen's leadership, in February 2017, the DOJ publicly released an extensive series of questions its prosecutors would ask when evaluating the effective-

14. As the decision to charge is also a function of the program, the cases that were prosecuted only make up a small set of the total number of evaluations. Moreover, as part of deferred prosecution agreements and non-prosecution agreements, firms are often required to make changes to their compliance program, which requires additional evaluation. See BRANDON L. GARRETT, *TOO BIG TO JAIL: HOW PROSECUTORS COMPROMISE WITH CORPORATIONS* (2014); Vikramaditya Khanna & Timothy L. Dickinson, *The Corporate Monitor: The New Corporate Czar?*, 105 MICH. L. REV. 1713, 1725 (2007).

15. See WILLIAM S. LAUFER, *CORPORATE BODIES AND GUILTY MINDS: THE FAILURE OF CORPORATE CRIMINAL LIABILITY* (2006) (arguing that compliance-driven ethics and gaming by corporate attorneys helps artificially bolster firm reputations and distance itself from misconduct by not appropriately assessing culpability), and Marie McKendall, Beverly DeMarr & Catherine Jones-Ridders, *Ethical Compliance Programs and Corporate Illegality: Testing the Assumptions of the Corporate Sentencing Guidelines*, 37 J. BUS. ETHICS 367 (2002) (providing an empirical analysis of programs that suggests that many programs seem to serve as window dressing).

16. Press Release, Dep't of Justice, *New Compliance Counsel Expert Retained by the DOJ Fraud Section* (Nov. 2, 2015) (on file with author).

ness of programs.¹⁷ For the first time in more than two decades since “effective compliance” entered the nomenclature of the legal and regulatory process, the Fraud Section’s “Evaluation of Corporate Compliance Programs” offered transparency into how officials would determine whether a firm has an effective program.

The Fraud Section’s document was conceived simply as a list of questions prosecutors would ask when evaluating compliance programs. Notably, it did not provide actual guidance as to what constitutes an effective program. However, many observers and law firms interpreted the document as a handbook defining what was meant by “effective compliance.” In particular, many firms believed that if they could adequately respond to the questions provided by the Fraud Section, this would affirm that their compliance program was effective—at least in the eyes of federal prosecutors and courts.¹⁸

While the Fraud Section’s document offered firms an extensive series of questions to consider when examining a program, it explicitly stated that the list was not necessarily inclusive and the document gave no indication as to what constituted a satisfactory reply. In fact, the document explicitly noted that evaluations would not be performed according to “a checklist nor a formula” and that prosecutors would continue to make “an individualized determination in each case.”¹⁹ Ultimately, the document did not intend to define what constituted an appropriate and sufficiently rigorous re-

17. U.S. DEP’T OF JUSTICE, CRIMINAL DIVISION, FRAUD SECTION, EVALUATION OF CORPORATE COMPLIANCE PROGRAMS (2017), <https://www.justice.gov/criminal-fraud/page/file/937501/download>.

18. For examples of firms noting that the document provided guidance, see *DOJ Issues New Compliance Program Evaluation Guidance*, BAKER MCKENZIE (Feb. 28, 2017), <https://www.bakermckenzie.com/en/insight/publications/2017/02/doj-issues-new-compliance-program/>; memorandum from Alison Tanchyk, Eric W. Sitarchuk, John J. Pease, III, *DOJ Issues New Guidance on Corporate Compliance Programs*, MORGAN LEWIS (Feb. 21, 2017), <https://www.morganlewis.com/pubs/doj-issues-new-guidance-on-corporate-compliance-programs?p=1#>; Carrie Penman, *New Guidance from the DOJ on Your Compliance Program*, NAVEX GLOBAL (Feb. 23, 2017), <https://www.navexglobal.com/blog/new-guidance-doj-your-compliance-program>; MICHAEL W. PEREGRINE, McDERMOTT WILL & EMERY, DOJ RELEASES IMPORTANT NEW COMPLIANCE PROGRAM GUIDANCE (Feb. 24, 2017), <https://www.mwe.com/~media/files/press-room/2017/02/weekly022417.pdf>.

19. U.S. DEP’T OF JUSTICE, *supra* note 17, at 1.

sponse when ascertaining whether a particular compliance initiative was effective.

As an example, the DOJ's memo asks, in a section on training and communications, "How has the company measured the effectiveness of the training?" To the extent that firms train employees, they do so with the intent of helping employees better understand and better comply with firm policies. However, as noted by a recent survey of compliance officers in *Compliance Week*, most firms rely on completion rates to assess the effectiveness of their training.²⁰ That is, if 90% or 95% of employees complete the required training, the compliance officer would deem the training "effective."

However, completion rates are not only unrelated to the training's effectiveness, but are also entirely divorced from the quality of the underlying content presented in the training. To appreciate the problem with using completion rates, imagine if teachers were evaluated by a similar metric: the effectiveness of the teacher will be based on how many final exams the teacher receives back. As is overtly clear, if a teacher receives back 95% of the total exams passed out to students, this has no bearing on whether the teacher appropriately conveyed concepts or students understood the material. As no one would evaluate teacher effectiveness based on the percentage of final exams turned in, neither should compliance office or prosecutor accept completion percentage as a metric of training effectiveness. Still, the Fraud Section's document does little to dissuade firms from using this metric since it does not provide an appropriate alternative metric or model.²¹

The frequency with which managers present unsuitable responses to questions posed in the Fraud Section's document, such as completion percentage as a means to show training effectiveness, reflects both norms in the compliance industry and firms' uncertainty over how to link data from compli-

20. DELOITTE, IN FOCUS: 2016 COMPLIANCE TRENDS SURVEY 11 (2016), <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/governance-risk-compliance/us-advisory-compliance-week-survey.pdf>.

21. The decision by the DOJ to not provide a specific metric of evaluation is not faulty, but rather a feature of the evaluation questions. In particular, if the DOJ provided a specific metric of evaluation to each question, firms may "game" that metric. Moreover, as discussed *infra* Part VI, univariate metrics cannot alone assess the effectiveness of compliance programs.

ance initiatives to program objectives.²² In this regard, the perceived challenges of measuring the impact of compliance programs today is similar to that faced by marketing executives throughout most the 20th century. While marketing managers could easily quantify the costs associated with an advertising initiative (such as a commercial or magazine campaign), quantifying the initiative's return in sales—the ultimate objective of any marketing initiative—once seemed untenable. As a result, marketing managers used their instincts and guidelines from prior experience to ascertain whether an advertisement effectively boosted sales. By the turn of the 21st century, the field of marketing began to move away from relying on intuitions and developed a variety of models and empirical techniques to assess the effectiveness of marketing campaigns. Today, few firms would embark on a multimillion-dollar advertisement campaign without quantitatively measuring its expected benefits and comparing it against the costs. In contrast, firms spend millions on their compliance products today without the assurance of similar impact assessments. Instead, firms may rely on intuition, attempts to “benchmark” their programs against those of other firms (rather than comparing them against objective standards), and a mixture of metrics that are disconnected from the ultimate objective of their programs. Ultimately, without better models to assess corporate compliance efforts, it is not only infeasible for prosecutors to fairly and consistently differentiate “paper” from effective programs, but it is also difficult for managers to understand the returns from investing in their own compliance programs.

III.

MEASURING COMPLIANCE EFFECTIVENESS

This article seeks to address this problem by proposing a framework to rigorously evaluate the effectiveness of corporate compliance programs. It begins by laying out the three objectives of compliance—to prevent misconduct, to detect misconduct, and to align corporate activities with regulation—and the various compliance initiatives that firms can undertake as part of their programs to support these objectives. Some of the

22. By norms in the industry, firms follow practices they hear from other firms rather than develop appropriate metrics anew. Such practices are supported by the use of benchmarks survey. See DELOITTE, *supra* note 20, at 2.

initiatives, like training and whistleblower hotlines, are already common in the compliance industry. Others drawn from more recent management literature, including advisory hotlines and behavioral nudges, are relatively less known in the field of compliance, but may be employed to support an effective compliance program. After describing the different initiatives available to firms, the article depicts the design of empirical models using data collected from programs to assess whether a compliance initiative is effectively contributing to the objectives of the compliance program. Notably, the framework is not “one size fits all,” but rather a process that can be consistently applied to determine whether the initiatives within a program are supporting the prevention of misconduct, detection of misconduct, or alignment with regulation.

A parsimonious model to rigorously evaluate the effectiveness of compliance programs can benefit both external entities tasked with evaluating programs (i.e., prosecutors, courts, and administrative agencies) and firms themselves.²³ For prosecutors and regulators, the framework permits a consistent assessment of compliance programs across firms despite differences in risk profiles, industry, and program implementation. Despite the presence of significant heterogeneity in firms and their programs, maintaining consistency in assessment is crucial to producing fairness in evaluation. To this end, it is designed to differentiate a “paper” compliance program from one of true substance. The structure of the models also does not need to be kept private (i.e., only known to prosecutors, courts, and regulators) since it does not readily avail itself to gaming of standards.²⁴

23. William Laufer, *Corporate Liability, Risk Shifting, and the Paradox of Compliance*, 52 VAND. L. REV. 1343, 1419 (1999) (describing the value of greater consistency in standards by noting that “moving beyond the rhetoric of ‘effective’ compliance requires consistency between organization assessments of effectiveness and those of regulators, prosecutors, and courts”).

24. If firms do not understand what they are being measured by, it is difficult to understand whether they are satisfying regulation. This lack of awareness of what serves to further benefit or penalize the firms seems especially unusual in that this matter relates to criminal law. The challenge in disclosure is that providing these metrics could spur the gaming of standards. However, what are proposed here are not specific metrics that can be easily manipulated but rather conceptual designs of measuring a compliance program.

For managers evaluating a program, employing empirical models addresses several challenges. First, managers often struggle with non-revenue generating areas when the return of investment is unclear. By quantifying the impact of investment in compliance initiatives, managers can provide more meaningful data to support and justify their investments in compliance. Moreover, since effectiveness is not binary, to the extent that a firm invests more, managers should understand what they are receiving in return. Second, empirical measurements, such as those in this article's framework, can help managers assess how a program functions and evolves over time.²⁵ Finally, if firms are faced with the need to demonstrate the effectiveness of their program to outsiders (e.g., prosecutors, regulators, or outside investors funding a firm), the models presented here can provide relevant evidence. Specifically, the models describe the kinds of data that managers should be collecting to demonstrate that their compliance program is not a mere "paper" program.

IV.

OBJECTIVES OF CORPORATE COMPLIANCE PROGRAMS

Governments confer upon firms both their ability to operate and certain beneficial rights (e.g., protection of property rights, limited liability).²⁶ Firms, however, often have interests that differ from those of the governments that provide these rights.²⁷ For instance, it may be profitable for a firm to provide financial services to a drug cartel, but servicing this client is averse to the public interest (e.g., against the public's moral norms, disruptive to public security). In return for the privi-

25. Linda Klebe Treviño et al., *Legitimizing the Legitimate: A Grounded Theory Study of Legitimacy Work among Ethics and Compliance Officers*, 123 *ORG. BEHAV. & HUM. DECISION PROCESSES* 186, 190–91, 195 (2014) (describing the challenges compliance officers have internally within firms in gaining and sustaining legitimacy).

26. See, e.g., KARTHIK RAMANNA, *POLITICAL STANDARDS: CORPORATE INTEREST, IDEOLOGY, AND LEADERSHIP IN THE SHAPING OF ACCOUNTING RULES FOR THE MARKET ECONOMY* (2015) (arguing that firms can conduct lobbying activities that can capture regulators in "thin markets").

27. Although maximizing shareholder value is often regarded as the objective of firms, this is only one of several possible objective functions. See Joseph L. Bower & Lynn S. Paine, *The Error at the Heart of Corporate Leadership*, *HARV. BUS. REV.*, May–June 2017.

lege of operating within a jurisdiction, firms agree to abide by certain rules and regulations or else face punishment.

Firms themselves are an amalgamation of individuals who voluntarily work for the entity in return for privileges (e.g., compensation, health care, professional development). Individuals within firms may have goals that differ from, or even conflict with, those of the organizations that employ them. Individuals' self-interested objectives—whether in the form of satisfaction or remuneration—may not align with firms' objectives. In seeking to achieve these goals, individuals may inhibit a firm's ability to operate within the confines of relevant rules and regulations within which the firm is bound to operate. When an individual employee deviates from the rules and regulations governing the firm, the employee places both himself and his firm at risk.²⁸

Compliance seeks to mitigate these differences in interests by better aligning the goals of firms and their employees with the goals of governments.²⁹ Regulatory bodies police both firms and their employees for deviation from laws and policy, but such bodies also recognize the practical limitations of efficiently monitoring the immense amount of activity taking place within firms that could run afoul of regulation. To reduce the burden on regulators while not compromising their desire to observe potential violations, regulators incentivize firms to create their own monitoring and reporting system.³⁰ By encouraging firms to engage in their own self-policing ef-

28. The legal theory of *respondeat superior* holds employers vicariously liable for their employee's tortious conduct when the conduct is in the course of their employment. See generally *Bailey v. Filco, Inc.*, 48 Cal. App. 4th 1552 (1996).

29. Embedded in this statement is the implicit assumption that managers in the firm view compliance with the law as the goal of this alignment. Some organizations (e.g., organized crime) may be more appropriately viewed as seeking to have members of the organization comply with goals that contradict regulation.

30. Baer, *supra* note 6, at 989 (describing how the flexibility afforded to firms in the DOJ's stance is not extended to a firm's relationship with its employees. Specifically, "one of the less discussed ironies of compliance regulation: whereas the government stresses its own flexible stance towards firms, because flexibility presumably encourages cooperation, it cannot and will not delegate the same flexibility to private firms in how they order their relationships with employees. To the contrary, the government's mercy is possible only if its private proxy, the corporate firm, adopts an entirely *unmerciful* stance toward its own employees.").

forts, regulators, prosecutors, and courts hope to better align the interests of firms and their employees with applicable laws, rules, and regulations.³¹

Compliance programs are internal firm structures and processes designed to support firms' efforts to achieve this concurrence. Compliance programs are expected to achieve three objectives.³² First, and most fundamentally, compliance programs seek to prevent misconduct from occurring. Recognizing that firms cannot design programs to prevent all misconduct from occurring (e.g., the conduct of a rogue employee), the second element of programs is a mechanism to detect deviant behavior if it does arise. Finally, programs need policies that align corporate behavior with applicable laws and regulations. Policies should not only describe the conduct that ought to be detected and prevented, but also outline the procedures for appropriate action if misconduct arises.

Although these three objectives—preventing misconduct, detecting misconduct, and aligning behavior with regulation—may appear related and overlapping, each is a distinct objective requiring different compliance initiatives to support it. For instance, employee training serves primarily to prevent misconduct by having individuals understand what the firm expects regarding their actions and professional behavior.³³ Whistleblower hotlines can serve to detect misconduct, but have limited efficacy in preventing misconduct before it arises. Interdependence of the objectives can arise as the capabilities and efficacy of one objective impacts others. For example, if a firm more capably prevents misconduct from initially occur-

31. See Jennifer Arlen, *The Failure of the Organization Sentencing Guidelines*, 66 U. MIAMI L. REV. 321 (2012) (describing the need for firms to internally detect and report wrongdoing and whether the guidelines optimally encourage such efforts).

32. Several memos by the Office of the Deputy Attorney General have described the objectives of compliance programs as part of the "Principles of Federal Prosecutions of Business Organizations." See, e.g., Memorandum from Larry Thompson, Deputy Att'y Gen. to Heads of Dep't Components, U.S. Att'ys (Jan. 20, 2003) (on file with author); Memorandum from Mark Filip, Deputy Att'y Gen. to Heads of Dep't Components, U.S. Att'ys (Aug. 28, 2008) (on file with author).

33. If training also includes components to help employees understand what to do if they observe misconduct by reporting it on a hotline, training could also serve to improve detection.

ring, the compliance program will rely less on detecting and sanctioning misconduct.

If a firm fails to create initiatives to support any of these objectives, the overall program will be imperiled. For instance, if a firm relies on a generic set of internal policies to align employee behavior with applicable laws and regulation, the firm will not be able to prevent and detect the most relevant violations. Consequently, the overall effectiveness of a program is the aggregation of the effectiveness that the firm has in achieving all three objectives.

V.

COMPONENTS OF A COMPLIANCE PROGRAM

The sentencing guidelines, memos by senior officials at the DOJ, and regulatory bodies all ascribe value to compliance programs. These writings do not, however, state the set of specific initiatives that firms ought to create to support an effective program.³⁴ Without specific policy guidelines, firms rely on guidance from a disparate set of third-party sources—including industry surveys, compliance consultants, and attorneys—to guide the set of practices in support of an effective program.³⁵

The possible initiatives a firm could implement continue to evolve as new technology and research suggests better or complementary initiatives to improve compliance. For instance, analytic detection systems which scan communications (e.g., employee e-mail) for potential violations is a recent in-

34. Some legislation alludes to the creation of particular initiatives for some firms (e.g., those with publicly traded securities). For example, Sarbanes–Oxley Act states that firms need systems to handle concerns about dubious financial practices. Securities exchanges may impose their own requirements. Sarbanes–Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (2002) (codified as amended in scattered sections of 15 and 18 U.S.C.). Further, Section 303A.10 of the *Manual for New York Stock Exchange Listed Firms* requires firms to publicly post a code of business conduct. The exception to this is training, which is noted by both the sentencing commission and some regulations (e.g., FCPA). See THOMAS FOX, DOING COMPLIANCE: DESIGN, CREATE, AND IMPLEMENT AN EFFECTIVE ANTI-CORRUPTION COMPLIANCE PROGRAM 51 (2014).

35. Geoffrey Miller, *An Economic Analysis of Effective Compliance Programs* 19 (N.Y.U. Law & Econ. Working Papers, Paper No. 396, 2014) (arguing that regulators should not micro-manage programs given the level of knowledge of the inner workings of specific firms).

novation that became possible only with advances in machine learning and hardware processing speed. Additionally, initiatives such as decision advisory hotlines or behavioral nudging are more recent ideas which are gaining increasing acceptance and adoption.³⁶ The following subsections outline possible initiatives under the three objectives of a compliance program.

A. *Prevention*

Employee Selection: Some individuals have a higher propensity to violate rules and regulations.³⁷ A firm can support prevention of misconduct by screening these individuals out during the hiring process or early stage employment. Background checks and focused questioning during the interview process—in order to assess an individual's enthusiasm for rule compliance—are strategies to prevent individuals from jeopardizing the firm's compliance.³⁸ Firms can also directly test new employees' willingness to comply with regulation. In one recent example, two large investment banks gave newly hired analysts examinations that covered material taught during initial training. The new analysts were told that looking up materials on the internet during the exam was forbidden. Moreover, instructors noted that failure on the exam would not result in firing; rather, the employee would simply be asked to retake the exam. This guidance was meant to foster a low-pressure examination environment. Nonetheless, over twenty analysts looked up answers on the internet during the exam, using the same company-owned computer that they were assigned to take the exam on. The inability of these new analysts to follow a simple rule suggested they would be even more inclined to break rules during a higher-pressure situation and potentially

36. See, e.g., Wellner, *supra* note 7.

37. Various theories in criminology, sociology, and economics suggest that there are different propensities to engage in white-collar crime. See EUGENE SOLTES, *WHY THEY DO IT: INSIDE THE MIND OF THE WHITE-COLLAR CRIMINAL* 47–63 (2016) for an overview of these theories.

38. A lack of self-control has been examined in a body of work exploring repeat white-collar offenders. See, e.g., MICHAEL GOTTFREDSON & TRAVIS HIRSCHI, *A GENERAL THEORY OF CRIME* (1990); Robert Davidson, Aiysha Dey & Abbie Smith, *Executives' "Off-the-Job" Behavior, Corporate Culture, and Financial Reporting Risk*, 117 J. FIN. ECON. 5 (2015); Mark Egan, Gregor Matvos & Amit Seru, *The Market for Financial Adviser Misconduct* (Nat'l Bureau of Econ. Research, Working Paper No. 22050, 2016).

run afoul of actual compliance rules.³⁹ The investment banks terminated these analysts for their misconduct during the test.⁴⁰ This is an example of firms utilizing a test to identify employees who demonstrate an inclination to deviate from the firms' compliance goals.⁴¹

Training: Educating employees about applicable firm rules and policies by which they are expected to abide in their professional conduct can potentially prevent and deter misconduct. It may serve to teach gatekeepers (e.g., financial controllers or accounts payable clerks) their specific duties. Such training is conducted in various formats including group sessions, one-on-one meetings, and web-based sessions. Firms may create the training program in-house or with support from external content providers. The focus of training varies depending on the industry and risk-profile of the firm but may include topics such as bribery, privacy, antitrust, and insider trading.

Decision Advisory Hotline: When encountering an unfamiliar situation that has regulatory implications or a difficult dilemma where the appropriate judgment is unclear, an individual may benefit from guidance. A decision advisory hotline seeks to proactively guide individuals in the midst of decision-making to an appropriate course of action.⁴² Although such hotlines do not identify situations that may require differential judgment, the hotline can intercede prior to an individual making an adverse decision that may have detrimental consequences for both the individual and the firm.⁴³

39. For discussion of fraud triangle, see SOLTES, *supra* note 37, at 84–85.

40. Julia La Roche, *This Is How the Goldman Sachs Analysts Who Got Fired Were Cheating*, BUS. INSIDER (Oct. 19, 2015), <http://www.businessinsider.com/goldman-sachs-analysts-fired-for-cheating-2015-10>.

41. Dennis Campbell, *Employee Selection as a Control System*, 50 J. ACCT. RES. 931 (2012) (describing how the selection of employees can be used as a means to align incentives with organization).

42. PAUL HEALY, V.G. NARAYANAN & PENELOPE ROSSANO, GUIDING PROFESSIONAL ACCOUNTANTS TO DO THE RIGHT THING (Harv. Bus. Sch. Pub., Case No. 9-115-028, 2015).

43. Identifying the moral decision is one of the most challenging components of this decision-making process. See MAX BAZERMAN & ANN TENBRUNSEL, BLIND SPOTS: WHY WE FAIL TO DO WHAT'S RIGHT AND WHAT TO DO ABOUT IT (2011); FRANCESCA GINO, SIDETRACKED: WHY OUR DECISIONS GET DERAILED AND HOW WE CAN STICK TO THE PLAN (2013).

Behavioral Nudging: Organizations can help foster an environment that inclines individuals to select the more appropriate course of action during the decision-making process. Often these approaches are designed as behavioral “nudges” that have gained credence in public policy.⁴⁴ Similar conceptual designs within organizations can motivate behavior that is more compliant with regulation. For example, prompting individuals to consider the appropriateness of an action when they face a decision where competing objectives may motivate dishonesty could help deter people from engaging in misconduct (e.g., signing at the beginning rather than at the end of the document).⁴⁵ In this way, improving the “choice architecture” under which decisions are made within a firm could help prevent misconduct.⁴⁶

B. *Detection*

Monitoring systems: data analytics can help detect deviations from firm policies. The focus of such monitoring (e.g., payments, third-parties, conflicts of interest, contract management, etc.) varies depending on the risks facing the firm.⁴⁷ As

44. See CASS SUNSTEIN & RICHARD THALER, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* (Yale Univ. Press 2008). Tom Haugh argues that behavioral nudging should be used in compliance within organization settings. Todd Haugh, “*Cadillac Compliance*” Breakdown, 69 STAN. L. REV. ONLINE 198, 204 (2017).

45. One prominent example is to have individuals sign at the beginning rather than end of documents. See Lisa Shu et al., *Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-reports in Comparison to Signing at the End*, 109 PROC. NAT’L ACAD. SCI. U.S. 15197, 15197–15200 (2012).

46. An even broader argument for the use of programs and incentives aligned with recent advances in behavioral ethics research is described by Haugh, *supra* note 44 (arguing that adopting programs that mimic the criminal law have serious problems and compliance initiatives should focus on designing them with a greater focus on how employees will be impacted by the programs). For a discussion of the psychological research related to unethical behavior in organizations, see Linda Klebe Treviño, Niki A. den Nieuwenboer & Jennifer J. Kish-Gephart, *(Un)Ethical Behavior in Organizations*, 65 ANN. REV. PSYCHOL. 635 (2014).

47. This field is often referred to as “forensic data analytics.” For a pragmatic discussion of how these system work, see ERNST & YOUNG, *GLOBAL FORENSIC DATA ANALYTICS SURVEY 2016: SHIFTING INTO HIGH GEAR: MITIGATING RISKS AND DEMONSTRATING RETURNS* (2016), [https://www.ey.com/Publication/vwLUAssets/ey-forensic-data-analytics-survey-2016/\\$FILE/ey-forensic-data-analytics-survey-2016.pdf](https://www.ey.com/Publication/vwLUAssets/ey-forensic-data-analytics-survey-2016/$FILE/ey-forensic-data-analytics-survey-2016.pdf), and Aditya Misra, *Proactive Fraud Analytics*, INTERNAL AUDITOR (Apr. 20, 2016), <https://iaonline.theiia.org/2016/Pages/>

the collection and analysis of large datasets becomes increasingly feasible, such systems increasingly include analytics software that proactively identifies risks, which can then be targeted for further investigation.⁴⁸

Whistleblower Hotline: When employees sense or observe actions that deviate from policy, firms can offer individuals the opportunity to report that violation.⁴⁹ Such systems rely on providing employees some combination of telephone hotline, e-mail address, and physical drop boxes to report alleged violations.⁵⁰ Hotlines are often designed to be anonymous and confidential reporting tools that protect callers' privacy. As such systems provide allegations of misconduct, these reports need to be tracked and investigated in order to verify the data acquired by the hotline.

C. *Aligning Corporate Activities with Regulation*

Risk Assessment: Firms that operate across multiple industries, jurisdictions, or business lines face wide-ranging risks in complying with civil and criminal laws, rules, and regulations. A risk assessment attempts to ascertain which facets of the business and its operations pose the greatest risk of deviation. The assessment provides a summary of those risks at the firm or business unit level and assesses the likelihood that the firm could face such risks under its current operations. Risk assessment identifies the various risks employees are likely to face across different parts of the firm, allowing firms to focus resources and attention on those areas.

Formal Code of Conduct and Policies: Codes of conduct describe the behaviors and actions that firms seek to prevent

Proactive-Fraud-Analysis.aspx. Monitoring systems can impose additional costs and crowd out incentives to comply. See, e.g., Donald Langevoort, *Monitoring: The Behavioral Economics of Corporate Compliance with the Law*, 2002 COLUM. BUS. L. REV. 71, 74 (2002); Jennifer Lerner & Philip Tetlock, *Accounting for the Effects of Accountability* 125 PSYCHOL. BULL. 255 (1999).

48. Dan Torpey et al., *Fraud Triangle Analytics: Breaking the Status Quo in E-mail Review*, FRAUD MAGAZINE (May–June 2010), <http://www.fraud-magazine.com/article.aspx?id=4294967554>.

49. Whistleblowing is among the most significant ways of exposing major fraud cases. See Alexander Dyck et al., *Who Blows the Whistle on Corporate Fraud?*, 65 J. FIN. 2213, 2214 (2010).

50. DEBBIE TROKLUS & SHERYL VACCA, COMPLIANCE 101: HOW TO BUILD AND MAINTAIN AN EFFECTIVE COMPLIANCE AND ETHICS PROGRAM (2d ed. 2015).

and policies which individuals are expected to follow as employees of the firm. This code includes the standards that employees should observe when seeking to make decisions about firm policy. In the event that alleged misconduct arises and is detected, the code provides guidance about resolution methods.⁵¹

Accountability: Compliance programs require assessments and monitoring to evaluate whether they are achieving their prescribed goals. Appointing an individual who has chief responsibility for the functioning and effectiveness of the program provides accountability for the program. Moreover, senior parties (e.g., CEO, board members) need to stay apprised of the program's progress to provide the appropriate resources and support.⁵² Beyond ensuring program leadership, accountability includes the set of procedures that firms implement to ensure the accountability of individuals (e.g., terminate, demote, change remuneration) in the event that a deviation from firm policies occurs.

VI.

MEASURING COMPLIANCE PROGRAMS

Measuring the effectiveness of the overall compliance programs is often regarded as infeasible since one cannot observe the level of misconduct that would have occurred in the program's absence. To repeat, as described in Part II, compliance programs have three distinct, but complementary objectives that causally contribute to a program's overall effectiveness – the ability for a program's initiatives to support prevention, de-

51. Prior research on codes of conduct have found limited and often mixed evidence on the efficacy of codes of conduct on behavior although most rely on self-reporting rather than direct analysis of behavior. See Mark Schwartz, *The Nature of the Relationship Between Corporate Codes of Ethics and Behaviour*, 32 J. BUS. ETHICS 247, 249 (2001); Gary Weaver et al., *Corporate Ethics Practices in the Mid-1990's: An Empirical Study of the Fortune 1000*, 18 J. BUS. ETHICS 283, 287 (1999); Donald Cressey & Charles Moore, *Managerial Values and Corporate Codes of Ethics*, 25 CALIF. MGMT. REV. 53 (1983).

52. Following *In re Caremark Int'l Inc. Derivative Litig.*, directors face the possibility of civil liability for failure to implement appropriate compliance systems. In particular, directors must "exercise a good faith judgment that the corporation's information and reporting system is in concept and design adequate to assure the board that appropriate information will come to its attention in a timely manner as a matter of ordinary operations." 698 A.2d 959, 970 (Del. Ch. 1996).

tection, and alignment. By focusing on these narrower objectives that indicate the program's overall effectiveness, more precise measurement of a program's impact becomes feasible.

A foundational management principle is that one cannot manage what one cannot measure. Thus, if a program or initiative cannot be measured, it cannot be effectively managed.⁵³ Metrics are designed to provide firms and their managers with the ability to assess progress and, when needed, to diagnose problems and create remedies. To accurately assess performance, measures need to closely track the desired outcome. To the extent that a change in a measure only weakly relates to the outcome, the measure does not adequately capture the relation between actions taken and desired objective.

Measurement designs in compliance programs often violate this principle as firms create a plethora of metrics that are poorly or ambiguously linked to outcomes of failed attempts to capture program effectiveness. As an example, in January 2017, the Office of the Inspector General at the Department of Health and Human Services convened a meeting to develop better metrics to measure the effectiveness of compliance programs within health care organizations. Their report produced over 550 different measures.⁵⁴ Although the report noted that not all of these measures would be appropriate or needed for all programs, ascertaining which metrics are appropriate to evaluate effectiveness remained hopelessly ambiguous. Firms can readily fall prey to tracking vast numbers of different metrics, only some of which actually contribute to understanding the program effectiveness appropriately.⁵⁵ Management

53. See ROBERT S. KAPLAN & DAVID P. NORTON, *STRATEGY MAPS: CONVERTING INTANGIBLE ASSETS INTO TANGIBLE OUTCOMES* (2004) (describing how you cannot measure what you cannot describe, and cannot manage what you cannot measure by focusing on measuring the performance of intangible assets).

54. See DEP'T. OF HEALTH & HUMAN SERVS., OFF. OF INSPECTOR GEN., *COMPLIANCE EFFECTIVENESS ROUNDTABLE, MEASURING COMPLIANCE PROGRAM EFFECTIVENESS: A RESOURCE GUIDE* (2017) (describing, in the aggregate, 582 different compliance metrics that firms could consider using to evaluate effectiveness).

55. The difficulty with interpreting ambiguous metrics is described during interviews with compliance officers. Treviño et al., *supra* note 25, at 192. As one of their compliance respondents noted: "Compliance professions aren't typically judged by probably the majority of metrics that other executives follow. Ours are different, I mean, it's how quickly are we responding to

scholar Robert Simons describes this risk of creating too many performance metrics, noting: “As you add more and more measures . . . you pay an opportunity cost. People have less time to focus on what really matters. The truly critical variables become lost in an overload of extraneous measures.”⁵⁶

A. *Measurement Bias: Differentiating “Paper” Initiatives from Substantive Initiatives*

Appropriate metrics need to accurately track and describe how a compliance initiative contributes to more effectively fulfilling one of the objectives of a compliance program. There is a myriad of ways in which firms can create compliance initiatives—intentionally or accidentally—that have little substance. Clear examples include hotlines that are inaccessible (thus, they do not generate any calls) and generic “off-the-shelf” compliance manuals unrelated to the firm’s business.⁵⁷ Appropriately designed performance metrics should detect these weaknesses and deem the initiative as weak or ineffective. However, there are several common biases which can arise that lead to false or misleading evidence of the effectiveness of an initiative. To further clarify the challenges inherent in designing measures for compliance systems, it is useful to demonstrate the common measurement problems that arise which can falsely suggest that a compliance initiative is effective. To illustrate, each fallacy is presented as a seemingly appropriate response to one of the questions in the DOJ’s Evaluation of Corporate Compliance Programs.

hotline calls, what’s our training completion, our code of conduct. Those are typical metrics or performance factors that are considered by most individuals, so I think those are unique to compliance. It’s kind of harder sometimes to justify our existence when some of those things are a little more nebulous than the flat-out revenue or expense numbers that some of the other people are judged against.” *Id.*

56. ROBERT SIMONS, SEVEN STRATEGY QUESTIONS: A SIMPLE APPROACH FOR BETTER EXECUTION 73 (2010).

57. The Securities and Exchange Commission noted that among the five most frequent compliance topics identified in deficiency letters sent to investment advisory, “off-the-shelf” generic manuals and out of date manuals were two of the most common deficiencies. See Sec. & Exch. Comm’n, Off. of Compliance Inspections & Examination, *The Five Most Frequent Compliance Topics Identified in OCIE Examinations of Investment Advisors*, 6 NAT’L EXAM PROGRAM RISK ALERT, Feb. 7, 2017, at 2, <https://www.sec.gov/ocie/Article/risk-alert-5-most-frequent-ia-compliance-topics.pdf>.

Incomplete metrics: Output measures that capture a limited portion of the desired variation in the output are incomplete metrics and can produce biased inferences. Section 8 of the DOJ's Evaluation of Corporate Compliance Programs asks "Has the company ever terminated or otherwise disciplined anyone for the type of misconduct at issue?" and "Have there been specific examples of actions taken as a result of compliance and ethics considerations?"⁵⁸ To respond to these questions, firms can provide instances of employees who have been terminated, denied promotion, or have faced reductions in remuneration in order to suggest that the firm is holding employees accountable for compliance breaches. The appropriate statistic to assess whether compliance breaches are enforced consistently is the difference between the number of employees involved in recorded violations and the total number of employees who have been disciplined for misconduct.⁵⁹ In particular, understanding the instances in which employees are *not* disciplined is more informative with respect to how a firm enforces accountability for violations. Consider, for instance, a mid-sized firm that presented evidence showing that it sanctioned nine people last year for compliance-related violations. To the extent that the firm only had nine violations last year, this would suggest a high degree of sanctioning related to violations. If the firm had fifty violations and those sanctioned were only the most junior employees with the lowest revenue generating potential, this would differently depict the firm's standards of holding employees accountable. As most firms are continuously making choices about when and how to sanction employees for potential violations of policy, it is necessary to examine the instances in which the firm either does not impose sanctions, provides lesser sanctions, or sanctions lower-lever employees in order to assess the consistency of enforcement.

Invalid metrics: Data collected from an initiative can be uncorrelated with effectively supporting an objective. When there is no apparent causal link between the metric and the objective of the initiative, the metric is invalid. In Section 6 of

58. U.S. DEP'T OF JUSTICE, *supra* note 17, at 6.

59. See U.S. SENTENCING GUIDELINES MANUAL § 8B2.1 (U.S. SENTENCING COMM'N 2015) ("The organization's compliance and ethics program shall be promoted and enforced consistently.").

the Evaluation of Corporate Compliance Programs, the DOJ asks “How has the company measured the effectiveness of the training?”⁶⁰ According to a survey in *Compliance Week*, 50% of responding firms indicated that they used training completion rates to measure their training’s effectiveness. However, as described earlier in the article, there is no causal link between completion and effectiveness. As another example, many firms track total expenditures on a program as part of their evaluation of program effectiveness. However, the amount spent has little causality to the program’s actual effectiveness.⁶¹ While this does not mean that understanding expenditures lacks value, since it serves a legitimate role for corporate budgeting, aggregate expenditures provide little insight into the tracking of program effectiveness.⁶² All metrics to support the effectiveness of a compliance initiative need a causal link between the metric and the objective of the compliance program.

Ex post evaluation metrics: Well-specified metrics provide a contemporaneous assessment of the effectiveness of an initiative. Metrics whose quality can only be assessed in the future (i.e., with the benefit of hindsight) cannot serve as valid metrics and serve only as ex post evaluation metrics. In Section 2 of the Evaluation of Corporate Compliance Programs, the DOJ asks about the steps firm leaders have taken to demonstrate their commitment to compliance. For example, “How have senior leaders, through their words and actions, encouraged or discouraged the type of misconduct in question?” As a metric to indicate leadership’s commitment, a firm may provide the number of speeches and forums where leaders have discussed the value they and other senior leaders place on compliance and ethics. However, these statements and actions are not made in the setting in which they are tested. That

60. U.S. DEP’T OF JUSTICE, *supra* note 17, at 5.

61. It would be reasonable to conclude that no expenditures on a compliance program would imply that it is ineffective through its non-existence. However, greater expenditures could contribute to additional “window dressing” rather than strengthen a program to become more effective. Declining expenditures could imply that the program is being cut and becoming less effective or that the program is simply becoming more efficient while maintaining or even increasing effectiveness.

62. See KRISTY GRANT-HART, HOW TO BE A WILDLY EFFECTIVE COMPLIANCE OFFICER 42 (2016) (providing additional examples, such as “number of compliance policies” and “number of training hours completed on a per-employee basis”).

is, to understand substantive commitment, one would want to see if there is continued commitment to those statements by executives in their private and intrafirm correspondence, i.e., outside of the public spotlight.

Take, for example, a case questioning the impact of Goldman Sachs' public compliance and ethics statements. Executives at Goldman Sachs had long emphasized the importance of business integrity and following both the rule and spirit of regulation in their public statements. In the firm's Business Principles and Standards, for instance, it stated: "we are dedicated to complying fully with the letter and spirit of laws, rules, and ethical principles that govern us. Our continued success depends upon the unswerving adherences to this standard."⁶³ In this case, metrics measuring leadership commitment to compliance would have helped produce a picture of a firm with a healthy approach to compliance and integrity.

In 2010, Ilene Richman, a shareholder, sued Goldman Sachs and its executives for allegedly defrauding the firm's shareholders by not abiding by these statements about the firm's ethics and integrity.⁶⁴ The suit followed a recent admission by the firm that it made a mistake in marketing one of its transactions to clients and paid a \$550 million penalty.⁶⁵ This suit served as an evaluation of the firm's commitment to its prior statements about the standards and principles to which it stated it adhered. Goldman and its executives responded, in an effort to have the suit dismissed, that "the vast majority of the supposed 'misstatement' alleged in the complaint—regarding the firm's 'integrity' and 'honesty'—are nothing more than classic "puffery" or statements of opinion."⁶⁶ By the firm's and its executives' own admission, Goldman Sachs was not substantively committed to those statements at the time they were

63. See GOLDMAN SACHS, *Our Business Principles*, in ANNUAL REPORT 111 (2003), http://www.goldmansachs.com/our_firm/investor_relations/financial_reports/annual_reports/2003/pdf/GS03AR_businessprinciples.pdf. For additional statements by executives, see *Richman v. Goldman Sachs Group, Inc.*, 274 F.R.D. 473 (S.D.N.Y. 2011).

64. *Richman*, 274 F.R.D. 473.

65. Press Release, Sec. & Exch. Comm'n, Goldman Sachs to Pay Record \$550 Million to Settle SEC Charges Related to Subprime Mortgage CDO (July 15, 2010) (on file with author).

66. Brief for Defendant, *Richman v. Goldman Sachs Group, Inc.*, 868 F. Supp. 2d 261 (2012) (No. 10:cv-0395-UA).

made. Yet, any metric created at the time those statements were made would have suggested that they *were* committed. In this way, measures seeking to indicate substantive commitment through “words and actions” can effectively only become ex post evaluation measures once the commitment is tested and one is able to distinguish commitment that is “puffery” from substantive commitment.

Conflating legal metrics with compliance metrics: Compliance creates legal commitments for employees and managers. However, those legal commitments do not necessarily spur behavioral commitment.⁶⁷ For example, in Section 4 of the Evaluation of Corporate Compliance Programs, the DOJ asks: “How has the company assessed whether these policies and procedures have been effectively implemented?” Firms routinely have employees sign agreements affirming that they have read and understood the policies. While such signatures provide the legal basis necessary for certain actions (e.g., firing an employee, reducing benefits with violation, etc.), their existence does not signify or forge an alignment between policy and behavior. Employees may not have actually read the policy (perhaps it is a lengthy document that is quickly signed before moving on to other work), the policy may be vague, the language may be legalistic, technical or dense, or there may be implicit understandings amongst employees—communicated informally and privately within the firm—that the policy does not have to be followed.⁶⁸ As the goal of the policies from an effective compliance program is to align behavior, metrics showing that all employees signed a particular policy is not indicative of whether the policy itself was effective in better aligning behavior with regulation and firm policies.⁶⁹

67. Individuals are able to undermine or overlook legal commitment through a variety of techniques including the use of rationalizations and the employment of loopholes. See SOLTES, *supra* note 37.

68. Some researchers describe this tendency to not appropriately focus as a state of mindlessness. See, e.g., Ellen Langer, Arthur Blank & Benzion Chanowitz, *The Mindlessness of Ostensibly Thoughtful Action: The Role of “Placebic” Information in Interpersonal Interaction*, 36 J. PERSONALITY & SOC. PSYCHOL. 635 (1978).

69. An extension of this concern is when employees interpret legal rules and the absolute boundaries of conduct. Thus, if one finds a loophole or means to circumvent the rule, then the conduct is acceptable. However, this presumes that a firm’s own internal rules and regulations are complete which is seldom the case. See, e.g., Todd Haugh, *The Criminalization of Compli-*

Self-selection and self-reporting bias: Surveying employees to understand their perceptions and knowledge is a popular diagnostic tool for compliance officers. In analyzing survey data, both self-selection and self-reporting bias can confound interpretation. Some sub-groups which may be correlated with the desired measure may choose not to participate (e.g., unethical people may not report seeing rule violation). Similarly, even when surveys are mandated, attention and desire to be accurate can vary by sub-group. Second, an individual's tendency to be honest may be correlated with the type of question. That is, those who are dishonest will not express this sentiment as a survey respondent. Anonymity may help minimize this concern, but respondents may not trust that such anonymity or confidentiality will be maintained at all times.⁷⁰

In Section 6 of the Evaluation of Corporate Compliance Programs, the DOJ asks, "How has the company assessed whether its employees know when to seek advice and whether they would be willing to do so?"⁷¹ To measure employees' willingness to seek guidance about compliance policies, firms will often survey employees. To create a valid survey, the sample of respondents needs to be representative. However, suppose employees who have not encountered violations are those that respond to the survey and they express openness to come forward, whereas those who have seen violations but are less comfortable simply do not respond. Such self-reporting will bias the inferences from the survey. Similarly, if only select groups—by region or seniority—participate, the external validity of the survey will be limited and not applicable to the whole firm.⁷² Consequently, analysis validating the quality of

ance, 92 NOTRE DAME L. REV. 1215, 1265 (2017) (critiquing Intel's compliance program that was viewed as "best practice," but was later noted by the New York Attorney General as where "the actual effect of the program was to school Intel executives in cover-up, rather than compliance"). See also David Yoffie & Mary Kwak, *Playing by the Rules: How Intel Avoids Antitrust Litigation*, HARV. BUS. REV., June 1, 2001, at 119.

70. Firms can set up anonymous reporting systems that can become undermined by actions of management. See, e.g., Ben Martin, *How the Whistleblowing Scandal at Barclays Unfolded*, TELEGRAPH (Apr. 10, 2017, 8:04 PM), <https://www.telegraph.co.uk/business/2017/04/10/whistleblowing-scandal-barclays-unfolded/> (the case of Barclays).

71. U.S. DEP'T OF JUSTICE, *supra* note 17.

72. See THOMAS COOK & DONALD CAMPBELL, QUASI-EXPERIMENTATION: DESIGN & ANALYSIS FOR FIELD SETTINGS (1979).

the data from a survey needs to accompany the interpretation of the metrics produced from such surveys.

B. *Empirically Evaluating Compliance Effectiveness*

Evaluating compliance programs through empirical measurement enhances clarity and reduces discretion in interpreting whether an initiative is progressing in a desired manner. Notably, evaluating the effectiveness of compliance initiatives requires modeling, not benchmarking, programs.⁷³

To illustrate the limitations of univariate comparison of compliance data to other firms (“benchmarking”) consider the annual *Ethics & Compliance Hotline Benchmark Report* produced by one large compliance solution provider. The report is cited as providing “data and analysis to help you benchmark your program . . . and enhance your ethics and compliance program effectiveness.”⁷⁴ The report provides a variety of metrics related to hotlines (e.g., median and range estimates of the report volume per 100 employees). Comparing one firm’s data against these benchmarks provides little insight into a hotline’s effectiveness. For instance, such comparisons ignore the broader differences that make individual firms unique. In particular, an accurate comparison of hotline volumes would need to hold all other factors that could impact volume—size, age, industry, risk profile, and complexity—constant. However, univariate comparisons fail to hold these other dimensions constant during the period of comparison. In this way, comparison against benchmarks often generates more questions than it does unambiguous inferences. If a firm’s call volume is lower than peer volume, it may imply that the firm has fewer violations (an attractive outcome) or it could simply indicate that employees are less knowledgeable or comfortable using the hotline as compared with other firms (an undesirable outcome). Benchmarking does not provide any clear

73. Firms may rely on approaches that appear effective on average. However, such approaches do not take into account differences that may arise between types of firms and employees who may cause effectiveness and ineffectiveness to deviate. Linda Klebe Treviño et al., *Managing Ethics and Legal Compliance: What Works and What Hurts*, 41 CALIF. MGMT. REV. 131 (1999) (providing answers to common questions about what makes compliance effective based on aggregated evidence).

74. NAVEX GLOBAL, 2017 ETHICS & COMPLIANCE HOTLINE & INCIDENT MANAGEMENT BENCHMARK REPORT (2017).

means of separating these two competing hypotheses. Consequently, without an ability to interpret the inferences from the comparison, benchmarking does not facilitate a rigorous evaluation of program effectiveness.

A model to evaluate effectiveness needs to account for differences across time and other factors that can influence output metrics. The model also needs to clearly connect the effort, actions, and resources invested in that initiative with the objective of that initiative. In a number of instances, regression models are well suited to this task for their ability to explain the impact of one variable, while holding others constant. Moreover, the models can be based on data generated from compliance programs to provide a clear and rigorous means of evaluating different initiatives.⁷⁵

To illustrate the process of evaluating programs using regression-based modeling techniques, several different compliance initiatives are examined. Each of these regression models is designed to be run at the firm level with data that is accessible and to produce inferences that connect the specific initiatives to overall compliance effectiveness.

1. *Whistleblower Hotlines*

Whistleblower hotlines offer firms the opportunity to detect misconduct by providing a means for employees to report violations. Interpreting statistics generated from hotlines is regarded as difficult since increases and decreases in hotline volume can be interpreted differently (e.g., high number of calls indicates either more potential violations or, alternatively, greater employee comfort in using the hotline).⁷⁶ To over-

75. For an excellent basic introduction to regression modeling, see JEFFREY WOOLDRIDGE, *INTRODUCTORY ECONOMETRICS: A MODERN APPROACH* (6th ed. 2015).

76. See, e.g., Treviño et al., *supra* note 25, at 192. In their interviews with compliance officers, the authors describe an instance of this difficulty in interpreting hotline data: “If helplines go up, well maybe that’s a good thing, or maybe that’s a bad thing depending on how you look at it. It may be a good thing because more employees are coming to you. Then your boss says, ‘Well, looks like you don’t need any more money for training because you’re properly educating folks. The numbers went up.’ And we may look at it and say, ‘But we need people to handle the calls and deal with those investigations.’ So because we don’t have the same model of history in our metrics that other functions do, it’s not always an easy business case to make when you need additional resources.”

come this ambiguity and eliminate conflicting interpretations, the portion of the variation attributable to differences in alleged misconduct needs to be identified while controlling for the other factors associated with the tendency for employees to use the hotline.

To identify the association between alleged misconduct during a period and call volume, a firm can employ an ordinary least square regression (OLS) model with call data.⁷⁷ The base model specification for evaluating the effectiveness of a whistleblowing hotline would be:

FIGURE 1:

$$\begin{aligned} \text{Volume}_i = & \alpha + \sum_{a=1}^A \gamma_a (\text{Hotline Attributes})_{ai} \\ & + \sum_{b=1}^B \varphi_b (\text{Firm Attributes})_{bi} + \sum_{c=1}^C \beta_c (\text{Period})_t \\ & + \varepsilon_i \end{aligned}$$

As explanatory variables in the model, one would include variables to control for hotline and firm-level attributes. Hotline attributes would include data on the accessibility of calling (or e-mailing) the hotline, responsiveness of the hotline, and prior usage of the hotline. Firm attributes would include data on the number of people that could call the hotline, jurisdictions where the firm operates, and firm performance. Period is an indicator variable of the time frames that assesses the variation in alleged misconduct.⁷⁸ The dependent variable, volume, is a measure of the amount of call volume (in total or by type) during a particular time window. The variable of interest is the coefficient, β , which represents the association between the level of alleged misconduct and hotline call volume. In particular, if the coefficient β is significantly greater than zero in a particular period, the model would indicate that more allegations of misconduct in that period spurred greater hotline call

77. For a discussion of regression modeling, see JEFFREY WOOLDRIDGE, *ECONOMETRIC ANALYSIS OF CROSS SECTION AND PANEL DATA* (2d ed. 2010).

78. The number of periods needs to be less frequent than the data collection period for the model to have sufficient degrees of freedom.

volume. In effect, this would support the idea that the hotline is functioning as desired in that the greater volume of calls reflects people reporting more potential violations and not simply that people are more comfortable with using the hotline or that there are simply more people available to call the hotline. This model is also flexible enough to examine how changes in the hotline (e.g., changing the accessibility, utilizing a new service) influences call volume. By including variables to indicate when those changes occur, one can assess the relative impact of those changes in system processes on hotline call volume.

2. *Training*

Training serves to improve employees' awareness of policies and of potential issues they may encounter, as a means of preventing future violations. Although the ultimate goal of training is to mitigate misconduct by altering behavior, a more immediate goal is to improve understanding of a firm's policies. Since a better understanding of policy may not necessarily translate into changes in behavior (e.g., lack of retention, impact, etc.), assessments of both differences in understanding and in behavior should be examined subsequent to training to evaluate the impact of training.⁷⁹

The quality of understanding firm policies can be assessed through testing. If the assessment would have generated the same results prior to training because the training was ineffective, the assessment was too easy, or the material was already known, then the training did not convey a better understanding of policies to employees. Thus, assessments must compare results before and after training.

Statistical improvements in understanding can be assessed by examining a paired one-sided t-test based on the difference in the pre- and post-learning.⁸⁰ In particular, the null hypothesis is that there is no difference (i.e., training did not impart additional understanding).

79. Greater moral reasoning may translate into improved moral behavior. See LAWRENCE KOHLBERG, *THE PHILOSOPHY OF MORAL DEVELOPMENT: MORAL STAGES AND THE IDEA OF JUSTICE* (1981); but see SOLTES, *supra* note 37, at 360 (showing that behavior may be only weakly correlated with moral development).

80. For a discussion of t-statistics, see, for example, JOHN RICE, *MATHEMATICAL STATISTICS AND DATA ANALYSIS* (2d ed. 1995).

$$H_0: \text{Post Score} = \text{Pre Score}$$

$$H_1: \text{Post Score} > \text{Pre Score}$$

To avoid the potential that ordering effects will bias the measurement, questions and evaluation scenarios should be randomized in the pre- and post-training assessments.⁸¹ That is, from a bank of questions, some questions can be randomly assigned to the pre- and post-training assessment. The model does not specify when to assign the post assessment (or whether there should be more than one). However, an evaluation distributed immediately after the completion of training limits the ability to assess employees' retention of policies over time. Thus, an evaluation that takes place after (i.e., weeks or months) could be employed to better examine the retention of material from training. To the extent that individuals or groups are given different training materials or a new training module is deployed, similar comparisons between groups and old–new training materials can be made with independent t-tests.

The second component to assessing the impact of training in preventing misconduct is to evaluate its impact on behavior. In particular, are individuals and groups that receive training less likely to be involved in an incident or compliance violation? This could be examined with an OLS regression (or a probit model given the binary nature of the dependent variable) to examine the impact of training on violations, which would be modeled as and run on an individual level basis within the firm.⁸²

81. To address this concern, an assessment could be unintentionally designed where the easier questions or those already known are placed in the post-training assessment. In this case, assessments would show improvement, not from training, but rather the difference in question difficulty between the pre- and post-assessment.

82. This model assumes that taking training is uncorrelated with ability to evade detection. Incidents are those that are known to the firm. To the extent that incident does not perfectly detect, additional explanatory variables related to the likely of detection should also be incorporated.

FIGURE 2:

$$\begin{aligned}
 \text{Incident}_{it} = & \alpha + \sum_{a=1}^A \gamma_a (\text{Individual})_{ait} + \sum_{b=1}^B \varphi_b (\text{Group})_{bit} \\
 & + \beta (\text{Training})_{it} + \varepsilon_i
 \end{aligned}$$

As explanatory variables, factors related to the individual include position, gender, age, and past conduct. Group variables would include the propensity of individuals within a group who completed the training and their proclivity for committing a compliance violation. Within a group, it would also be possible to control for the tendency to detect an incident by incorporating frequency of reporting and prior period incidents.⁸³ The variable of interest is the coefficient, β , which represents the association between training and incidents. In particular, if the coefficient β is significantly less than zero, the model would indicate that employees who have undergone training, conditional on other individual and group factors, are less likely to be involved in a compliance violation. Notably, this model can be run incrementally over time as new training is developed as a means of assessing how new training efforts incrementally impact employee behavior.⁸⁴ One caveat to this measure is that to the extent to which a firm is less effective in identifying violations, this measure will be relatively less informative in regards to the quality of training.

83. If the incidence of individual violations is infrequent, the model could be used to assess the violations at the group or unit level within a firm.

84. Learning and behavior can be assessed as whether employees feel subjectively more like the firm is ethical based on their personal observation and whether they understand and are willing to report misconduct. See Gary Weaver & Linda Treviño, *Compliance and Values Oriented Ethics Programs: Influences on Employees' Attitudes and Behavior*, 9 BUS. ETHICS Q. 315 (1999) for a discussion of measures. Using Weaver and Treviño's measures, Danielle Warren, Joseph Gaspar, and William Laufer conducted a univariate comparison of the impact of training over time. The above framework could be applied to these dependent variables, but would suggest employing a multivariate model to control for additional factors influencing respondents over time. See Danielle Warren, Joseph Gaspar & William Laufer, *Is Formal Ethics Training Merely Cosmetic? A Study of Ethics Training and Ethical Organizational Culture*, 24 BUS. ETHICS Q. 85 (2014).

3. *Individual Accountability*

According to the sentencing manual an “organization’s compliance and ethics program shall be . . . enforced consistently throughout the organization through . . . appropriate disciplinary measures for engaging in criminal conduct.”⁸⁵ To align firm policy with the sentencing guidelines and to prevent recidivism, firms need to hold employees engaging in misconduct accountable. Sanctions may include warnings, or reduction in duties, denial of promotion opportunities, adjustment in compensation, or termination. To achieve consistency in accountability across the organization, sanctioning should be based on the nature of the misconduct, rather than on individual characteristics. If, for example, high performers, older employees, or leading revenue-generators are sanctioned differently from other employees for the same offense, then the organization has not achieved consistent accountability.

A regression model can hold the characteristics that may contribute to a sanctioning outcome constant while varying the type of violation, to understand the relationship between different infractions and discipline outcomes. The basic model to describe this relationship is:

FIGURE 3:

$$Discipline_i = \alpha + \sum_{a=1}^A \gamma_a (Individual)_{ai} + \sum_{b=1}^B \varphi_b (Violation)_{bi} + \tau (Interactions) + \varepsilon_i$$

The dependent variable is a discipline outcome describing the nature of the action taken against the employee (e.g., warning, reduction in duties, denial of promotion, adjustment in compensation, or termination). The individual level explanatory variables are those related to the person such as performance, age, gender, and seniority. Violation is a variable describing the type and magnitude of the misconduct and whether it was direct involvement or failure of oversight. To the extent that perceived evidentiary reasons contribute to variation in sanctions, this could also be included as an additional variable.

85. U.S. SENTENCING GUIDELINES MANUAL § 8B2.1(b) (U.S. SENTENCING COMM’N 2015).

The coefficient ϕ would describe the marginal increase in sanctions for a particular offense as compared to the excluded violation. Interaction effects explore the incremental impact of individual characteristics with particular violation. For instance, the marginal impact on a particular discipline outcome based on a particular violation type given an employee's seniority. To the extent that the coefficients on interactions (τ) are greater than zero, the model would indicate that some employees with similar violations receive incrementally harsher ($\tau > 0$) or weaker ($\tau < 0$) sanctions for the same conduct.

4. *Analytic Detection Systems*

Analytic systems examine structured and unstructured data to identify potential misconduct. To the extent that a system can identify misconduct, each system could, in isolation, be viewed as effective. To assess the actual effectiveness of an analytics system, one must consider the externalities such systems impose on organizations more broadly. In particular, analytics potentially impact individual behavior and incentives as well as an organization's ability to investigate allegations of misconduct.

The first potential negative externality to examine is whether the analytic detection system crowds out reporting or adversely alters employee behavior. To the extent that individuals feel less inclined to observe norms on account of the fact that they are closely monitored (e.g., a financial trader will receive an alert if their positions exceed some conditions), conduct itself may become less compliant. Moreover, individuals may feel less of a need to report allegations of misconduct under the pretense that it will be "detected" by the analytics system. Consequently, the individual-level compliance and the desire to report potential misconduct through other channels (e.g., whistleblower hotlines) may decline once such a system is deployed. Whether the analytics system is viewed as effective then is based on whether the detection capabilities offered by the analytics system offset this decline in conduct and reporting.

A second consideration for analytic detection systems is the error rate. To the extent that data analytics are imperfect, the system will identify some false positives (Type 1 errors) and some false negatives (Type 2 errors). Type 1 errors consume investigation time, but ultimately are not cases of misconduct.

Type 2 errors are instances in which the detection software misses an explicit instance of misconduct. To the extent that a compliance division relies on analytics to detect compliance violations, Type 1 errors represent an inefficiency that diverts resources and reduces compliance quality while Type 2 errors create compliance failures through the reliance on imperfect analytic detection systems.⁸⁶

Ultimately, the overall effectiveness of an analytics system can be shown by how much previously undetected misconduct is detected now. At the same time, it is also necessary to consider whether there is misconduct that goes undetected (i.e., change in norms, individuals changing reporting practices, inefficiencies in compliance created by Type 1 and Type 2 errors) because of the analytics system in use.⁸⁷ Similar to the evaluation of training deployed to different groups, a regression model which compares changes in incidents and reporting between those with and those lacking detection systems (or archival data for the same division during, before, and after the deployment of the system) would provide insight into whether the aggregate impact of the system enhances or detracts from overall detection. To the extent that an analytics system leads to improvements in detection even after taking into account the potential adverse behavioral impacts, it would be appropriate to view such system as effective.

VII.

HOW “EFFECTIVE” IS EFFECTIVE?

Compliance effectiveness is not a binary condition. All programs have potential faults and the federal sentencing guidelines recognize that even effective programs can fail.⁸⁸ While a nonexistent program would be, by definition, entirely

86. For a more detailed discussion of Type 1 and Type 2 errors, see RICE, *supra* note 80, at 300–03.

87. Firms may utilize analytics systems to make compliance more efficient, which then makes Type 2 errors salient. That is, to the extent that a firm no longer conducts the same investigations as before and instead relies on the analytics system to guide its investigations, compliance violations that were detected previously and in the absence of the analytics system may no longer be detected, creating the potential for lower compliance when Type 2 errors arise.

88. U.S. SENTENCING GUIDELINES MANUAL § 8B2.1 (U.S. SENTENCING COMM’N 2004) (“The failure to prevent or detect the instant offense does

ineffective, even programs that are created with the intent of “window-dressing” could be marginally effective. Thus, the state of a program’s effectiveness lies on a continuum, with some firms’ programs being relatively more effective and some being relatively less effective. The question then is, how effective does a program need to be to compel prosecutors, courts, and regulatory agencies to offer the firm credit for its design? Put differently, how well does a compliance program need to prevent misconduct, detect misconduct, and align behavior with polices such that a failure of compliance (i.e., a regulatory violation) should be viewed as anomalous?

Memos written by numerous deputy attorneys general state that the evaluation of compliance programs should rest on “whether the program is adequately designed for maximum effectiveness in preventing and detecting wrongdoing.”⁸⁹ However, “maximum effectiveness” is an impractical and infeasible standard by which to evaluate firms’ compliance programs. Consider the fact that several professional services firms now offer advanced analytics programs that can cost millions of dollars to implement. Although the vast majority of firms would find at least some benefit to implementing such advanced analytics in their compliance programs to detect fraud, the costs for most firms are impractically high at present. Should a smaller private firm face higher sanctions, and receive less credit for their program, simply because they did not deploy the latest multimillion-dollar solution designed to detect early stage fraud in e-mail? If this was the accepted condition, only the largest multinational firms are capable of creating a legally sufficient compliance program. To avoid creating biases that favor crediting some firms over others, the criteria to evaluate effectiveness of programs needs to account for the fact that firms of different scales have different feasible options, some of which must be more limited in scope. Consequently, an appropriate evaluation of program effectiveness has to recognize differences in what firms can reasonably expect to implement based on their economic profile as well as

not necessarily mean that the program is not generally effective in preventing and detecting criminal conduct.”).

89. See Memorandum from Larry D. Thompson, *supra* note 32; Memorandum from Eric H. Holder, Deputy Att’y Gen., to all Component Heads and U.S. Att’ys (June 16, 1999) (on file with author); Memorandum from Mark Filip, *supra* note 32.

the fact that different types of firms have different compliance needs based on their industry, risk-profile, and location of operations.

To flexibly evaluate heterogeneous programs in a consistent manner, there are two necessary conditions that firms should satisfy: (1) the existence of initiatives to support each of the three objectives of an effective program and (2) rigorous evidence in support of the effectiveness of each of those initiatives. Conceptually, this approach supports “minimally sufficient” compliance program effectiveness.

To assess a compliance program, the first of the criteria evaluates whether the firm has an initiative that supports each of the three objectives of a compliance program. A firm that has a hotline, for example, but does not train its employees about policies and identifying misconduct, succeeds on detection but fails on prevention. Notably, this approach recognizes that firms do not need to implement every conceivable initiative to support each compliance objective. In practice, firms may not implement an initiative because they believe the costs are prohibitive or because the initiative is redundant to other initiatives. This approach recognizes that different firms will design their programs in different ways, and thus an evaluation of effectiveness needs to be flexible on the specific components of a program. Nonetheless, firms may not compromise on the objectives of a compliance program. To achieve minimal effectiveness, a program needs to have at least one compliance initiative supporting each of the three compliance objectives.

The second criterion for determining whether a firm’s compliance program is at least minimally effective is to have rigorous evidence that each component of the program is effective. That is, for each of the initiatives of the compliance program, the firm must be able to provide data and supporting analysis to indicate how it contributes to preventing, detecting, or aligning policies. In particular, this requirement differentiates compliance initiatives that exist “on paper” from those that actually serve to enhance compliance. As an example, a firm could hire an external service provider to create a whistle-blowing hotline to satisfy the creation of a compliance initiative to detect misconduct. Yet, if the firm does not communicate the availability of this hotline to employees or fails to investigate reports, then the hotline is an ineffective initiative

and does not support detection. As a result, a firm would not receive credit from prosecutors, courts, or regulators for instituting such a hotline.

Qualitative descriptions, benchmarks, and quantitative metrics that are not linked to outcomes cannot be interpreted as showing the effectiveness of an initiative since such assessments do not measure the initiative's impact. A firm that does not rigorously measure the effectiveness of its initiatives cannot evaluate their effectiveness. Notably, a lack of rigorous measurement not only inhibits outsiders (e.g., prosecutors, courts, or regulators) from ascertaining the effectiveness of a compliance initiative, but internal firm leaders themselves will also be unable to ascertain the program's impact.

The overall costs of a firm's compliance program should not be directly considered in evaluation for several reasons. First, if external examiners (e.g., prosecutors or regulators) examine expenditures as part of their evaluation, it implicitly encourages firms to focus on size and number of initiatives, rather than their effectiveness. To the extent that one firm can create an effective system that is less costly than another, equally effective, system, they ought to be rewarded for their efficient use of resources.⁹⁰ There is no *ex ante* reason to assert that the amount spent on compliance necessarily contributes to a more effective program.⁹¹ Second, expenditure-based evaluations discourage innovation in compliance. To the extent that firms are expected to spend at least as much as others in their industry or risk-class, this will only ratchet up program costs as new compliance solutions are created.⁹²

These two criteria offer a means to evaluate a program from the standpoint of external parties seeking to provide a consistent evaluation of programs. One concern, especially if the aggregate compliance expenditures are excluded from the

90. Lower costs could be driven by developing a less costly solution or culture (i.e., not direct cost).

91. Excessive spending and over-compliance may actually lead to less effective compliance. *See supra* note 61 and accompanying text.

92. Monitoring programs that increase employee surveillance, for example, may cause employees to try to bypass the system, leading firms to create even more advanced monitoring systems. *See* Robert B. Cialdini, *Social Influence and the Triple Tumor Structure of Organizational Dishonesty*, in *CODES OF CONDUCT: BEHAVIORAL RESEARCH INTO BUSINESS ETHICS* 44–57 (David M. Messick & Ann E. Tenbrunsel eds., 1996).

evaluation, is that such an assessment by prosecutors, courts, and regulators could lead to underinvestment in compliance programs.⁹³ In particular, focusing on a narrower set of compliance initiatives that can be more rigorously evaluated empirically could foster “legal compliance” at the expense of other initiatives focused on building integrity and culture that are critical to supporting firm wide compliance.⁹⁴ As Lynn Paine has observed, “even in the best cases, legal compliance is unlikely to unleash much moral imagination or commitment. The law does not generally seek to inspire human excellence or distinction. It is no guide for exemplary behavior.”⁹⁵

VIII.

THE CONCERN OF UNDERINVESTMENT

If regulators, courts, and prosecutors focus greater emphasis on rigorously measuring the impact of a narrower set of compliance initiatives (i.e., those that can be measured with data and then empirically evaluated) to evaluate program effectiveness, it will not lessen a firm’s effort or investment in compliance for three reasons.

First, even by focusing on a subset of all possible compliance initiatives to rigorously assess whether a firm has a minimally effective program, it would spur compliance efforts that are likely to equal or exceed many firms’ current programs.

93. This does not mean that costs ought to be included as noted. An additional reason to exclude the level of expenditures in an external evaluation is that it avoids having firm’s compliance spending efficiency become a direct function of the success of regulators and prosecutors in holding firms accountable for wrongdoing.

94. A potential limitation is that the model that relies on empirical assessment narrows the scope and type of initiatives that can show demonstrable effectiveness. Some initiatives and choices by a firm to support impactful compliance programs would thus be excluded. For instance, the commitment of leadership could not be demonstrated. While this may appear as a limitation of more rigorous measurement, this avoids the misclassification of initiatives based on biased measurement. *See supra* Part V. Moreover, from a prosecutor’s perspective, since the firm has failed the prosecutor can assume that system is imperfect, and so will assess the minimal level of program efficiency (rather than the maximum). This does not mean, as will be discussed, that more effective programs are not in the interest of the firm, but rather that these other conditions are not evaluated as part of the prosecutors’ or courts’ evaluation of a minimally sufficient effective program.

95. Lynn Paine, *Managing for Organization Integrity*, HARV. BUS. REV., Mar.–Apr. 1994, at 111.

For example, many firms currently provide training and believe that their training is effective in light of high completion rates. As has been stated by many commentators across a variety of different training approaches, much of this training is in fact ineffective.⁹⁶ When appropriately measured, many firms are likely to find that they need to revise and improve their training efforts in order to make it impactful.

More significantly, although focusing on the empirical evaluation of a select number of initiatives appears to narrow the scope of what is needed for an effective compliance program, many of these initiatives require additional embedded, organizational support. For example, while “tone at the top” may not be an initiative in itself, to the extent that a firm’s executives do support compliance, this will be implicitly captured in the effectiveness of other initiatives. In contrast, if a firm creates a code of conduct that is not supported by senior leaders, managers will find that employees don’t actually know how to interpret the code, which will impact behavior that is picked up in the empirical assessment.

Second, resources are constrained within organizations and officers and directors appropriately want to understand the value of compliance spending. Without measurement, it is difficult to sustainably justify spending on compliance and often programs become a function of the particular leaders within a firm. The quality of compliance thus may fluctuate from chief executive to chief executive. However, the waxing and waning of support for compliance undermines its effectiveness. Measuring what works in compliance helps support more sustainable programs since the impact of that investment is clear.

Finally, observe that “minimally effective” programs (as evaluated by prosecutors, courts, and regulators, when evaluating a firm in regard to potential reductions in sanctions) are not optimal programs from firms’ standpoint.⁹⁷ To see this,

96. The failure of compliance training may be strongest in the area of harassment. Claire Suddath, *Why Can't We Stop Sexual Harassment at Work?*, BLOOMBERG BUSINESSWEEK (Nov. 28, 2016), <https://www.bloomberg.com/features/2016-sexual-harassment-policy/#/>.

97. Formally adhering to the legal codes in a jurisdiction in a successful, albeit “technical” manner does not insulate a firm from the true costs associated with misconduct. As a result, rules-based compliance programs that focus simply on specific behaviors to be avoided are insufficient. *See e.g.*, Lynn

one only needs to see that the costs associated with misconduct are greater than those imposed by the DOJ, courts, or regulators. In particular, firms face private litigation costs, individual directors and officers face personal liability and reputation costs, and the firm faces reputational costs. In cases of misconduct, these non-criminal costs are often many times greater than the direct costs imposed by regulators or courts. Thus, the optimal level of compliance effectiveness is not simply that associated with reduction in criminal or reputational monetary penalties since many of the other costs still remain. The optimal level of investment for firms exceeds the expected direct costs associated with the regulatory/criminal fines.⁹⁸ Thus, while a firm could have a sufficiently effective program to satisfy regulators, prosecutors, and courts, this could still be an underinvestment in forestalling the broader negative consequences associated with compliance related failures.

IX.

CHANGING COMPLIANCE MEASUREMENT

The preceding sections describe a series of principles that managers, courts, and regulators can employ to better measure compliance effectiveness. The lack of a critical inquiry is why more effort has not been expended on improving measurement. Much of the regulation, sentencing guidelines, and memorandums from the DOJ that promote measurement is not new. In fact, in 1995 the United States Sentencing Commission (“USSC”) held a multi-day conference titled “Corporate Crime in America: Strengthening the ‘Good Citizen’ Corporation” in which senators, practitioners, and academics speak about improving compliance. At this conference, numerous participants raised the measurement issue.⁹⁹ However, more than two decades later, many firms have not advanced

S. Paine, Rohit Deshpandé & Joshua Margolis, *A Global Leader’s Guide to Managing Business Conduct*, HARV. BUS. REV., Sept. 2011.

98. Under both conditions, the probability of detection is assumed to be held constant.

99. Richard P. Conaboy et al., *Corporate Crime in America: Strengthening the “Good Citizen” Corporation*, Proceedings of the Second Symposium on Crime and Punishment in the United States (Sept. 7–8, 1995), https://www.uscc.gov/sites/default/files/pdf/research-and-publications/research-projects-and-surveys/economic-crimes/19950907-symposium/WCSYMPO_opt.pdf.

beyond the primitive measures that some suggested when the guidelines were created and discussed during USSC conference. What accounts for this lack of progress?

One explanation is that neither firms nor prosecutors possess a genuine desire to actually improve compliance. If there is no desire to manage programs more effectively, it would be no surprise that there is little desire to measure them more rigorously. William Laufer persuasively argued this view in describing the incentives of different parties in the compliance field. In this argument, firm managers, despite public language expressing a desire to improve compliance, really seek to do only the minimum needed to avoid sanctions. Moreover, while prosecutors may state a desire to enhance compliance, the fact that the DOJ recently hired only one individual to assist in evaluating programs underscores a timid and delayed response to the issue. As Laufer explains about the new DOJ memo describing the evaluation of effective compliance, “like all of the other nods and winks about what really matters to prosecutors. This is nothing more than an additional round in a pretend game of evaluation science with an ultracrepidarian’s hand.”¹⁰⁰

The lack of clear progress in the decades since “effective compliance” entered the nomenclature appears to offer some immediate support to this claim. Generally, other firm departments (marketing, for example) that historically did not measure their efficacy have long since advanced and incorporated empirical models to ensure the department’s effectiveness and efficiency. One would be hard pressed to find another area within a firm that consumes as much money and time as compliance while being subject to nearly zero internal scrutiny. While this could be interpreted as a lack of desire on the part of senior management, an alternative explanation suggests that it is a lack of appropriate knowledge that impedes compliance measurement and innovation.

Compliance programs are often viewed as sub-departments of a firm’s legal program. In many instances the chief compliance officer reports to the general counsel, and the budget for a compliance program is a component of the firm’s legal budget. Compliance programs’ expertise is also rooted in

100. See William S. Laufer, *A Very Special Regulatory Milestone*, 20 U. PA. J. BUS. L. (forthcoming 2018).

law as compliance officers are usually trained as attorneys.¹⁰¹ On the surface, the underlying legal focus of compliance programs and their leaders is due to the focus on complying with regulation and the law.

While the legal background provides the appropriate skills to understand the regulatory environment, attorneys' training and experience tend to exclude other critically important skills. Many attorneys, for instance, do not feel comfortable handling quantitative data. Few law schools offer empirically focused courses, and practicing law does not require understanding empirical data to any significant degree.¹⁰² Moreover, attorneys and legal departments within firms tend not to be measured in terms of efficiency and effectiveness like other parts of the firms are. Thus, the legal function differs from virtually every other division in a firm, wherein most senior leaders and managers have some capacity to work with data and tend to be measured with empirical metrics. Due to the lack of experience in developing empirical models and working with data, those who lead compliance programs are often unaccustomed to thinking in terms of measuring effectiveness and designing rigorous empirical models to support that measurement.

Left on their own, firms themselves are unlikely to instigate significant changes in compliance. To the extent that prior practices "worked" and seem to be aligned with other firms', risk-averse firms will not seek change. On the other side, prosecutors have limited time and resources. Focusing on corporate compliance falls outside both the primary objective and expertise of most prosecutors. Thus, neither firms nor prosecutors are likely to instigate an improvement in measurement or drastically change compliance efforts. Although this could appear to create an impasse, there are two possible mechanisms that could spur changes that do not rely on the singular efforts of either firms or prosecutors.

101. There is considerable debate among practitioners about who compliance ought to report to (e.g., CEO, Board, or General Counsel). *See, e.g.*, DLA PIPER, 2017 COMPLIANCE & RISK REPORT (2017), https://www.dlapiper.com/compliance_survey/.

102. NAT'L CONF. OF BAR EXAM'RS & AM. BAR ASS'N, COMPREHENSIVE GUIDE TO BAR ADMISSION REQUIREMENTS 2017 (2017), https://www.americanbar.org/content/dam/aba/publications/misc/legal_education/ComprehensiveGuidetoBarAdmissions/2017_comp_guide_web.authcheckdam.pdf.

When compliance was a rather narrow function and of inconsequential cost to firms, it was easy to overlook. But today, as compliance becomes increasingly costly for firms, some people within the firm (e.g., the CEO or the board of directors) have rising desire for compliance officers to justify these costs in order to maintain their budgets. Because of these high costs, the equilibrium whereby firms spend increasingly more on compliance without evidence of its effectiveness is becoming increasingly unstable. In this regard, to support their programs, compliance officers are seeking guidance about how to evaluate effectiveness. Prosecutors are well positioned to provide this guidance, despite their historical reluctance to do so.

As an example, at the Society of Corporate Compliance and Ethics' 2017 conference, four chief compliance officers from large financial services firms held a panel that discussed measuring the effectiveness of compliance training. One of the compliance officers, whose firm faced sanctions from the DOJ, sought feedback on her firm's revised training. The training was reduced significantly in length, but was believed to have greater impact on employees. She described seeking the DOJ's feedback on the revised training since it had previously been an area that drew scrutiny. Unfortunately, in her opinion, the DOJ was not willing to provide feedback other than to note the training needed to be effective.¹⁰³ This example illustrates an opportunity wherein prosecutors could help guide standards about what effective training entails (e.g., pre/post changes in learning, impact on behavior). Prosecutors need not provide an opinion on "effectiveness", but could simply educate and guide firms on how to more rigorously evaluate effectiveness their initiatives.

Thus, one path to improving compliance measurement is through the DOJ responding to firms' desire for greater transparency in how to measure compliance effectiveness. To the extent the DOJ offered guidance about the process, it would encourage firms to begin engaging in that measurement themselves.

103. Gary Collins, Managing Director, Compliance Management Division, BNP Paribas; Cassandra Knight, Head of Company Compliance, Morgan Stanley; Melinda Miller, VP Manager Regulatory Compliance, HSBC, Transforming Compliance & Ethics Learning – Lessons from 4 Financial Services Global Leaders, Remarks at the Society of Corporate Compliance & Ethics 16th Annual Conference (Oct. 18, 2017).

An alternative way to increase compliance measurement is through corporate monitors. As part of deferred prosecution agreements, monitors supervise firms and help assure that a firm is compliant with current regulation and aim to reduce the likelihood of future violations.¹⁰⁴ To this end, monitors are often tasked with designing and restructuring the compliance initiatives of a firm. Given the considerable latitude in their decision-making, monitors are well-positioned—from both a leadership and resource perspective—to implement more effective compliance initiatives and to rigorously measure their implementation. As monitors are independent and not guided by any overarching regulatory design, support for more rigorous measurement by monitors would not require any institutional change, but rather could simply arise out of the desire by individual monitors.¹⁰⁵

CONCLUSION

More than two decades after “effective compliance” entered discussions concerning corporate conduct, there continues to be significant ambiguity in understanding what “effective compliance” is on the part of both the legal system (i.e., courts, prosecutors, and regulators) and firms. By defining an effective program as one that 1) has initiatives to support the different components of a compliance program (i.e., prevention, detection, and alignment), and 2) is supported by rigorous evidence of its impact, this article seeks to provide a framework that can be consistently applied by firms and regulators to measure compliance programs.

From a practical standpoint, the compliance field continues to rely more heavily on intuition than empirical measurement to evaluate whether and how effectively programs are working. To the extent that data is used, the evidence is difficult to interpret (e.g., training completion percentage or hot-

104. Vikramaditya Khanna & Timothy Dickinson, *The Corporate Monitor: The New Corporate Czar*, 105 MICH. L. REV. 1713 (2007).

105. The effort to criminalize deviant business arose from the efforts of prosecutors and regulators around the country in the mid to late-20th century before becoming more widely accepted. In this regard, systematic changes in social norms in practices can often be more readily changed by individual actors rather than seeking immediate institutional changes. For a discussion of how the dispersed actions of different prosecutors and regulators changed business conduct, see SOLTES, *supra* note 37, at 33–44.

line volume). Yet, as this article describes, by using some basic models that are widely employed in the evaluation of social programs, marketing, and other business initiatives, the assessment of compliance effectiveness is quantifiable.

For firms, this ability to better assess the impact of compliance programs would not only serve a function when facing potential sanctions, but would more broadly facilitate better management of the immense resources invested in compliance. Many firm leaders continue to invest in additional compliance products and initiatives not out of confidence they are necessarily useful, but out of the fear that not doing so might impose additional costs on the firm. The root of this confusion is the inability to genuinely assess the effectiveness of different initiatives and solutions. To the extent that firms develop the capacity to measure their compliance programs, firms can focus on developing better and more efficient solutions, rather than simply building new systems on top of what has been previously developed.

Ultimately effective compliance programs may take on different forms than those prescribed by some narrowly focused legal teams. As Donald Langevoort has observed:

The natural response of the legally trained—the creation of more aggressive (perhaps lawyer-run) monitoring systems may also generate a host of unexpected costs. Some seemingly obvious cures will be worse than the disease, and that awareness may not come so easily. Compliance is indeed a struggle with no simple check the box solutions. Lawyers must come to understand thoroughly why this is so if they are to be good compliance engineers.¹⁰⁶

Better and more rigorous measurement is the first step to redefining compliance as a field focused on engineering solutions to encourage better behavior, rather than one simply looking for increasingly better ways to protect firms from litigation.

106. Langevoort, *supra* note 47, at 118. For a detailed and thoughtful discussion of the adverse consequences associated with the criminalization of compliance, see Haugh, *supra* note 44.