

Journal of Law & Commerce

Vol. 31 (2012-2013) • ISSN: 2164-7984 (online)

DOI 10.5195/jlc.2013.43 • <http://jlc.law.pitt.edu>

HOW STATISTICAL SAMPLING CAN SOLVE THE CONUNDRUM OF COMPENSATION DISCLOSURES UNDER DODD-FRANK

Michael Ohlrogge



This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License.



This site is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program, and is cosponsored by the University of Pittsburgh Press.

HOW STATISTICAL SAMPLING CAN SOLVE THE CONUNDRUM
OF COMPENSATION DISCLOSURES UNDER DODD-FRANK

Michael Ohlrogge^{*}

ABSTRACT

One of the more controversial measures of the Dodd-Frank bill is its requirement that companies report the ratio of their CEO's compensation to that of their median employee. Critics of this provision have claimed that for large companies with employees and subsidiaries throughout the world, compliance with this measure alone could cost millions of dollars a year, due to the difficulties in identifying the median employee. This paper demonstrates that the Securities and Exchange Commission, which is charged with implementing this provision, has the latitude to direct companies to calculate the figure using a statistical sampling procedure which would greatly reduce the costs of compliance while still achieving a satisfactory degree of reporting accuracy.

^{*} J.D. Candidate, Stanford Law School; Doctoral Student, Economics and Finance—Stanford Department of Management Science and Engineering. I would like to thank Brandon Rees and the AFL-CIO Office of Investment for sponsoring the internship during which I began work on this project, and Vineeta Anand for ongoing discussion as the project continued.

Additionally, I am indebted to Joseph Grundfest and A. Mitchell Polinsky for supervising and sponsoring the independent study in which I fully developed the methodology presented here. Finally, I thank Kathryn Austin for her encouragement and help editing this work.

TABLE OF CONTENTS

I.	Introduction	111
II.	Legal Justifications for a Statistical Sampling	
	Approach to 953(b)	115
	A. Interpreting the Text of Section 953(b).....	115
	B. <i>Chevron</i> Deference and the Use of Statistics in Administrative Agency Rules	116
	C. Section 953(b) Is “Silent or Ambiguous” on the Methods to Be Used in Calculating the Median of the Annual Total Compensation of All Employees of an Issuer	118
	D. The Use of Statistical Sampling to Calculate the Median Is a “Reasonable” Interpretation of Section 953(b) and Therefore Would Receive <i>Chevron</i> Deference	120
III.	Recommended Sampling Procedure and Robustness Analysis	122
	A. Overview of Procedure.....	122
	B. Procedural Details	123
	Conclusion	126
	APPENDIX A: Derivation of Formula for the Level of Confidence Between the kth Smallest and kth Largest Values in a Sample of Size N.....	127
	APPENDIX B: Derivation of Sizes of Confidence Bands for Simulated Companies.....	132

I. INTRODUCTION

Federal regulators have had more than a few headaches as they have sought to implement the remarkably broad but at times maddeningly vague provisions of the Dodd-Frank Wall Street Reform and Consumer Protection Act. The question of how to implement the provisions of Dodd-Frank dealing with executive compensation has proved to be particularly vexatious and controversial. Section 953(b) of Dodd-Frank instructs the U.S. Securities and Exchange Commission (SEC) to amend the executive compensation provisions of Regulation S-K, Item 402, so as to require companies to disclose the median annual total compensation of all employees apart from the chief executive officer, the total annual compensation of the chief executive officer, and the ratio between these two numbers.¹

This provision, which was introduced during the final stages of negotiation over the Dodd-Frank bill and received relatively little attention prior to the bill's passage,² has since become one of the more hotly debated and contentious issues concerning the implementation of the new law. Critics of the provision have claimed that it will saddle corporations with millions of dollars a year in additional compliance costs, hurting competitiveness and discouraging job growth just when the economy needs these most. The crux of the matter lies in section 953(b)'s somewhat peculiar wording, calling upon corporations to calculate median employee compensation using the same comprehensive metric by which they currently tally executive compensation. The exacting requirements of this metric currently encompass more than 40 pages of detailed procedures for tabulating the precise value of nearly every conceivable benefit that a corporation can bestow upon its employees.³

In numerous meetings with SEC officials and hundreds of pages of comment letters submitted to the agency, many companies and trade

¹ Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, Pub. L. No. 111-203, tit. IX, sub. E, § 953(b), 124 Stat. 1904 (2010).

² H.R. REP. NO. 112-42, pt. 1, at 1-2 (2011).

³ 17 C.F.R. § 229.402 (2011).

associations have expressed their concerns with this provision.⁴ They have argued that if taken literally, § 953(b) would require them to apply this exhaustive calculation (which includes such elements as net present value of defined benefit pension plans—a potentially tricky accounting figure to generate) to hundreds of thousands of different employees, working for hundreds of different subsidiaries in dozens of different countries, paid in dozens of different currencies and administered through divergent and incompatible payroll and accounting systems. Many companies have worried that without such exhaustive calculations, it will be impossible to identify the precise median employee—that is, the employee who receives more compensation than 50% of the company’s employees and less compensation than 50% of the company’s employees.⁵

Some of those concerned about the implementation of 953(b) have gone so far as to seek complete repeal of the measure. On March 14th, 2011, Rep. Nan Hayworth (R-NY19) introduced H.R. 1062, the Burdensome Data Collection Relief Act, directed explicitly and exclusively

⁴ See, e.g., Letter from Timothy J. Bartl, Senior Vice President and Gen. Counsel, Ctr. on Exec. Comp., to Elizabeth M. Murphy, Sec’y, U.S. Secs. and Exch. Comm’n (Nov. 11, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-79.pdf>; Written Testimony, Darla C. Stuckey, Senior Vice President, Policy and Advocacy, Soc’y of Corporate Sec’y and Governance Prof’ls (Sept. 24, 2010), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-20.pdf>; and Letter from Mark E. Warren, Vice President, Tax and Fin., Retail Indus. Leaders Ass’n, to Elizabeth M. Murphy, Sec’y, U.S. Secs. and Exch. Comm’n (Oct. 27, 2010), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executive-compensation-43.pdf>. For a full list of submitted comments and meetings with SEC officials, see generally <http://www.sec.gov/comments/df-title-ix/executive-compensation/executive-compensation.shtml>.

⁵ See, e.g., Letter from Timothy J. Bartl, Senior Vice President and General Counsel, Ctr. on Exec. Comp., to Elizabeth M. Murphy, Sec’y, U.S. Secs. and Exch. Comm’n (Sept. 1, 2010), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-8.pdf> (arguing that “[b]ecause the definition of median means ‘midpoint,’ depending on how the phrase ‘all employees’ is defined, companies could be required to calculate pay as specified by the proxy rules for each individual employee globally and then determine the median of those values. For large employers, this means they will have to accurately calculate pay for tens of thousands and in some cases, hundreds of thousands of employees to determine the median.”) See also Letter from Mark E. Warren, Vice President, Tax and Fin., Retail Indus. Leaders Ass’n, to Elizabeth M. Murphy, Sec’y, U.S. Secs. and Exch. Comm’n (Oct. 27, 2010), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executive-compensation-43.pdf> (arguing, “[a]t the extreme, this would require an issuer to calculate the salary, bonus, stock awards, option awards, non-equity incentive plan compensation, change in pension value and nonqualified deferred compensation earnings, and all other compensation for each employee—potentially tens or even hundreds of thousands of individuals for the largest U.S. employers.”).

at repealing § 953(b) of Dodd-Frank.⁶ This bill was favorably reported by the House Financial Services Committee on June 22, 2011, and is currently awaiting a vote by the full House of Representatives.⁷

Supporters of 953(b) have been no less vigorous in their efforts. Numerous institutional investors, representing hundreds of billions of dollars of corporate equity holdings, have for some time been frustrated by skyrocketing rates of executive compensation, which they see as being unrelated to or even sometimes negatively correlated with company performance.⁸ Many such investors, joined by a host of corporate governance and financial accountability organizations and private citizens, have submitted more than two hundred of their own comment letters to the SEC.⁹ Proponents argue that the provision will provide a much-needed tether to reality for executive compensation. They claim that the CEO to median worker compensation ratios will enable investors and executive compensation committees to compare CEO compensation to something

⁶ Burdensome Data Collection Relief Act, H.R. 1062, 112th Cong. (as reported by H.R. Comm. on Fin. Servs., June 22, 2011), <http://www.govtrack.us/congress/bill.xpd?bill=h112-1062>.

⁷ *Id.*

⁸ JIM COLLINS, GOOD TO GREAT: WHY SOME COMPANIES MAKE THE LEAP . . . AND OTHERS DON'T 49 (2001) (noting that amongst the companies he studied who had exceptional stock performance over sustained periods, CEO compensation was on average lower than at companies whose performance was more mediocre.). See also Lucian Bebchuk & Yaniv Grinstein, *The Growth of Executive Pay*, 21 OXFORD REV. ECON. POL'Y 7 (2005) (noting that there is no indication that increases in company size or improvements in performance can explain a large portion of increases in executive compensation). See also Council of Institutional Investors, *Corporate Governance Policies*, 10–11, <http://www.cii.org/CouncilCorporateGovernancePolicies> (adopted Dec. 21, 2011) (noting that executives should not be excessively paid and that their compensation should be “reasonable and rational with respect to critical factors such as . . . compensation paid to other employees.”).

⁹ See, e.g., Letter from Stu Dalheim, Director, Shareholder Advocacy, to Elizabeth M. Murphy, Sec'y, U.S. Secs. and Exch. Comm'n (May 27, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-75.pdf> (noting that a high CEO to median worker compensation ratio can indicate increased risks of employee turnover); Letter from Timothy Smith, Senior Vice President, Director of ESG S'holder Engagement, Walden Asset Mgmt., to Elizabeth M. Murphy, Sec'y, U.S. Secs. and Exch. Comm'n (Apr. 29, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-68.pdf> (noting that “[h]igh levels of disparity [between CEO and median worker pay] may breed mistrust in the workplace that dampens productivity”); see also Letter from Rick Wartzman, Exec. Dir., Drucker Institute, to Elizabeth M. Murphy, Sec'y, U.S. Secs. and Exch. Comm'n (Feb. 17, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-60.pdf> (submitted Feb. 17, 2011) (noting that famed management guru Peter Drucker advocated no more than a 25:1 ration between CEO and average worker compensation, noting that higher ratios impede teamwork, particularly between the CEO and middle-management).

other than simply the sky-high compensation packages of other CEOs.¹⁰ Naturally, these proponents of the provision have also tended to dispute claims by corporations and trade associations that the provision will be prohibitively expensive to implement.¹¹

As is often the case, however, the debate between proponents and opponents of the measure has generated more heat than light. Many of those commenting in opposition to the proposal have urged patently unworkable solutions, such as that the SEC simply ignore the law and implement the provision in ways directly contrary to the specific wording of section 953(b). Thus, for example, one group of compensation consultants wrote to the SEC requesting that the agency implement the law using the mean rather than the median worker compensation, as the law explicitly requires.¹² Likewise, the Center on Executive Compensation, a trade association that deals with matters of executive compensation, requested that the SEC ignore certain aspects of the compensation calculation procedures in Regulation S-K, Item 402, even though the law explicitly requires that such procedures be used.¹³ Proponents of the measure, for their part, have generally just asserted that claims of excessive costs are overblown, while failing to offer a convincing picture of how the law can indeed be implemented at a reasonable cost.¹⁴

This article proposes a way out of the morass that discussions of Dodd-Frank's section 953(b) have currently reached. The Securities and Exchange Commission can implement section 953(b) by instructing

¹⁰ See, e.g., Letter from Timothy Smith, Senior Vice President, Dir. of ESG S'holder Engagement, Walden Asset Mgmt. to Elizabeth M. Murphy, Sec'y, U.S. Secs. and Exch. Comm'n (Apr. 29, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executive-compensation-68.pdf>.

¹¹ See, e.g., Letter from Americans for Financial Reform to Elizabeth M. Murphy, Sec'y U.S. Secs. and Exch. Comm'n (Mar. 23, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-63.pdf>.

¹² Letter from Frederic W. Cook & Co., Inc. to Chairman Shapiro, Chairman, U.S. Secs. and Exch. Comm'n (Aug. 13, 2010), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-6.pdf>.

¹³ Letter from Center on Executive Compensation to Elizabeth M. Murphy, Sec'y, U.S. Secs. and Exch. Comm'n (Sept. 1, 2010), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-8.pdf>.

¹⁴ See, e.g., Letter from Walden Asset Management to Elizabeth M. Murphy, Sec'y, U.S. Secs. and Exch. Comm'n (Apr. 29, 2011), <http://www.sec.gov/comments/df-title-ix/executive-compensation/executivecompensation-68.pdf>.

companies to identify their median employee through a process of statistical sampling. Such sampling is already widely used by numerous other administrative agencies, and the Supreme Court has recently upheld the discretion of agencies to determine how technical measures such as medians are to be calculated. In addition to laying out the case for the legal permissibility of such an interpretation by the Securities and Exchange Commission, this article proposes the outlines of a specific sampling methodology which would allow a satisfactory degree of precision while minimizing compliance costs.

II. LEGAL JUSTIFICATIONS FOR A STATISTICAL SAMPLING APPROACH TO 953(B)

A. Interpreting the Text of Section 953(b)

Section 953(b) instructs the SEC to amend 17 C.F.R. § 229.402 (Regulation S-K, Item 402) to require issuers to disclose, in any filing described in 17 C.F.R. § 229.10(a) (Regulation S-K, Item 10):

(A) the median of the annual total compensation of all employees of the issuer, except the chief executive officer (or any equivalent position) of the issuer; (B) the annual total compensation of the chief executive officer (or any equivalent position) of the issuer; and

(C) the ratio of the amount described in subparagraph (A) to the amount described in subparagraph (B).¹⁵

The Act further states that “[f]or purposes of this subsection, the total compensation of an employee of an issuer shall be determined in accordance with section 229.402(c)(2)(x) of title 17, Code of Federal Regulations, as in effect on the day before the date of enactment of this Act.”¹⁶ CFR Section 229.402(c)(2)(x) provides a formula for the disclosure of chief executive officer total compensation in company proxy statements.

Section 953(b) notably calls for disclosure of “median” employee compensation levels. In contrast to the average or “mean” statistic, the

¹⁵ Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, *supra* note 1.

¹⁶ *Id.*

median is considered to be a more reliable statistical description of skewed distributions, such as a sample of income data when there are highly compensated outliers in the population.¹⁷

While Section 953(b) is clear on how the total compensation of an employee is to be calculated, it does not specify how the “median” is to be measured. As discussed below, Section 953(b)’s silence on the term “median” gives the SEC the necessary regulatory flexibility to permit companies to use statistical sampling to determine the median.

B. Chevron Deference and the Use of Statistics in Administrative Agency Rules

The U.S. Supreme Court has recognized a two-step test under which courts are to defer to the interpretations of a law supplied by an administrative agency tasked with implementing that law. Under the *Chevron* doctrine, a reviewing court will ask first whether the meaning of a statute is clear or whether “the statute is silent or ambiguous with respect to the specific issue.”¹⁸ If the statute is silent or ambiguous, the court will secondly ask whether the agency’s interpretation is “reasonable” and whether it is “based on a permissible construction of the statute.”¹⁹ If both of these requirements are met, the court will defer to the agency’s interpretation of the statute, rather than supplying the court’s own.²⁰

In *Zuni Pub. Sch. Dist. No. 89 v. Dep’t of Educ.*, the Supreme Court established that *Chevron* deference is pertinent when an administrative agency issues regulations that specify the procedure by which a statutorily mandated statistical test is to be performed. In *Zuni*, the statute in question called for the Secretary of Education to make certain funding decisions that required identifying school districts in each state “with per-pupil expenditures . . . above the 95th percentile or below the 5th percentile of such expenditures . . . in the State.”²¹ Because the law at issue in *Zuni* did not specify how these percentiles were to be calculated, the Secretary of

¹⁷ See GEORGE W. SNEDECOR & WILLIAM G. COCHRAN, STATISTICAL METHODS 136 (8th ed. 1989).

¹⁸ *Chevron, U.S.A., Inc. v. Natural Res. Def. Council*, 467 U.S. 837, 843 (1984).

¹⁹ *Id.*

²⁰ *Id.* at 843.

²¹ *Id.* at 84 (quoting 20 U.S.C. § 7709(b)(2)(B)(i) (2000)).

Education promulgated regulations stating that the percentiles were to be calculated by ranking every student in the state based on how much funding that student received, and then finding which districts the lowest 5% and the highest 5% of students were in and excluding those.²²

The plaintiffs in *Zuni* were several school districts that received less funding under the Secretary's method of calculating percentiles than they would have under an alternate methodology, which they claimed was the only permissible reading of the statute.²³ The plaintiffs argued that this statutory language required the Secretary to rank the *districts* in order from lowest per-pupil funding to highest and calculate the percentiles treating each district as a single unit, regardless of the number of students in it.

Nevertheless, the Court found that the statute, by simply calling for a calculation based on certain percentiles, left sufficient ambiguity to permit the agency to issue regulations providing procedures for calculating those percentiles.²⁴ As part of its justification for this stance, the Court quoted from a number of dictionaries and professional manuals to indicate that the term "percentile" is open to a variety of meanings and calculation techniques.²⁵ Notably, one of these sources, quoted directly by the Court, was the *Concise Oxford Dictionary of Mathematics*, which stated that in some instances, percentiles can be "applicable to . . . a large sample ranked in ascending order."²⁶

The Court further found that the procedures provided by the regulation were a reasonable interpretation of the statute.²⁷ In support of this conclusion, the Court reasoned that calculation methods such as this are "the kind of highly technical, specialized interstitial matter that Congress often does not decide itself, but delegates to specialized agencies to decide."²⁸ Finally, the Court examined the legislative history of the statute, and found nothing to suggest that the Secretary's interpretation and

²² *Id.*

²³ *Zuni*, 550 U.S. 81, 88–89 (2007).

²⁴ *Id.* at 99.

²⁵ *Id.* at 95.

²⁶ *Id.* (quoting CHRISTOPHER CLAPHAM & JAMES NICHOLSON, *CONCISE OXFORD DICTIONARY OF MATHEMATICS* 378–79 (3d ed. 2005)).

²⁷ *Id.*

²⁸ *Id.* at 90.

calculation technique would run counter to the intent of Congress in passing the law.²⁹

C. Section 953(b) Is “Silent or Ambiguous” on the Methods to Be Used in Calculating the Median of the Annual Total Compensation of All Employees of an Issuer

Section 953(b), by calling for reporting of median employee compensation but not specifying how that median is to be calculated, “is silent or ambiguous with respect to [that] specific issue.”³⁰ Where the drafters of Section 953(b) wanted to be explicit about calculation methodologies, they clearly knew how to do so. Section 953(b)(2) specifically provides for the exact values to be used in calculating the total value of the median employee’s compensation. That Congress did not use this same level of specificity to describe the statistical procedures for identifying the median suggests that this is “the kind of highly technical, specialized interstitial matter that Congress often does not decide itself, but delegates to specialized agencies to decide.”³¹

The only way in which this language could be viewed as specific enough to provide the “unambiguously expressed intent of Congress,”³² would be if the word “median” possessed a single, clear definition that explicitly included procedures for calculating it in a situation such as this. The Supreme Court in *Zuni* recognized that there may be multiple ways of analyzing a large set of multi-variable figures, and thus rejected the notion that a single term could fully encapsulate the complexities of modern statistical practices.³³ As the Court remarked:

We are not experts in statistics, but a statistician is not needed to see what the dictionary does not say. No dictionary definition we have found suggests that there is any *single* logical, mathematical, or statistical link between, on the one hand, the characterizing data (used for ranking purposes) and, on the other hand,

²⁹ *Zuni*, 550 U.S. at 90.

³⁰ See *Chevron*, 467 U.S. at 842–43.

³¹ See *Zuni*, 550 U.S. at 94.

³² *Chevron*, 467 U.S. at 843.

³³ *Zuni*, 550 U.S. at 95.

the nature of the relevant population or how that population might be weighted for purposes of determining a percentile cutoff.³⁴

Moreover, the fact that Section 953(b) calls for the SEC to implement its disclosure provisions by regulation indicates that Congress anticipated the SEC would play a role in shaping the implementation specifics for this provision. Section 953(b) could have bypassed regulation altogether by simply making its disclosure requirements mandatory as a matter of law. Alternatively, Section 953(b) could have provided specific language to insert into the regulation, stipulating that the SEC was to insert that language and no more. Yet, Section 953(b) does neither of these.

Instead, Section 953(b) calls for its provisions to be incorporated into Reg. S-K Item 402, which is itself a complex, yet at times flexible document, created through years of discretionary rulemaking by the SEC. Throughout Reg. S-K, Item 402 the SEC provides clarification for calculation techniques.³⁵ Item 402 also provides for exceptions and flexibility in certain circumstances.³⁶

Indeed, it would seem quite unusual for new provisions to be added to Item 402 in a way that denied the SEC any discretion in issuing the specific instructions and guidance that characterize the document. There is nothing to suggest that this is what Congress intended Section 953(b) require. Thus, in no way does Section 953(b) “unambiguously express[] [the] intent of Congress,” with regards to every methodological detail for performing the statistical calculations it mandates. Rather, Section 953(b) identifies a clear objective of information to be disclosed and then provides great specificity with regards to some aspects of its application while leaving others to the practical experience and technical expertise of the SEC to implement.

³⁴ *Id.* at 96 (emphasis in the original).

³⁵ *See, e.g.*, 17 C.F.R. § 229.402 (2011), Instructions to Item 402(c)(2)(viii) (providing procedures for calculating the value of interest on deferred compensation).

³⁶ *See, e.g., id.* at Instructions to Item 402(a)(3) (exempting reporting on executives making less than \$100,000/year).

D. The Use of Statistical Sampling to Calculate the Median Is a “Reasonable” Interpretation of Section 953(b) and Therefore Would Receive Chevron Deference

Given then that Section 953(b) contains ambiguity with regards to calculating the median, the next question under the *Chevron* analysis is whether the SEC’s resolution of that ambiguity by implementing a statistical sampling methodology would be a “permissible construction of the statute.”³⁷ As society and administrative processes have become more complex and data driven, the use of statistics within government functions has increased. Numerous laws³⁸ and even larger numbers of regulations³⁹ now mention the use of statistical calculations. Because statistics is a vast and often highly technical subject matter, it is not surprising that Congress has generally not attempted to precisely define in legislation how particular calculations are to be performed.⁴⁰

By the standards established in *Chevron* and *Zuni*, a regulation implementing Section 953(b) by use of a statistical sampling technique to calculate median employee compensation would clearly be a permissible interpretation of the Dodd-Frank Act. Again, *Zuni* is very closely analogous in that it upheld an agency’s procedures for calculating a statistic whose use was required by statute.⁴¹ In fact, the Secretary of Education’s regulations in question in *Zuni* permitted the use of statistical sampling for counting the number of pupils in each school district. The regulations provided that the number of pupils in each district, an essential number for determining the

³⁷ *Chevron*, 467 U.S. at 842–43.

³⁸ See, e.g., 12 U.S.C. § 1703 (2006) (referencing median house price); 42 U.S.C. § 300g-3 (2006) (referencing water safety standards calibrated to the 90th percentile of sampling); 42 U.S.C. § 1395x (2012) (capping payments for health services at 105% of the median payments for such services); etc.

³⁹ See, e.g., 40 C.F.R. § 799.9120 (2011) (providing for the use of medians in toxic substance tests); 7 C.F.R. § 246.7 (2011) (restricting certain payments to families below 50% of the median income in an area); 10 C.F.R. § 835.2 (2011) (requiring calculation of median particle size within an aerosol); etc.

⁴⁰ See, e.g., 42 U.S.C. § 8624 (2006) (restricting certain federal aid payments to families below 60% of state median income, but not specifying calculation procedures for that median figure). See also 45 C.F.R. § 96.85 (2011) (a regulation implementing 42 U.S.C. § 8624 and specifying the procedures for calculating the required median figures).

⁴¹ *Zuni*, 550 U.S. 81.

percentiles, was to be calculated “in accordance with whatever standard measurement of pupil count is used in the State.”⁴²

Such state methods in *Zuni* often use sampling and estimation techniques.⁴³ Furthermore, not only did the procedures in *Zuni* also involve the potential use of statistical sampling, the regulation in question actually delegated to a group of third parties, the states, the task of deciding whether and how the sampling would be conducted.⁴⁴ Thus, in many ways *Zuni* upheld a regulation that provided for a far broader range of interpretation of the original statute than would be the case in this situation, in which the SEC could specify the procedures for the statistical sampling rather than delegating the decision to third parties.

Section 953(b) and the statute at issue in *Zuni* are far from the only instances in which statutes call for the use of statistical measures without specifying procedures for their calculation, thereby leaving administrative agencies to fill in the specifics via regulation. For instance, 42 U.S.C. § 8624 provides for a Low Income Energy Assistance Program, funded by the federal government but administered by states. The law specifies that payments cannot go to households with incomes above “60 percent of the State median income.”⁴⁵ The law says nothing more about this “median income” figure or how it is to be calculated. A regulation, however, promulgated by the Department of Health and Human Services, specifies that this calculation can be made using “state median income estimates.”⁴⁶ The federal government, based on a sampling procedure conducted by the U.S. Census, generates these state median income estimates.⁴⁷

⁴² *Id.* at 103; Appendix A.

⁴³ *See, e.g.*, Michigan’s provisions, MCLS § 388.1606(4) (Bender 2011) (providing for estimation of school district populations based on attendance figures for all pupils for every school day statewide, but rather on two annual “pupil membership count day[s]”).

⁴⁴ *Zuni*, 550 U.S. at 103; Appendix A.

⁴⁵ 42 U.S.C. § 8624 (2006).

⁴⁶ 45 C.F.R. § 96.85 (2011).

⁴⁷ *See* State Median Income Estimate for a Four-Person Family: Notice of the Federal Fiscal Year (FFY) 2011 State Median Income Estimates for Use Under the Low Income Home Energy Assistance Program (LIHEAP), 75 Fed. Reg. 26780–82 (May 12, 2010); *see also* U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. CENSUS BUREAU, *Design and Methodology, American Community Survey* (2009), available at http://www.census.gov/acs/www/methodology/methodology_main/ for a description of the Census’ sampling methodology and statistical estimation techniques.

An additional reason why statistical estimation techniques of the median are a reasonable interpretation of Section 953(b) is that Reg. S-K, Item 402, the regulation that Section 953(b) modifies, already makes use of estimation techniques for other figures it requires issuers to disclose. For instance, the instructions to Item 402(c)(2)(vii), Section 2(A), provides for the reporting of the actuarial present value of accumulated pension benefits. Such an actuarial valuation is by definition an estimate, based upon educated guesses concerning life expectancy, interest rates, and so on. The fact that Item 402 already makes use of estimation techniques for reported figures further supports the notion that the additional Item 402 reporting requirements under Section 953(b) could be calculated using estimation techniques.

The use of sampling techniques to implement Section 953(b) is also a reasonable statutory interpretation because it comports with well-established practices in statistical calculations of median income. Calculations of median income are in fact one of the most common instances in which statistical sampling techniques are used to estimate a median value.⁴⁸ As noted above, such sampling and estimation is used by the U.S. Census in its calculations. The Bureau of Labor Statistics also uses a similar methodology.⁴⁹

III. RECOMMENDED SAMPLING PROCEDURE AND ROBUSTNESS ANALYSIS

A. Overview of Procedure

This section presents a simple procedure that will enable any company, through random sampling, to produce an estimate of its median compensation with a high degree of confidence. The degree of precision of this estimate can be made a function of the sample size, and the SEC in consultation with businesses and the public can select the size that best balances the costs and benefits of this mandated reporting.

Any company that wishes to estimate its median compensation can accomplish this by taking a random sample of its employees, calculating the

⁴⁸ See SNEDECOR & COCHRAN, *supra* note 17.

⁴⁹ See *Technical Notes for May 2010 OES Estimates*, U.S. BUREAU OF LABOR STATISTICS, available at http://www.bls.gov/oes/current/oes_tec.htm for a description of its methodology.

compensation for each of those employees, and finding the median of that sample. A company can conduct such a sampling by assigning a unique identifying number to each of the company's employees and then using a computer to randomly select a given number of those identifiers. More complicated procedures, such as stratified sampling, will be unnecessary, regardless of the size of a company, how many countries it operates in, or how many subsidiaries it has.

The sample median generated through this process, for any odd-sized sample, will be a median-unbiased estimate of the company's true median compensation.⁵⁰ This means that the sample median will be no more likely to overestimate the true median than it is to underestimate it, and thus does not contain any "bias" to err in one direction more often than another.⁵¹ Even a sample as small as 199 individuals will enable a company to achieve a 90% confidence level that the true median will be between the 89th and 111th entries in the sample, ranked by compensation.⁵² This will enable most companies to produce an estimate of their true median compensation that is, at a 90% confidence level, within \$1300 of the actual median compensation. Larger samples would allow even greater levels of confidence and precision, as detailed below.

B. Procedural Details

Through the use of basic laws of probability, as outlined in Appendix A, it can be shown that if a company randomly selects a sample of N employees and computes the total compensation for each employee, then the company's level of confidence that the true median compensation will be between the k th largest and the k th smallest value in the sample is given

⁵⁰ M. Mahamunulu Desu & R.H. Rodine, *Estimation of the Population Median*, SCANDINAVIAN ACTUARIAL J. 67–70 (1969).

⁵¹ George Brown, *On Small-Sample Estimation*, 18 THE ANNALS OF MATHEMATICAL STAT. 583 (1947).

⁵² See generally MORRIS H. DEGROOT & MARK J. SCHERVISH, PROBABILITY AND STATISTICS 487, 491–93 (4th ed. 2010) (providing a technical discussion on the meaning and interpretation of confidence intervals).

by: $1 - \sum_{j=0}^{k-1} \binom{N}{j} \left(\frac{1}{2}\right)^{N-1}$.⁵³ Using this formula, whose meaning and derivation

are explained in Appendix A, the following table lists different sample sizes and corresponding levels of confidence and precision that they yield for an estimate of the true median. Thus, for example, if $N = 199$, and $k = 89$, a company could take a sample of 199 employees, rank them in ascending order of compensation, and thereby achieve 90% confidence that its true median compensation will be between the compensation of the 89th and the 111th individuals in the sample.

Table 1—Exact k-Values for Confidence Intervals		
Sample Size	90% Confidence Interval	95% Confidence Interval
	k - value	k - value
N = 99	k = 42	k = 40
N = 199	k = 89	k = 86
N = 399	k = 184	k = 180
N = 499	k = 232	k = 228
N = 999	k = 474	k = 469

These figures, however, only indicate that the median is likely to be between certain ranked values of the sample. They do not indicate how wide the range is likely to be in absolute dollar terms. Of course, any company that applies this methodology would immediately be able to determine the level of compensation of the k th lowest and k th highest paid employees in their sample, and so the dollar value that describes the width of the interval would be easy to report. Nevertheless, in determining the appropriate values to set for N and k , the SEC will likely want to prospectively consider how wide of intervals, in dollar terms, different values of N and k will produce. Clearly, this will be a matter for

⁵³ See generally JOHN A. RICE, MATHEMATICAL STATISTICS AND DATA ANALYSIS 395–97 (3d ed. 2007) (providing a general procedure for calculating confidence intervals for medians, upon which the methodology in this article is based).

cost/benefit consideration that the SEC will be best poised to conduct with input from companies and the public.

The ideal would be for private companies or trade associations with access to employee compensation databases to work with the SEC as it implements this rule by using their databases to test the results from different values of N and k . Barring this, however, it is possible to generate simulated databases of employee compensation levels in order to estimate how wide in dollar terms the confidence levels produced by this procedure would be, were it applied to actual companies.

Using reasonable assumptions about the distribution of a company's compensation, as detailed in Appendix B, it can be shown that the confidence bands around the sample median would be of the size given in the table below:

Sample Size	Margin of Error 90% Confidence Level	Margin of Error 95% Confidence Level
N = 99	+/- \$1905	+/- \$2444
N = 199	+/- \$1286	+/- \$1668
N = 399	+/- \$938	+/- \$1178
N = 499	+/- \$842	+/- \$1035
N = 999	+/- \$607	+/- \$725

Thus, as can be seen, a relatively small sample can generate a narrow band that predicts the population median with a high degree of confidence.

The advantage of this procedure is that it is robust to different possible distributions of an employer's compensation. Although the dollar value that designates the size of the confidence intervals may vary from company to company, all companies will have the same level of confidence for the interval between the k th smallest and k th largest entry in a sample of a given size N . Nevertheless, it is certainly plausible that a particular company, with more detailed knowledge of its unique distribution of employee compensation, may be able to devise another sampling procedure that achieves predictions of comparable precision and accuracy to those specified here. Thus, the SEC could also specify in its rulemaking that if any company devises another procedure that it can demonstrate generates a median-unbiased estimator of the median within an $X\%$ confidence interval

of +/- \$Y of the estimated median, with the SEC selecting values for X and Y to best fit its cost/benefit analysis, then the company can substitute that estimator and procedure for this one specified here.

CONCLUSION

Section 953(b) does not specify how issuers must calculate the median of the annual total compensation of all employees. Because the median is a statistical term used to describe a set of observations, it is reasonable for the SEC to permit issuers to sample their employee populations to calculate the median. This approach will provide accurate information to investors with reduced compliance costs for issuers.

APPENDIX A: DERIVATION OF FORMULA FOR THE LEVEL OF CONFIDENCE
BETWEEN THE KTH SMALLEST AND KTH LARGEST VALUES IN A SAMPLE OF
SIZE N

The median is the value in a distribution that is greater than or equal to exactly half of the other values and less than or equal to exactly half of the other values. This special feature of the median lends itself very well to establishing a precise confidence interval for a company's true median compensation, based on taking just a small sample of employees and measuring their total compensation.

To start out with, suppose a company randomly and independently selects a sample of fifty employees and wants to know what the probability is that the true median compensation level is somewhere between the compensation levels for the lowest paid employee and the highest paid employee. For every individual in the sample, there is a 50% chance that that individual's compensation is below the true median and a 50% chance that that individual's compensation is above the true median.⁵⁴ Therefore the chance that *all fifty employees* in the sample have compensation above the true median (in other words, the probability that the true median is less

than the smallest value in the sample) is given by $\left(\frac{1}{2}\right)^{50}$. Likewise, the chance that *all fifty employees* in the sample have compensation below the true median (in other words, the probability that the true median is greater

than the largest value in the sample) is also given by $\left(\frac{1}{2}\right)^{50}$. Furthermore these two events are mutually exclusive—it is impossible for all fifty employees in the sample to have compensation that is both above *and* below the actual company median. Therefore, the probability that either of these conditions holds is simply the sum of the probabilities that each of them individually holds, which is given by:

⁵⁴ Technical note: When sampling from an unknown continuous distribution, such as that of a company's employees, the value of any given observation is considered to be a continuous random variable. In statistical theory, the probability that a continuous random variable will assume any single value is considered for technical reasons to be zero. Thus, this section will refer to the probability that an

$$\left(\frac{1}{2}\right)^{50} + \left(\frac{1}{2}\right)^{50} = 2 \cdot \left(\frac{1}{2}\right)^{50} = \left(\frac{1}{2}\right)^{49}.$$

In other words, if a company samples 50 employees and determines their individual compensations, then the probability that the company's true median compensation will not be somewhere between the lowest and highest paid individuals is given by $\left(\frac{1}{2}\right)^{49} = .000000000000002$. Therefore, the company can achieve 99.9999999999998% confidence that the true median compensation will be somewhere between the lowest paid individual in the sample and the highest paid individual. By using the lowest and the highest compensations in the sample, the company will be able to achieve a 99.9999999999998% confidence interval for its true median compensation.

Clearly then, by forming an estimate based on the smallest and largest compensations in a sample, a company can achieve a very high confidence level. In practice, however, there is likely to be quite a wide range between the lowest and the highest compensated individuals, even in a 50-person sample. Thus, achieving 99.9999999999998% confidence that the true median is somewhere within such a wide range will not be tremendously informative. This procedure can be easily extended, however, to give confidence percentages for smaller intervals within a sample.

To start with, suppose that a company wants to know the probability that the true median will be somewhere between the 2nd and the 49th observations in its sample, when those observations are ranked from smallest to largest. As with before, it is useful to break this down into pieces. First off, consider the probability that the company calculates compensations for a sample of fifty employees and that *forty-nine out of the fifty employees* in the sample have compensation above the true median (in other words, the probability that the true median is below the compensation of at least forty-nine out of the fifty employees). This equals the probability that the company would draw 50 observations, all 50 of which are greater than the population median (since if all 50 are greater than the true median, then it is also true that 49 are greater than the true median), plus the

individual employee has compensation that is *less than* the true median, rather than the probability that an employee has compensation that is *less than or equal to* the median. *See id.* at 396.

probability that the company would draw 50 observations, with exactly 49 of them greater than the true median. The first of these two probabilities was already calculated above as $\left(\frac{1}{2}\right)^{50}$.

For the second probability, there are 50 different combinations that could produce this because any of the fifty employees in the sample could be the lone employee whose compensation is less than the true median. For any given employee in the sample, the chance that their compensation is less than the true median is 50%, and the chance that all other 49 employees have compensations greater than the true median is $\left(\frac{1}{2}\right)^{49}$. Thus, the probability that that one particular employee's compensation is below the true median, *and* that all other employees have compensations above the true median is given by: $\left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right)^{49} = \left(\frac{1}{2}\right)^{50}$. But, since there are fifty different employees for whom this could be true, the probability that *any one* employee has compensation below the true median, while all others have compensation above the true median, is given by: $50 \cdot \left(\frac{1}{2}\right)^{50}$.

In situations such as this, where all of the different possible combinations that could satisfy a condition are added up, mathematicians use the binomial coefficient to represent the number of possibilities. In the case of calculating how many different combinations of one employee can be drawn out of a sample of 50 employees (i.e. how many different ways there could be exactly one employee whose compensation is below the sample median), the binomial coefficient is written as: $\binom{50}{1}$, which is read as "fifty choose one." Technically speaking, the binomial coefficient is defined as $\binom{n}{j} = \frac{n!}{j!(n-j)!}$; for $n \geq j \geq 0$, where the factorial operator "!" signifies $n! = n(n-1)(n-2)\dots(1)$ and where by convention $0! = 1$. Many standard calculators and computer programs also have built-in functions for the binomial coefficient, and in this case, it is easy to calculate that $\binom{50}{1} = 50$. Thus, the probability that exactly one employee in a sample will

have compensation below the true median is given by the number of possible combinations that could produce this (fifty), times the probability

of any one of the possibilities $\binom{50}{1} \left(\frac{1}{2}\right)^{50}$. In other words, the probability that exactly one employee in a sample will have compensation below the true median is given by: $\binom{50}{1} \cdot \left(\frac{1}{2}\right)^{50} = 50 \cdot \left(\frac{1}{2}\right)^{50}$, exactly as before but now

written using the binomial coefficient.

In fact, the binomial coefficient is implicitly present in the calculation of the probability that exactly fifty employees have compensation above the true median (in other words, the probability that exactly zero employees have compensation below the true median). In this case, the binomial coefficient is represented as $\binom{50}{0}$ (note that there is only one way to

choose exactly zero items out of a sample of fifty). Thus, the probability that all fifty employees will have compensation below the true median is

given by $\binom{50}{0} \cdot \left(\frac{1}{2}\right)^{50} = 1 \cdot \left(\frac{1}{2}\right)^{50}$, again, exactly as before.

Putting all of these calculations together then, the probability that the true median will be less than the 2nd ranked sample compensation (i.e. the

second lowest) is given by $\binom{50}{0} \left(\frac{1}{2}\right)^{50} + \binom{50}{1} \left(\frac{1}{2}\right)^{50} = \sum_{j=0}^{k-1} \binom{50}{j} \left(\frac{1}{2}\right)^{50}$,

where here, $k = 2$ because the procedure is calculating the probability that the true median is less than the 2nd ranked sample.

As with above, the problem is symmetrical and therefore the probability that the true median will be greater than the 49th ranked employee in the sample (i.e. the second highest) is the same as the probability that the true median will be less than the 2nd ranked employee in the sample. Thus, the probability that the true median will be *between* the 2nd and the 49th ranked employees in the sample is given by

$1 - 2 \cdot \sum_{j=0}^{k-1} \binom{50}{j} \left(\frac{1}{2}\right)^{50} = 1 - \sum_{j=0}^{k-1} \binom{50}{j} \left(\frac{1}{2}\right)^{49}$, which equals 0.99999999999991.

Thus, a company would be able to report a 99.999999999991% confidence interval for their true median compensation by giving the interval bounded

by the second lowest and the second highest compensations in a sample they took of 50 employees.

Finally then, this formula can be completely generalized. For a sample size of N , the confidence level that the true median is between the k th

smallest and k th largest value in the sample is given by: $1 - \sum_{j=0}^{k-1} \binom{N}{j} \left(\frac{1}{2}\right)^{N-1}$.

APPENDIX B: DERIVATION OF SIZES OF CONFIDENCE BANDS FOR
SIMULATED COMPANIES

In order to generate simulated employee compensation databases, it is necessary to make certain assumptions about the underlying distribution of employee compensation within companies. It is widely known that income within the population as a whole tends to be distributed in a log-normal form.⁵⁵ In such a distribution, there are a large number of employees who make low to medium salaries and a small number of employees who make very large salaries. Given the distribution of income within the population as a whole, it is a reasonable assumption that intra-company compensation distributions are also approximately log-normal.⁵⁶

In formal mathematical terms, a log-normal distribution is created by taking a constant value and raising it to the power of each value found within a normal distribution. Thus, if Y is a normally distributed variable, with a mean of μ and a variance of σ^2 , then $Z = e^Y$ will be a log-normally distributed variable, that is, $\ln(Z) \sim N(\mu, \sigma^2)$.⁵⁷ Thus, the only parameters whose values must be assumed in order to generate a simulated employee database with log-normally distributed compensation, are μ and σ^2 .

The mean or expected value of a log-normally distributed population is given by $E(z) = e^{\mu + \sigma^2/2}$.⁵⁸ The median of a normal distribution is equal to its mean, and because the exponential function does not change the ordering of numbers it is applied to, the median of a log-normal distribution is simply e^μ .

Given these properties, an approximately 100,000 person company with a log-normal wage distribution that has a mean compensation of around \$33,000 per year, a median compensation of around \$5000 per year, and a CEO paid around \$15,000,000 per year, would have approximate

⁵⁵ See, e.g., Ronald Oaxaca & Michael Ransom, *Using Econometric Models for Intrafirm Equity Salary Adjustments*, J. ECON. INEQUAL. 230 (Dec. 2003).

⁵⁶ *Id.*

⁵⁷ R. CARTER HILL, WILLIAM E. GRIFFITHS & GUAY C. LIM, *PRINCIPLES OF ECONOMETRICS* 103 (3d ed. 2008).

⁵⁸ *Id.* at 104.

values $\mu=8.5$ and $\sigma^2=3.8$. In order to generate the dollar estimates for the margins of error provided in Table 2, these values were inserted into a Monte Carlo simulation that created 10,000 simulated companies with log-normally distributed compensation according to values of $\mu = 8.5$ and $\sigma^2=3.8$. Each of these 10,000 simulated companies was then randomly sampled, using sample sizes as indicated in the table. Finally, the difference between the k th smallest and the k th largest values in each of the samples was taken, and the median of the 10,000 differences from the simulated companies was calculated. The margins of error in the table are each one-half the size of these median figures.