July 5, 2015
Brent J. Fields
Secretary
U.S. Securities and Exchange Commission
100 F Street NE Washington, DC 20549-1090

Re: Statistical Sampling Procedures for Section 953(b) of the
Dodd-Frank Wall Street Reform and Consumer Protection Act

Dear Mr. Fields,

I am writing as a JD candidate at Stanford Law School and as a Ph.D. Candidate in Stanford University's Department of Management Science and Engineering to offer to the Commission a few brief follow-up comments and suggestions, building off of my Comment Letter to the Commission on September 25, 2012, in which I suggested a statistical sampling methodology the Commission could use to implement Section 953(b) of the Dodd-Frank Wall Street Reform and Consumer Protection Act. These are based on my continued study of both legal and mathematical topics pertaining to the implementation of statistical sampling to estimate median compensation in corporations as stipulated in Section 953(b) of Dodd-Frank.

Comment 1:

First, I want to commend the Commission on their proposing release from last year on this regulatory topic. As you are well aware, this is a very challenging regulation to implement in a sensible, efficient, and effective way, and I think that the Commission did well rising to the many challenges it presented.

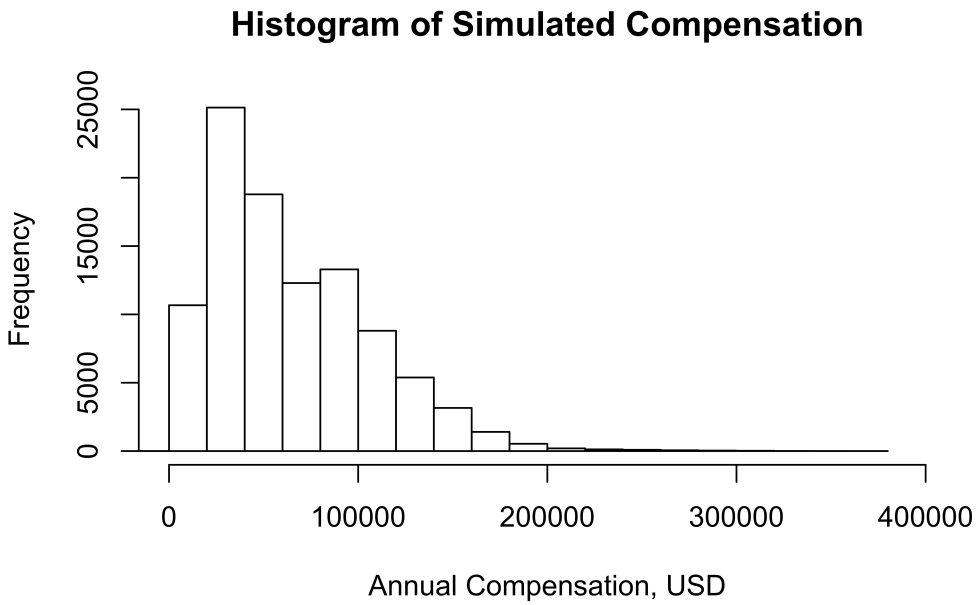
Comment 2:

Secondly, I would like to update the Commission on some recent findings that I have had from my further work on the theory and practice of statistical sampling as applied to median estimation. In my original comment letter to the SEC, I described a simple random sampling procedure that could be used to estimate median employee compensation and offered some calculations of sizes of confidence intervals that could be achieved using different sample sizes and under certain assumptions about the distribution of a company's worker pay. Although the procedure I presented is still an effective and valid means for companies to estimate median employee compensation (see more on this below), there are a host of other more sophisticated techniques that can achieve far greater precision than simple random sampling and with only a fraction of the sample size needed.

For instance, in studies that I have done, based on company structures with stratified employee groups (that follow distributions much more complicated than a simple log-normal distribution), I have found techniques that can achieve suitable levels of precision using as few as 100 to 400 individuals in the simple random sample. Specifically, for example, in one simulation study, I supposed that a company with 100,000 total employees was composed of five different segments, corresponding to either different divisions, different geographic areas, or other similar segmentations. I assumed that the five divisions accounted for 20%, 30%, 30%, 19%, and 1% of the total workforce. Thus, for example, the division with 1% might represent upper management, the lower divisions might represent lower paid classes of employees,

employees in other countries, etc. I assumed that each of the divisions had randomly distributed base compensations with corresponding means of $20,000, $40,000, $80,000, $120,000, and $200,000 per year in USD and standard deviations of $5,000, $10,000, $20,000, $30,000, and $60,000.

I further assumed that employees in the divisions received additional compensation that could account for bonuses, employee benefits, government-provided benefits of the type that have been discussed by the SEC and in comment letters as being permissible to include in compensation figures for foreign employees, or other such uncertain and variable compensation. I assumed that these random elements of the employees' "bonus" compensation were normally distributed within each of the five groups, with corresponding means of $4000, $8,000, $16,000, $24,000, and $40,000 and with standard deviations of $2,500, $5,000, $10,000, $15,000, and $30,000. Overall, this resulted in employees with salaries as low as a few hundred dollars per year to as high as hundreds of thousands of dollars per year. I did not include any estimate of the CEO's compensation within these simulations, since the primary concern in implementing this regulation is in estimating the median compensation for all employees apart from the CEO.[1] The overall histogram of this simulated company with stratified groups of employee compensation is given below.



[1] Many companies might also have a very small number of non-CEO employees making somewhere between $400,000 and perhaps $2,000,000 per year. Yet, for the purposes of median estimation, these "extreme values" have essentially no impact on the overall median compensation – indeed, it is the fact that the median is not influenced by extreme values that makes it a popular statistic to use to describe distributions such as employee compensation. Thus, whether an employee makes $400,000 or $4,000,000 per year has little impact on either the actual median or the median estimation, and thus the fact that non-CEO employees in this simulated company have a max compensation of around $400,000 does not represent a significant simplification.

Overall then, the total compensation for this company still follows a shape that is roughly similar to that of a log-normal distribution (with a large portion of the total employees towards the bottom of the distribution and a long "tail" showing a smaller number of employees with compensations much greater than those in the bulk of the company). Nevertheless, the actual distribution is much more complicated than that of a simple log-normal distribution, as evidenced by the detailed description of my construction of this simulated company's compensation given above.

Starting from this more complicated distribution of company pay, I applied a set of more sophisticated and specialized sampling techniques to estimate median compensation. The mathematical details of these specialized techniques are somewhat more complicated than those that I spelled out in my original letter to the SEC. Thus, I have omitted them from this letter for the purposes of brevity and also because a hallmark of such techniques is that the details of their implementation generally will vary somewhat depending on a company's particular circumstances. Nevertheless, I would be happy to provide the SEC with additional information upon request if it would be useful to your efforts.

Below, I have presented a table which compares different sample sizes and the standard deviations of the median estimators when using both these more specialized sampling techniques as well as simple random sampling. In all cases, the median estimators are approximately normally distributed and unbiased. This means that, for instance, a median estimator with a standard deviation of $1,000 will be within plus or minus two standard deviations, or in other words $2,000, of the true median compensation of a company approximately 95% of the time.

	Standard Deviation (for median estimator)	
Sample Size	Specialized Sampling Techniques	Simple Random Sampling
100	2242	6980
200	1548	4821
300	1261	3905
400	1059	3371
500	958	2999
600	893	2728
700	802	2513
800	749	2340
900	702	2210
1000	665	2087
1100	640	1997
1200	607	1907
1300	573	1826
1400	560	1767
1500	542	1708
1600	532	1649
1700	518	1600
1800	497	1557
1900	480	1517
2000	466	1483

There are several salient points to note about this table. First, with sample sizes as low as 100 to 400 employees, these specialized techniques can estimate median compensation with a standard deviation between $2,242 and $1059, respectively. This means that a sample size of just 100 employees will yield a median estimate within about $4,484 of the true median 95% of the time, and a sample size of just 400 employees will yield a median estimate within about $2,118 of the true median 95% of the time. The second thing to note is how much improvement in terms of efficiency is available through these specialized techniques, as compared to simple random sampling. Nevertheless, even simple random sampling can perform well, given a sufficiently large sample size.[2]

This example that I have detailed above is of course but one of many. I have conducted many other simulation studies, using a variety of constructions of company compensation and have obtained similar results to those reported here in each one.

These more sophisticated techniques, whose results I present above, will generally require the skill and judgment of someone with experience in random sampling designs to fit to a company's specific situation. If a company does not have such expertise in-house, they would need to hire or consult with someone with such expertise to assist them. Nevertheless, for companies for whom the sampling itself is relatively expensive, this can be a valuable tradeoff. For companies not wishing to deal with such sophisticated techniques, the option of simple random sampling remains a viable alternative.

Comment 3:

As a final comment, there is one small but important error in the statements made in the Commission's proposing release from October 1, 2013 on the implementation of section 953(b) of Dodd-Frank. On page 60,594 of the proposing release (as reported in the Federal Register), the Commission asserts that there would be some companies with relatively complex business and structures and that for such companies, "all [statistical sampling techniques] would be more complicated than simple random sampling." This statement is unambiguously false.

The Commission provides no direct arguments in favor of this statement, but does cite an article, S. Gross. Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods. American Statistical Association*, 181–184. (1980). Yet, this article that the Commission itself cites explicitly states that simple random sampling (i.e. taking a simple random sample and finding the median of that sample) is a valid technique for estimating medians in any situation, even if it is not always the most efficient technique. Specifically, the article states on p.181 that "[s]ample medians have long been recognized as simple robust alternatives to sample means, for estimating location of heavy-tailed or markedly skewed populations from simple random samples." The article goes on to state that "[a]lthough the

---

[2] The standard deviations for the median estimator using simple random sampling given here are somewhat larger than those for the simple random sampling estimator in my original letter to the SEC on September 25, 2012. This is to be expected – the simulated company that I created here has a more dispersed compensation structure than the simulated company that I used in my original letter. As I discussed in that original letter, the precision of a statistical estimator will be influenced significantly by how dispersed the underlying population being sample is. Thus, the original standard deviations I presented for the estimator were indeed correct for the simulated population I assumed in that letter and the ones I present here are correct for this simulated population. The fact that a given estimator yields different standard deviations when applied to different populations does not invalidate the estimator – in fact, quite the opposite – every legitimate estimator will yield different standard deviations when applied to different populations.

sample median did not emerge as 'best' estimate in many nonstandard populations simulated in the study, its robustness in small samples for medium and large deviations from normality was clearly demonstrated."

It was quite odd, therefore, to see the Commission citing this article in support of its assertion that for some companies, all statistical sampling techniques that could estimate median employee compensation would be "more complicated than simple random sampling." My original comment letter to the SEC from September 25, 2012 also addresses this topic in some depth, providing both a mathematical proof and additional citations[3] supporting the validity of simple random sampling as a robust technique to estimate median compensation in any population.

It is important that the Commission correct this issue in its comments in its final release, lest issuers believe that simple random sampling would not be a valid technique for them to employ and that instead they must, in order to comply with the regulation, employ more complicated techniques This is an important issue for minimizing compliance costs for this provision. For many companies, it may well be optimal to invest the resources in developing a more sophisticated sampling technique that can account for the company's unique and complex distribution of compensations, and that can minimize the sample size and maximize the efficiency of the sampling technique they implement. Nevertheless, for other companies, particularly those who may have a relatively low cost of sampling or who do not wish to engage an experienced statistician to assist them with implementing a more complex sampling procedure, simple random sampling remains a technique that can be easily and simply implemented and that can produce acceptable precision. The Commission should leave companies the discretion to choose whether it is optimal to spend more money on developing a more complex sampling procedure that requires a smaller sample size or to spend less money developing a simpler procedure that may require a larger sample size.

Conclusion:

Once again, I commend the commission for its excellent work on this challenging regulatory situation. If there is anything further that I can do to be of assistance to the commission, please do not hesitate to contact me, either by phone at [redacted] or by email at [redacted]

Yours,
Michael Ohlrogge

[3] M. Mahamunulu Desu & R. H. Rodine, *Estimation of the Population Median*, SCANDINAVIAN ACTUARIAL JOURNAL 67-70 (1969:1-2); John A. Rice, MATHEMATICAL STATISTICS AND DATA ANALYSIS 395-97 (3rd ed. 2007).