Portfolio Media. Inc. | 860 Broadway, 6th Floor | New York, NY 10003 | www.law360.com
Phone: +1 646 783 7100 | Fax: +1 646 783 7161 | customerservice@law360.com

The Use Of Statistical Sampling Post-Duran

Law360, New York (August 12, 2014, 11:35 AM ET) --

Sarah Butler

In its recent decision in Duran vs. U.S. Bank (Samuel Duran, et.al. v. U.S. Bank National Association. Super. Ct. No. 2001-35537), the California Supreme Court affirmed the Court of Appeal's ruling and remanded a new trial on liability and damages in this misclassification suit. While certainly a victory for the defendant, the ruling is also a victory for the rigorous application of scientific methods for statistical sampling.

It is useful to review the criticisms of the sample in the Duran matter and evaluate what steps the Supreme Court thought should have been taken. This decision provides insight not only as to the specific sampling issues that could occur in a misclassification case, but also the ways in which scientifically reliable sampling may be useful in labor and employment cases.

The Duran Sample

The Duran case was a misclassification case involving the employees of U.S. Bank National Association ("USB") working as business banking officers, or BBOs.[1]

In December 2001, a class action was filed alleging that USB's BBOs were misclassified as exempt and should not have been denied overtime pay. The original named plaintiff in the case was replaced by three alternate class representatives. In March 2005, these three were also replaced and two plaintiffs (Duran and Fitzsimmons) remained as named plaintiffs throughout trial. The four named plaintiffs who were replaced all indicated that they in fact met the exemption requirements and worked out of the office at least 50 percent of their workday.

A class of 260 BBOs was certified in 2005 and both sides proposed a trial management approach for the liability issues. Plaintiffs suggested a sample and survey of class members. Defendants objected and argued instead for special master hearings of individualized testimony on liability and damages.

The court agreed with the plaintiff and allowed for a sample. But, rather than rely on plaintiffs' expert to design and implement the sample, the court decided to randomly select 20 class members to extrapolate liability and damages to the entire population of BBOs. The court also decided to allow the two named plaintiffs to be part of the estimate, bringing the total effective sample size to 22.

After a procedural change by plaintiffs caused a second request for opt-ins to be sent out to the class, four of the sampled employees opted out and one employee was excluded.[2] These five eliminated employees were replaced with alternates. From the testimony of the 19 sampled employees who appeared at trial[3] and the two named plaintiffs, the trial court determined that all BBOs were misclassified. Plaintiffs' expert used information from the 21 employees and estimated an average of 11.86 weekly overtime hours with a margin of error of plus or minus 5.14 hours.

Sampling Errors as Cited by Supreme Court

In its review of the evidence from the sample used in the lower court, the California Supreme Court indicated, "[T]he sample relied upon must be representative and the results obtained must be sufficiently reliable to satisfy concerns of fundamental fairness." The court went on to describe a number of methodological flaws that made the evidence from the 21 employees unreliable.

Small Sample Size

The California Supreme Court indicated that the "arbitrarily" selected sample size of 20 was too small. Correctly, the state high court focused not on the overall number but rather the size relative to the possible variability in the population of BBOs. The court pointed out that, "It is impossible to determine an appropriate sample size without first learning about the variability in the population." Determining the correct sample size is in part a function of how much underlying variability or differences exist in the total population; more variability in a population requires a larger sample to ensure a reasonable margin of error.

There are a number of possible ways potential variability can be assessed in a labor case: existing data, time sheets, other personnel records, or, as the Supreme Court indicated, surveys. Not surprisingly, USB objected strongly to the use of a survey to determine liability and there are a number of difficulties associated with using surveys in labor cases.[4] But in this case, a survey may have been helpful to demonstrate that significant variability existed and a larger sample should have been taken. In fact, a survey could have potentially demonstrated that the amount of time spent out of office and the amount of overtime worked varied not only between employees but also within a single employee's work history.[5] Most employees have some variation in their schedules and for some — seasonal workers or new hires — the variation may be substantial. Depending on the type of estimates being calculated, individual level variation coupled with variation across different employees can create large margins of error and may demonstrate that the sample needs to be very large before reliable estimates can be made.

Key Lesson

Determining the appropriate sample size is not simply a matter of convenience, as the California Supreme Court pointed out, it needs to be based on scientific principles and some understanding of the underlying population's distribution. Surveys or other data points may demonstrate that variability across employees render small or simple random samples unusable. If the variation in the population is significant a large sample size may be required to generate reliable results.

Sample Was Not Random

The California Supreme Court also correctly indicated that plaintiffs' results were made unreliable due to

nonresponse bias and selection bias.

Individuals who are selected for the sample may not be included or may not participate for many reasons (e.g., contact information for the selected individual may be incorrect or out of date or the respondent may be unavailable during the data collection period). These noncontacts can increase the nonresponse rate, but may not affect the reliability of the data.

Nonresponse becomes a problem when nonrespondents differ in systematic and meaningful ways from the study participants and the difference introduces bias into the results. To evaluate for potential nonresponse bias, an expert can compare the characteristics of those who responded with those who did not. For example, did nonrespondents come from particular stores or geographic locations? Did nonrespondents all work during particular time periods or have particular job descriptions? Showing that nonrespondents differ in systematic and meaningful ways from those included in the sample can be an indication that nonresponse bias is affecting the data and may be undermining the reliability of the results.

The Duran sample was also affected by selection bias. Because sampled members were allowed to opt out — four of the originally sampled opted out — and the two named plaintiffs were included in the total, the final pool of respondents cannot be said to be representative of the total population of class members. Selection bias occurs when individuals or data points are selectively included or excluded from the sample.

In the Duran matter, 20 percent of the original sample opted out of providing evidence at trial and 10 percent of the data — the two named class members — came from employees who opted in, as named plaintiffs.[6] The sample opt-out rate was substantially different than the 2 percent opt-out rate for the remainder of the class (i.e., five out of 235). Additionally, the court removed from the originally selected sample a worker whose work habits deemed too different from other employees.[7]

Key Lesson

A sample with nonrespondents who are systematically and materially different from respondents may generate biased results. An expert can evaluate the characteristics of the two groups to determine if meaningful differences exist. Additionally, the sample cannot be based on selective criteria that includes or excludes members of the population if the sample is going to be used to make population estimates.

Large Margin of Error

Even if the sample had been selected according to scientific standards, the California Supreme Court indicated that the margin of error on the average overtime hours worked from the 21 employees was unacceptably large. Using the data from the 19 sampled employees who appeared for court and the two named plaintiffs, plaintiffs' expert calculated an average of 11.86 hours of overtime per week with a margin of error of 5.14 hours. With a 95 percent confidence interval, this means the estimated overtime hours worked would have been between 6.72 and 17 hours. Simple math would suggest that such a large margin of error could have very different cost implications for a class of 260 workers.[8]

The sample size of 20 was selected without knowing anything about the variability in overtime hours and the end result was an unacceptably large range for potential damage estimates. A pilot sample could have helped to establish what sample size would be needed to get a particular margin of error. In some circumstances, a court may want to select a neutral expert to help determine what sample size

would yield an acceptable margin of error. While the acceptable margin of error will largely depend on issues specific to the case, a relative error rate of 43 percent (i.e., the ratio of the error to the estimate or 5.14 to 11.86 overtime hours) is large given the potential impact in possible dollars.

Key Lesson

Small sample sizes, even when selected correctly, may lead to large margins of error. A pilot sample may help to determine how large a sample is needed and a statistical expert can help decide what sample size is necessary.

Lessons in Sampling

The Duran decision provides a number of useful pointers in how best to develop and implement a sample in employment litigation. While the California Supreme Court refuses to weigh in on whether such techniques can be used to establish liability, the court clearly indicates that, if sampling is used, it must adhere to scientific standards. Courts must pay careful attention to the potential sources of variability and of bias that can render results from a sample unreliable.

—By Sarah Butler, National Economic Research Associates Inc.

Sarah Butler is vice president of NERA Economic Consulting and is located in the firm's San Francisco office.

[1] The job of a BBO is to create new business and their success was evaluated by the number of new accounts, not time spent in a branch location. According to USB, BBOs were required to visit new and potential customers and were considered by USB to be exempt from overtime compensation, according to the outside sales exemption in Labor Code Section 1171, which indicates that to be exempt an employee must spend 50 percent or more of her working hours outside of the office.

[2] This employee's job description differed so greatly from other sampled employees that the court removed him from the sample.

[3] One sampled employee did not respond to trial subpoena and was treated by the lower court as though he was nonexempt.

[4] Surveys in labor typically violate the "double-blind" rule (i.e., survey respondents in labor cases tend to be aware of the intention of the research and may participate or not participate in survey depending on their interests in the case). Similarly, respondents with a vested interest in the case outcome may undereport or overreport data that may be difficult to verify.

[5] In Flores v. Lamps Plus, a survey submitted by plaintiffs was actually used by the court as evidence to decertify a class. The survey demonstrated variability in relation to employee meal and rest breaks and showed no consistent pattern of denial.

[6] The court also refused to consider the testimony of four original named plaintiffs who were later

replaced because these employees did not appear to represent plaintiffs' claims.

[7] While the lower court asserted this worker was "too different," it actually had no empirical way of knowing if this worker actually represented work habits of a sizable portion of the population. A larger sample would have allowed the court to determine whether this employee was truly unusual.

[8] If we assume California's minimum wage ($8) and time-and-a-half for overtime ($12) and that each worker worked 25 weeks with owed overtime, the range of damages owed to class could be between $524,000 and $1.3 million.
