**Portfolio Media. Inc.** | 648 Broadway, Suite 200 | New York, NY 10012 | www.law360.com
Phone: +1 212 537 6331 | Fax: +1 212 537 6371 | customerservice@portfoliomedia.com

# Challenges In RIF Analyses

Law360, New York (March 25, 2009) -- In "Reductions-In-Force: Best Practices" (Employment Law 360, Jan. 27, 2009), authors Jonathan C. Wilson and John M. Farrell of Haynes and Boone LLP state:

"It is ... imperative to perform a statistical impact study once preliminary decisions are made. This study should be part of the oversight review, before final implementation. ... Reductions-in-force [RIF] require planning and careful consideration in order to avoid costly legal battles that destroy the cost of the anticipated savings."

A proactive pre-RIF adverse impact analysis as proposed by Messrs. Wilson and Farrell is indeed a powerful instrument in the management of post-RIF litigation risk.

Companies thinking of reorganizing through RIFs should perform certain statistical tests to determine whether any protected group is over-represented in the pool selected for the RIF and, if so, whether it is statistically significant enough to support a claim of disparate impact.

These statistical significance tests, which can be used for both pre-RIF proactive analysis and post-RIF litigation, are complex; using the wrong test may indicate disparate impact when none exists.

In this article, we describe these tests, discuss their relative merits and illustrate the importance of using the appropriate one. Our examples consider gender as the protected class, but the argument applies equally to other protected classes, such as minorities, workers age 40 or older, etc.

## Approaches to Analyzing Disparate Impact

To test for disparate impact, the hypothetical firm, Company XYZ, should first divide the employees into one of the four categories below. This assignment of data is called a 2x2 (retained-selected, male-female) contingency table.

- selected women - retained women - selected men - retained men

Next, the company should calculate the difference between the actual number of women selected for the RIF to the number of women one would expect to be selected based on the proportion of women in the company's workforce.

If the actual number is greater than the expected number, the company should then perform a statistical significance test called the Chi-Square Test (CST).

A significance level of 0.2 then would mean that there is only a 20 percent chance that the observed over-representation of women in the RIF selection process occurred by random chance and that the test is 80 percent confident that chance is not the reason for the over-representation.

The courts have found a significance level of 0.05 or less (five percent or less) for the data to be consistent with a claim of disparate impact on women. For exposition purposes only, the discussion below will use the 0.05 significance level as the threshold to establish discrimination.

The CST is meaningful only when the sample is large and not sparsely or unevenly distributed across the four categories. Otherwise, it may give wrong or misleading results, indicating, for example, that the RIF had a disparate impact on women when in reality it did not.

Being aware of this shortcoming can make the difference between an employer prevailing at trial or paying millions in damages.

If the sample size is small or the data is sparsely or unevenly distributed, a more appropriate test is the Fisher's Exact Test (FET), a valuable tool for labor economists proposed by R.A. Fisher during the Great Depression and used primarily in biomedical research.

The advantage of the FET over the CST is that unlike the CST, the FET does not make any parametric assumptions about the underlying distribution.

Parametric assumptions consider the data to have a certain stylized pattern, for example, coming from a normal ("bell-shaped") distribution.

The FET does not make any such assumptions but instead relies on the empirical distribution (i.e., the observed distribution of the data).

Computing the FET is an intensive process, but it is now routinely reported in the output of many standard statistical software packages.

## Applying the CST and FET Tests

Going back to the above example, suppose that Company XYZ has 50 male and 50 female employees and wants to reduce its workforce by 20.

A completely gender-neutral selection process would select 10 male and 10 female employees for the RIF. Let's say that the company selects 14 female and 6 male employees for reduction, and retains 44 male and 36 female employees.

The CST probability of this distribution is 0.046 which, at less than 0.05, would support a claim that the RIF selection process has an adverse impact on women.

But this would be the wrong conclusion. The probability from the more accurate FET is 0.078, a value greater than 0.05, and one that does not support an adverse impact claim.

## Using Regression Analysis

Even confronted with evidence of statistically significant over-representation of women in the RIF selection, Company XYZ may justify the selection criteria used if they are consistent with business necessity.

In this event, the additional criteria would force the company to replace the simple 2x2 contingency table framework with a multiple regression analysis where additional factors can be considered to model the selection process. Multiple regression analysis can control for all characteristics — for example, the employee's education level, performance ratings, job function and division assignments — that were considered during the RIF selection process.

Let's assume that Company XYZ has three divisions and that the supervisors of each used different selection criteria. To aggregate data from the three divisions, consider the whole workforce as a single entity and use a single FET would be inappropriate.

Instead, in this context, a Standard Logistic Regression (SLR), a special type of multiple regression analysis, is traditionally used.

The SLR multiple regression method is used when the outcome variable is binary (retained-selected, in our example) and multiple factors — gender and the division of an employee — are considered as determinants of the outcome.

Other additional criteria considered during the RIF selection process could be added to the model to explain the selection process.

Like the CST, the SLR makes parametric assumptions about the underlying distribution and is meaningful only when the sample is large and not sparsely or unevenly distributed across different categories.

Because Company XYZ has three divisions, the number of categories increases from four to 12 (selected women in division one, retained women in division one, selected men in division one, etc.) and the data must now be presented as a 2x2x3 (retained-selected, male-female, divisions 1-3) contingency table.

As more criteria are added to the model, the number of categories increases geometrically. When the size of a layoff increases, the number of additional criteria considered in the selection process normally increases, making it likely that the data will be sparsely or unevenly distributed.

Therefore, the traditional SLR approach to model more realistic — and hence more practical and complex — selection processes is often problematic.

However, with recent developments in statistical theory, and the advent of powerful computers and numerical algorithms, it is now possible to perform the FET's analogue in logistic regression.

Company XYZ would still model the same selection process but leave out the parametric distributions and rely on the observed empirical distribution.

Like the FET, the Exact Logistic Regression (ELR) technique is appropriate for small sample sizes of data, or data that are sparsely or unevenly distributed.

Let's now return to the example to illustrate the value of the ELR relative to that of the traditional model. Assume that XYZ's three divisions have the following demographic and RIF distributions:

Division 1: Total 30 with 15 male and 15 female. Selected 3 with 1 male and 2 female.

Division 2: Total 40 with 20 male and 20 female. Selected 3 with 1 male and 2 female.

Division 3: Total 30 with 15 male and 15 female. Selected 14 with 4 male and 10 female.

The aggregated data is identical to the earlier example. Assuming that gender and division are the determinants of the RIF selection process, the SLR model will find the probability of a female over-representation to be 0.031, which would support the conclusion of adverse impact.

The more appropriate ELR model, however, finds the probability of the relationship between RIF selection and being female to be 0.052, which, since it is greater than 0.05 would not support the conclusion of adverse impact.

So if one incorrectly used an SLR model, one would conclude falsely that the RIF is having a disparate impact on female employees.

The failure of the CST and SLR with small sample size or sparsely or unevenly distributed data is also illustrated by comparing the results of both models when testing for adverse impact using the firm's overall workforce, with gender being the only determinant of the RIF selection process.

Since the two models use the same data to measure the same relationship, you would hope they would produce identical results.

Unfortunately, according to the CST, the probability of the female over-representation is 0.046, while according to the SLR it is 0.051. The FET and ELR are not marked by this inconsistency; both models estimate the probability to be 0.078.

One more general problem with the SLR: it will not return a result if the number of employees in any category is zero (e.g., no women were selected for the RIF in division one).

Therefore, even when the sample size is large, SLR may still not be appropriate. The ELR technique, however, does not have this problem and remains a valid option, even when the number of employees in one or more categories is zero.

## Conclusion

As this example displays, when companies or their attorneys conduct a pre-RIF adverse impact analysis or analysis for litigation, the choice of statistical test is vitally important.

Executives and attorneys involved in RIF-related issues should not blindly accept a test because it is commonly used or proposed by an expert without understanding that the limitations of such tests can erroneously indicate the existence of adverse impact where none exists. It can make the difference between employers prevailing at trial or paying millions in damages.

They should be especially careful when using large sample parametric models for samples that are either small or sparsely and unevenly distributed.

As father of the FET R.A. Fisher noted, such an application is not only like using "a cannon to shoot a sparrow," but also a cannon that "misses the sparrow!"

--By Debo Sarkar (pictured) and Mark Gustafson, Analysis Group Inc.

*Debo Sarkar is a vice president with legal consulting firm Analysis Group in the New York office. Mark Gustafson is a vice president with the firm in the Los Angeles office.*

*The opinions expressed are those of the authors and do not necessarily reflect the views of Portfolio Media, publisher of Law360.*

---