Properly Responding To The EU Pay Transparency Directive: Limiting False Positives And Unnecessary, Distortionary Action

Brian Levine, PhD[*] and Tatu Westling[**]

In the near future, reporting on pay equity issues will be required under the EU Pay Transparency Directive (EUPTD). In each country where the employee headcount exceeds a statutory threshold—150 employees under the EUPTD—companies will need to calculate and document gender pay gaps calculated within "worker categories." Where the gaps are larger than 5%, companies will need to coordinate with worker representatives to resolve those differences. Given the potential costs and issues associated with this mandated resolution process, it is imperative that companies assess pay equity now. They should do so in the right way so that they can set up a process to identify areas of genuine risk (as opposed to "false positives") and act proactively before the reporting deadline arrives. Two steps will be discussed, both of which are critical to such an assessment.

Two steps will be of critical importance for such an assessment. First, legitimate factors driving pay differences should be taken into account. This is a step allowed for by the EUPTD and achievable using multiple regression, long the standard for pay equity review. Legitimate factors taken into account will be associated with "skills, effort, responsibilities, and working conditions" and are best if readily defensible to government auditors, worker representatives, and the

[*]*BRIAN LEVINE* is Partner and Pay Equity Leader at Merit Analytics Group LLC. He has extensive experience in People Analytics, with a deep specialty in supporting DEI and pay equity efforts. Mr. Levine spent 23 years as a workforce scientist at Mercer, where, after launching the firm's pay equity offering, he held various leadership roles for more than a decade, including leading the analytics business, and, most recently, driving the commercialization of pay equity for the global firm. He has been the principal pay equity consultant for more than 20 of the *Fortune* 100, has led hundreds of pay equity studies, most global in scope, and has significantly evolved related methodologies over his career to ensure companies can achieve real progress and document that progress in their public reporting. Mr. Levine regularly meets with executive leaders and boards to review analysis results. His work crosses industries, though he has been heavily engaged in work for the financial services, technology, retail, consumer, and life science sectors. Mr. Levine frequently publishes in the professional press and speaks at conferences. In 2021, he published his research assessing compensation philosophies within and across firms, from an examination of more than 50 pay equity studies, in *The Journal of Total Rewards*; he later presented that research at the WorldatWork Rewards' 22 conference in Atlanta. He has been quoted in many publications, including *The New York Times, The Wall Street Journal,* and *The Washington Post.* Mr. Levine has Ph.D. and M.S. degrees in Economics and a B.S. in Industrial and Labor Relations with honors, all from Cornell University. He has served on the adjunct faculty of both New York University and Baruch College.

[**]*TATU WESTLING* is the Organization Development and Workforce Planning Lead at Wärtsilä. Prior to his current role, he was Director, People Analytics at UPM—The Biofore Company. Before industry, Mr. Westling spent 15 years in consulting, advising clients in the people strategy and analytics domains. Most of his career he has been based in Helsinki and worked with large firms headquartered in Finland. At Mercer, he also worked in the Workforce Strategy and Analytics group based in New York. Mr. Westling has a M.Soc.Sc. degree in Economics. He completed Ph.D. courses in the Finnish Doctoral Programme in Economics (FDPE). Both are from the University of Helsinki.

© 2026 Thomson Reuters

courts. Second, worker categories need to be thoughtfully defined to ensure that similarly valued employees are compared and that those categories are not too small—as, in the latter case, risks will be erroneously identified. Indeed, the purpose of this paper is to make plain why small worker categories will lead to wasted efforts and expenditure, directed to address non-existent problems.

Companies should exercise the discretion they have to define worker categories by ensuring they are set defensibly and optimally. While the temptation will be strong to define worker categories narrowly (for example, by job or job family within grade), the result will be a surprising number of false positives. This is because the absolute standard established by the EUPTD (gaps larger than +/-5%) plays out differently than the old standard of statistical significance. Under the old standard, small groups would be unlikely to show statistically significant differences due to large potential variance. Under the new absolute difference standard, small groups can be linked frequently to large gaps driven by that same, large potential variance and idiosyncratic cases. And these gaps will be frequently found against men as well as women. Spending money to address pay issues in such circumstances will not address real underlying issues and may well create new misalignments.

To better identify real issues, define worker categories to be broad employee groups with low pay variance (that is, pay grades over jobs). Such a definition will also be highly defensible, in its alignment to the EU's similar-value criterion.

THE LOGIC: REPORTING UNDER THE EUPTD WILL SHOW FALSE POSITIVES IN SMALL GROUPS

Most organizations under the EUPTD rules will show false positives. These are cases where analysis shows that there are worker categories with gaps that are larger than 5% in absolute value, even though the organization truly has no systematic pay gap against either gender. Any organization which has a compensation scheme where pay is allowed to vary around a compensation grade mid-point (for example, the widely used 90% to 110% or 75% to 125% rules around the mid-point), will likely have false positives. Most organizations with white-collar workforces (or workforces not covered by collective labor agreements) will have such intended pay variation. For these organizations, there are two factors which contribute to false positives: the headcount of the worker category and the intended pay variance.

All else equal, smaller headcount in a worker category will yield more false positives. The reason is that, in smaller groups, the idiosyncratic pay variation dominates. A few high- or low-paid individuals can substantially move the group average. For instance, in a group of five women and five men, even one very highly paid woman can result in a 5% gap against men. In groups with large headcounts, the individual variations serve to cancel each other out.

Further, all else equal, a compensation scheme with larger pay variation yields more false positives. A scheme with a 75% to 125% rule will provide more false positives than a 90% to 110% rule. The reason is straightforward: there is simply more room for idiosyncratic variation in pay. Importantly, this variation is present among both genders, and, therefore, the worker-category gaps can go in either direction.

These observations have practical relevance. First, many organizations—even large multinationals—might have quite small presences in many European countries. In these countries, total headcounts can easily be less than 100. Second, when these employees are scattered among numerous worker categories, counts will be even smaller. Finally, most organizations have

© 2026 Thomson Reuters

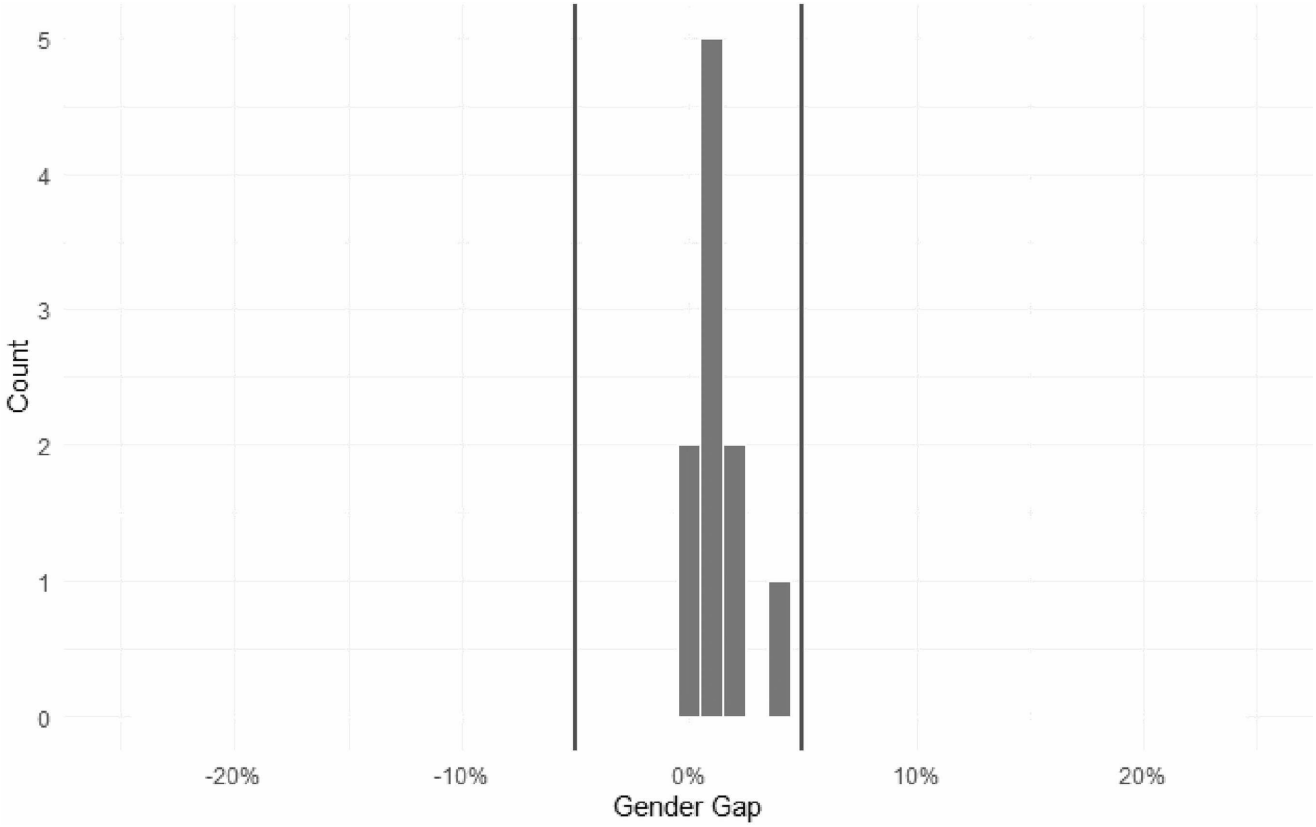
compensation schemes with 75% to 125% or even 70% to 130% rules. As a result, many organizations under the EUPTD rules will end up with false-positive gaps.

AN EXAMPLE: WHEN EUPTD REPORTING SHOWS FALSE POSITIVES

To examine the potential of required reporting under the EUPTD to generate false positives, the authors simulated data for a workforce of 5,000 where pay varies across 10 pay grades but not across 10 job families or genders. Pay is assumed to vary by about 10% across adjacent pay grades, and the standard deviation of pay is assumed to be about 20%. Women are assumed to be more heavily represented in lower pay grades, as is often the real-world case.

Per the design, analysis of gender pay gaps for each of the 10 grades—each with 500 employees—show that the absolute value of no gap to be larger than 5%, as can be visualized in Figure 1:

Figure 1: Histogram Showing Frequency of Pay Grades, by Size of the Gender Gap

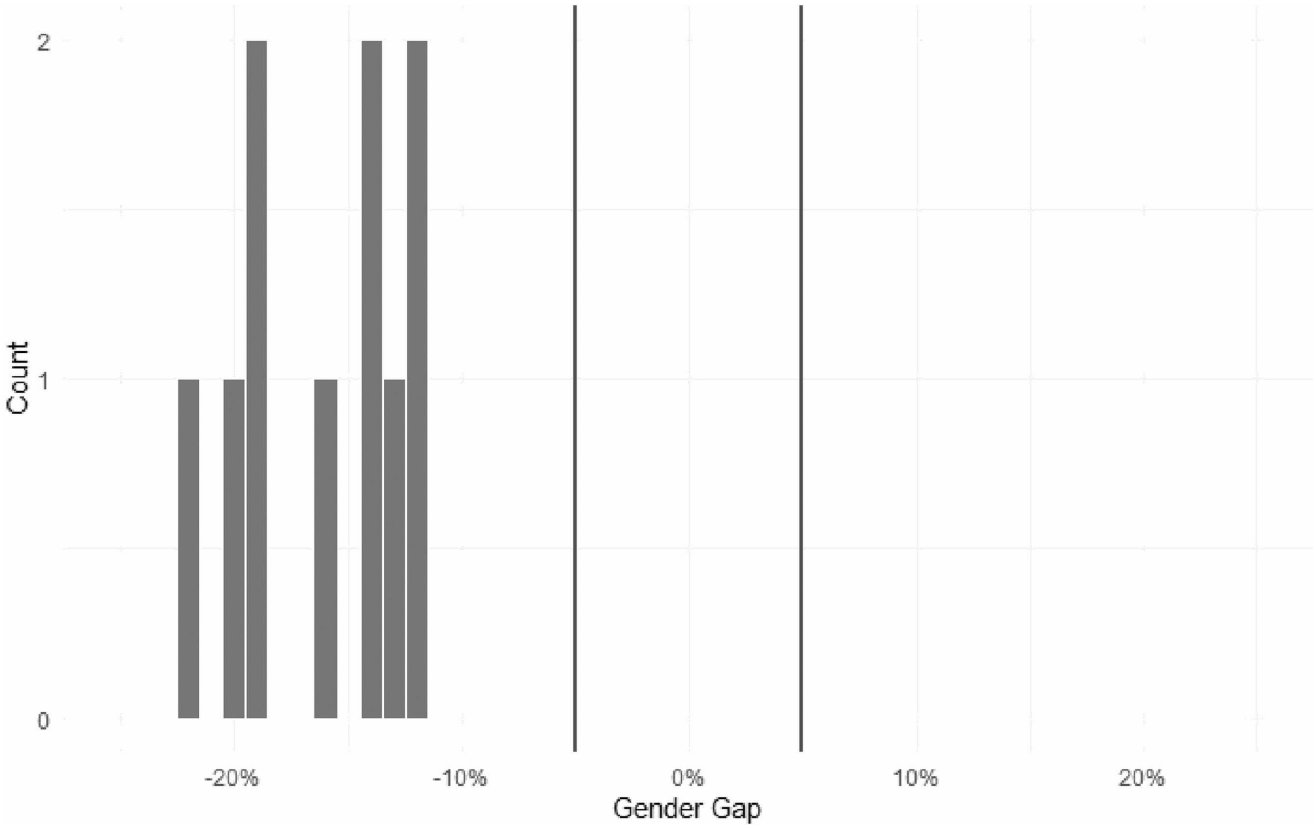


Further, no gap is statistically significant at the 5% level.

If instead, one were to look at the gender pay gaps by job family, without controlling for grade, all of the gaps would be less than 5%, and larger than 5% in absolute value, as seen in Figure 2:

© 2026 Thomson Reuters

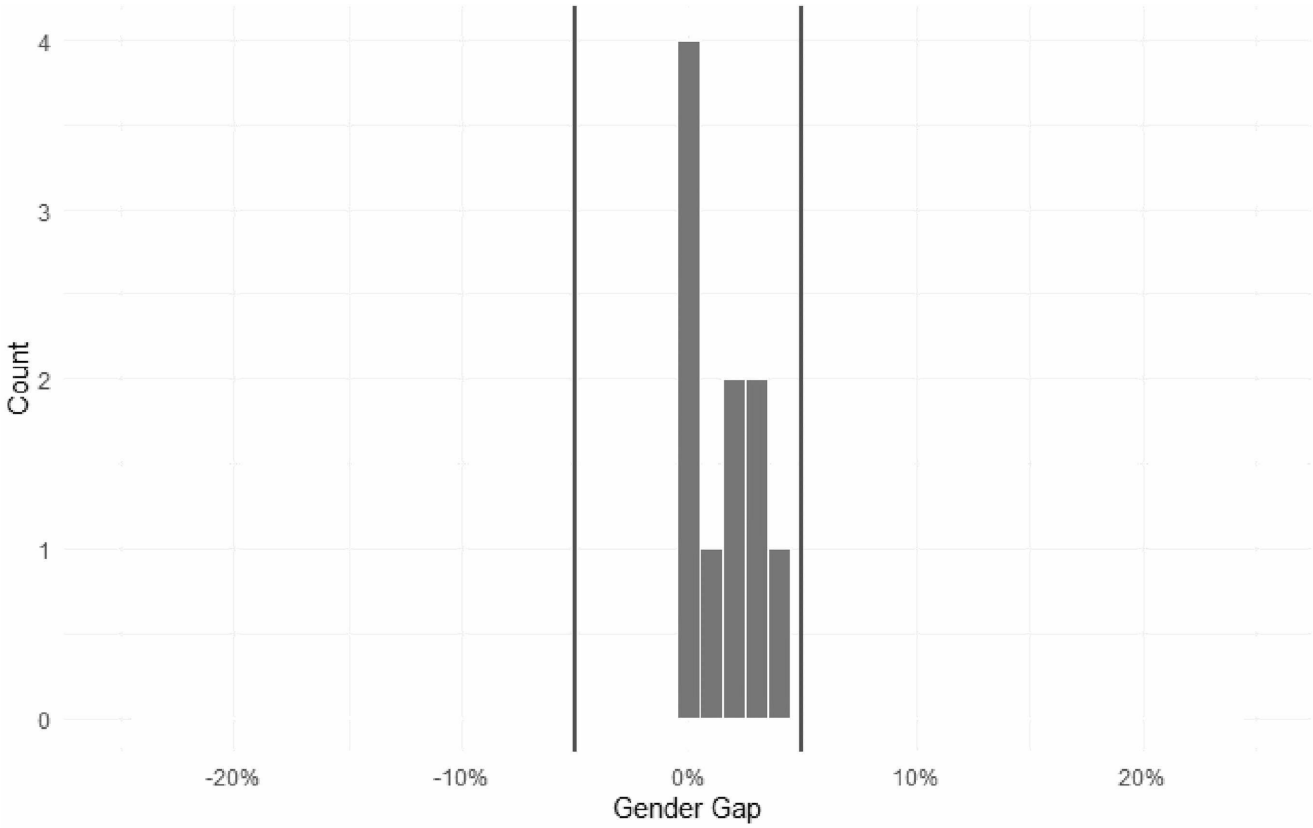
Figure 2: Histogram Showing Frequency of Job Families, by Size of the Gender Gap



Relatedly, all of these gaps are statistically significant at the 5% level. The implication is that one needs to account for key factors that would legitimately drive the gaps (such as differences in roles occupied) before their calculation, as is allowed by the EUPTD. In fact, when a regression is used to account for pay grades before calculating gaps, the job family gaps are no longer larger than 5% in their absolute value, as seen in Figure 3:

Figure 3: Histogram Showing Frequency of Job Families, by Size of the Gender Gap, After Accounting for Pay Grades

© 2026 Thomson Reuters

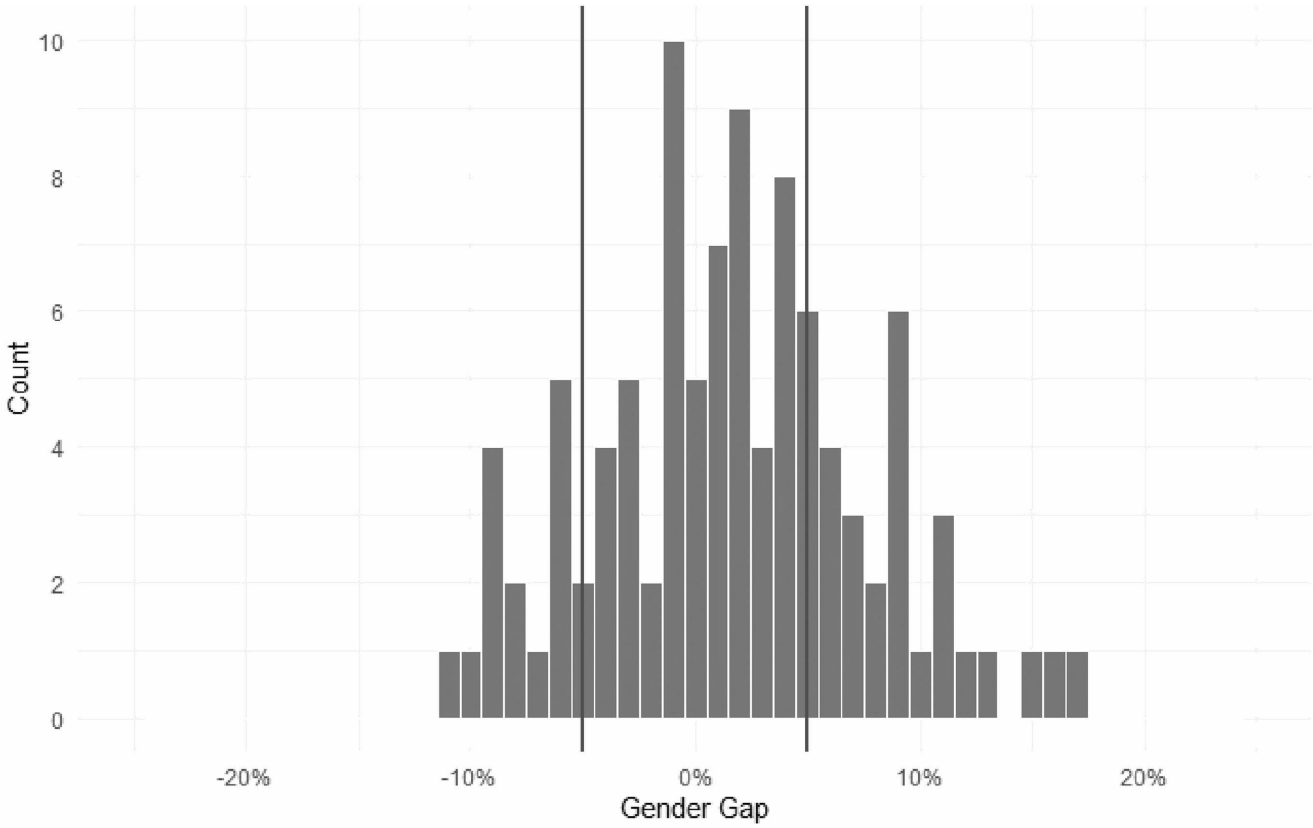


Again, no gap is statistically significant at the 5% level.

Critically, when one looks at gaps more narrowly, by job family and pay grade, one would not expect to see any differences, because, by design, there are no systematic differences by gender. Yet, when the population is split this way—into 100 worker categories of 50 employees each—almost half of them show differences by gender, as seen in Figure 4:

Figure 4: Histogram Showing Frequency of Job Family-Pay Grades, by Size of the Gender Gap

© 2026 Thomson Reuters



The simulated example shows that almost half of the worker categories show differences by gender. In real-life situations many groups will be even smaller than 50. It is likely that in such situations, more than half would show false positives.

Statistical significance as a criteria would tell a different story. In the above case, if statistical significance at the 5% level was used to determine areas of risk, only three areas would be identified—which is about as expected across 100 tests where the true gap is zero. While it may not matter given the specifics of the EUPTD, statistical significance is clearly seen to be a superior indicator of a true issue than the absolute 5% threshold.

The issue of false positives under the 5% threshold standard is much more of an issue in small workforce groups. This reinforces the recommendation that broad workforce groups, all else equal, will lead to more accurate measurement and better direction for remediation efforts.

ADDRESSING ISSUES IDENTIFIED

In traditional pay equity analysis, adjusted pay gaps can be addressed by making pay adjustments for "negative outliers"—employees paid well below what the regression model predicts they would be paid. For analyses conducted under the EUPTD, one can look to the same outliers to reduce the magnitude of the adjusted gap in a given worker category to be less than 5%. This and the fact that the EUPTD allows for legitimate factors to be taken into account before calculating gaps that must be addressed in a worker category support conducting EUPTD analysis concurrently with the regular pay equity review.

© 2026 Thomson Reuters

One legitimate factor that will be relevant to international companies will be unique compensation arrangements for expatriates. As the EUPTD applies to all employees in working in a given country, expatriates must be included in the review but might reasonably be excluded given specific contractual arrangements. If an employee is an EU citizen working in a different EU country, there would be some reasonable question about the country to which the employee should be assigned to for analysis; the most effective assignment would position the employee in the country that is most applicable for compensation setting.

Companies should strive to align the timing of analyses and associated methodologies across pay equity efforts, wherever feasible, to realize synergies and drive efficiencies.

CONCLUSION

While the EUPTD is focused on an important issue, the criteria for evaluation—the magnitude of the gap in worker categories—is problematic. This is because gaps larger than 5% in absolute value are likely to occur by chance, especially if employee counts in categories examined are small. Analyses will more accurately identify areas of risk in cases where the worker categories are broad. Relatedly, broader worker categories will limit pressures to make pay changes to address spurious gaps—pay changes that do not address true issues and that, themselves, can cause misalignments. That said, one should strive to define worker categories based on similarity of value in roles, leveraging pay grades or bands to ensure that those categories are defensible. Where employees must be combined across roles of differential value, it is important to account for differences between employees using regression analysis. In many organizations, where the headcount in EU countries is small, it will be challenging to escape the false positive problem, and readiness to explain employee-specific compensation outcomes will be critical.

Statistical significance would have been a better condition for evaluation for the EUPTD than the absolute gap criterion, because, again, gaps larger than 5% in absolute value are likely to occur by chance. Notably too, some gaps that are smaller than 5% will be statistically significant and reflective of actual, detectable bias. Statistical significance can still play a role, pointing proactive organizations to areas of risk even when they the gaps are less than 5%; statistical significance can serve as a secondary screen.

A further, interesting fact is that the EUPTD can fail to identify the real, million-Euro problems: systemic, organization- or country-wide gaps. It is, itself, focused on pockets of issue and, given the criteria, pockets that will be more idiosyncratic (and random) than systematic (and reflective of actual bias). Broader pay equity analysis is still a recommendation to counter such broader risk.

A final point: the many above caveats aside, it is important to consider your practices and whether any refinement is necessary before running any "EUPTD tests." If your grades are not well-defined or your bands are too broad, the looming EUPTD deadline suggests that now is the time to review and refine the compensation structure.

REFERENCE: CODE USED TO GENERATE TEST DATA AND ASSOCIATED ANALYSIS

```
library(ggplot2)

library(dplyr)

set.seed(123) # for reproducibility

# Parameters

n_people <- 5000

n_grades <- 10
```


© 2026 Thomson Reuters

```
people_per_grade <- n_people / n_grades

# Create data

pay_data <- data.frame(

id = 1:n_people,

pay_grade = rep(1:n_grades, each = people_per_grade)

)

pay_data$job_family <- round(runif(nrow(pay_data)) * 9+1, digits=0)

# More men in higher grades

pay_data$gender <- sapply(pay_data$pay_grade, function(g) {

prob_male <- 0.4 + (g - 1) * 0.05 # increases with grade

ifelse(runif(1) < prob_male, 0, 1)

})

# Checks

pay_data %>% group_by(pay_grade) %>% summarize(count=n(), pct_fem=mean(gender)) %>% View()

pay_data %>% group_by(job_family) %>% summarize(count=n(), pct_fem=mean(gender)) %>% View()

# Generate random pay according to log specification, variance about +/- 20%

pay_data$logpay <- 11 +.1 * pay_data$pay_grade + rnorm(nrow(pay_data),0,.2)

pay_data$pay <- exp(pay_data$logpay)

## LOOP BY GRADE

gender_coeffs_grade <- data.frame(

pay_grade = integer(),

count = integer(),

gender_coef = numeric(),

p_value = numeric()

)

for (i in 1:10) {

tempdata <- subset(pay_data, pay_grade == i)

if (nrow(tempdata) > 1) {

# model <- lm(logpay ~ gender + job_family, data=tempdata)

model <- lm(logpay ~ gender, data=tempdata)

gender_coeffs2 <- data.frame(

pay_grade = i,

count = nrow(tempdata),

gender_coef = coef(summary(model))[”gender”, “Estimate”],

p_value = coef(summary(model))[”gender”, “Pr(>|t|)”]

)

gender_coeffs_grade <- rbind(gender_coeffs_grade, gender_coeffs2)

}

}

# Histogram

ggplot(gender_coeffs_grade, aes(x = gender_coef)) +

geom_histogram(binwidth = 0.05, fill = “blue”, color = “white”) +

geom_histogram(binwidth =.01, fill = “steelblue”, color = “white”) +
```


© 2026 Thomson Reuters

```
geom_vline(xintercept = -.05, color = ‘‘red’’, linewidth = 1) +

geom_vline(xintercept =.05, color = ‘‘red’’, linewidth = 1) +

scale_x_continuous(limits = c(-.25, .25), labels = percent_format()) +

labs(x = ‘‘Gender Gap’’, y = ‘‘Count’’) +

theme_minimal()

## LOOP BY JOB FAM

gender_coeffs_jf <- data.frame(

job_family = integer(),

count = integer(),

gender_coef = numeric(),

p_value = numeric()

)

for (j in 1:10) {

tempdata <- subset(pay_data, job_family == j)

if (nrow(tempdata) > 1) {

# model <- lm(logpay ~ gender + pay_grade, data=tempdata)

model <- lm(logpay ~ gender, data=tempdata)

gender_coeffs2 <- data.frame(

job_family = j,

count = nrow(tempdata),

gender_coef = coef(summary(model))[’’gender’’, ‘‘Estimate’’],

p_value = coef(summary(model))[’’gender’’, ‘‘Pr(>|t|)’’]

)

gender_coeffs_jf <- rbind(gender_coeffs_jf, gender_coeffs2)

}

}

# Histogram

ggplot(gender_coeffs_jf, aes(x = gender_coef)) +

geom_histogram(binwidth =.01, fill = ‘‘steelblue’’, color = ‘‘white’’) +

geom_vline(xintercept = -.05, color = ‘‘red’’, linewidth = 1) +

geom_vline(xintercept =.05, color = ‘‘red’’, linewidth = 1) +

scale_x_continuous(limits = c(-.25, .25), labels = percent_format()) +

scale_y_continuous(breaks = seq(0, 2, by = 1)) +

labs(x = ‘‘Gender Gap’’, y = ‘‘Count’’) +

theme_minimal()

## LOOP BY JOB FAM, WITH GRADE CONTROL

gender_coeffs_jf_grade_control <- data.frame(

job_family = integer(),

count = integer(),

gender_coef = numeric(),

p_value = numeric()

)

for (j in 1:10) {

tempdata <- subset(pay_data, job_family == j)

if (nrow(tempdata) > 1) {
```


© 2026 Thomson Reuters

```
model <- lm(logpay ~ gender + pay_grade, data=tempdata)

# model <- lm(logpay ~ gender, data=tempdata)

gender_coeffs2 <- data.frame(

job_family = j,

count = nrow(tempdata),

gender_coef = coef(summary(model))[''gender'', ''Estimate''],

p_value = coef(summary(model))[''gender'', ''Pr(>|t|)'']

)

gender_coeffs_jf_grade_control <- rbind(gender_coeffs_jf_grade_control,gender_coeffs2)

}

}

# Histogram

ggplot(gender_coeffs_jf_grade_control, aes(x = gender_coef)) +

geom_histogram(binwidth =.01, fill = ''steelblue'', color = ''white'') +

geom_vline(xintercept = -.05, color = ''red'', linewidth = 1) +

geom_vline(xintercept =.05, color = ''red'', linewidth = 1) +

scale_x_continuous(limits = c(-.25, .25), labels = percent_format()) +

labs(x = ''Gender Gap'', y = ''Count'') +

theme_minimal()

## LOOP BY GRADE AND JOB FAM

gender_coeffs_grade_jf <- data.frame(

pay_grade = integer(),

job_family = integer(),

count = integer(),

gender_coef = numeric(),

p_value = numeric()

)

for (g in 1:10) {

for (j in 1:10) {

tempdata <- subset(pay_data, pay_grade == g & job_family == j)

if (nrow(tempdata) > 1) {

model <- lm(logpay ~ gender, data = tempdata)

gender_coeffs2 <- data.frame(

pay_grade = g,

job_family = j,

count = nrow(tempdata),

gender_coef = coef(summary(model))[''gender'', ''Estimate''],

p_value = coef(summary(model))[''gender'', ''Pr(>|t|)'']

)

gender_coeffs_grade_jf <- rbind(gender_coeffs_grade_jf, gender_coeffs2)

}

}

}

# Histogram 1

ggplot(gender_coeffs_grade_jf, aes(x = gender_coef)) +
```


© 2026 Thomson Reuters

```
  geom_histogram(binwidth =.01, fill = ‘‘steel-
blue’’, color = ‘‘white’’) +

  geom_vline(xintercept = -.05, color = ‘‘red’’,
linewidth = 1) +

  geom_vline(xintercept =.05, color = ‘‘red’’,
linewidth = 1) +

  scale_x_continuous(limits = c(-.25, .25), labels
= percent_format()) +

  scale_y_continuous(breaks = seq(0, 10, by =
2)) +

  labs(x = ‘‘Gender Gap’’, y = ‘‘Count’’) +

  theme_minimal()

  # Histogram 2

  ggplot(gender_coeffs_grade_jf, aes(x =
p_value)) +

  geom_histogram(binwidth =.01, fill = ‘‘dark-
cyan’’, color = ‘‘white’’) +

  geom_vline(xintercept =.05, color = ‘‘red’’,
linewidth = 1) +

  scale_x_continuous(limits = c(0, 1)) +

  # labs(x = ‘‘Gender Gap’’, y = ‘‘Count’’) +

  labs(x = ‘‘Gender Gap P-value’’, y = ‘‘Count’’)
+

  theme_minimal()
```

The views and opinions expressed in this article are solely attributable to the authors and do not reflect those of the authors' companies. Further, no findings are attributable to pay analysis performed at any specific company.

© 2026 Thomson Reuters