ROUGH EDITED COPY

>> Hi, welcome. My name is Joanna and I am representing the mid Atlantic region and the Washington D.C. aisle G. I am pleased to be here virtually, but certainly wish that we were altogether in national harbor, Maryland where the conference was supposed to be held. I am actually a Maryland resident and it is a really beautiful venue. So the next time they come back to the Maryland D.C. or mid Atlantic region, I hope you can make it.

I do have a few housekeeping items before we get started with our feature presentation today. This session is being recorded. A transcript will be made available on the N-I-L-G website by the end of the week. There is also a copy of the presentation that will be sent out afterwards to the Webinar attendees, so don't worry about that.

You should have also received an e-mail from Anthony or Tony with instructions for today's session, which included a link to closed captioning if you need it. So those instructions are also in the chat feature of the go-to meeting should you need that.
This session is expected to take 90 minutes, give or take, so we're estimating about 75 minutes of presentation and 15 minutes at the end of Q&A. So the questions and answers will be taken at the end, however throughout the course of the session you can use the chat functionality in go-to meeting box. I will provide the questions to the speakers at the end of the presentation, so no fear, put your questions in that chat feature and we will go through them and queue them up at the end.

Before I introduce our speakers, I want to thank our sponsors.

>> Thank you to sponsors. You see the sponsors on the screen. Thank you all for supporting the N-I-L-G and this Webinar series which was brought to us for free thanks to our sponsors.
Let me introduce our lead speaker for today's Webinar, assessing small group branch services. If we could go to the next slide. Doctor Robert or Bob as we know him is is the acting enforcement. He oversees the statistical program and servers as the expert technical advisor in systemic discrimination cases.
Bob, thank you and welcome. I would like to turn this over to you.

>> Thank you to the sponsors for retaining the session. The original plan, since we have social scientists stationed around the country, was for us to have our annual conference here and then attend at the national harbor and present this panel so that we could speak to contractors directly outside the context of an audit. We have had these conversations facilitated by NILG and others, and it is helpful to have that conversation outside of an audit. We are happy to discuss that today, and listen to those that issue and work with these policies on a daily basis and looking at

contractor data on a daily basis.
However, since it is now virtual format, we weren't able to assemble all of the branch expert services, so you will hear from 5/8s of them. It will be jam packed, I don't want to spend too much time on introduction. I will say that what this session does not try to achieve is to establish policy or indicate a preferred technique. You'll see lots of disclaimers throughout this presentation, but that does allow us to explore some fertile ground that was identified as the round table in many discussions and members of the NILG. So we came up with these three topics. They're somewhat connected, but as you have seen from the title, and I'm sorry we didn't get the title and topic out sooner. We decided to address a small group analysis cohort type analysis and then global analysis. We have bios distributed throughout the presentation to introduce the new topic and the person speaking for you to follow along.

So I want to kick it off and start can Timothy. He works nationwide. Timothy will discuss what the definition of small group is and then some of the techniques that might be available for analyzing those small groups.

>> Hi, I am a mathematician. We often encounter situations where employees end and a specific group is small. We also see situations where companies analyzing their own data incorrectly when it is small. Incorrect data can lead to wrong conclusions. This is because if the number of employees is not large enough, classical statistical measures will for a fixed period of time will make it less reliable. As uncertainty measures associated with the specifics of variance and ranges, are based on large simple properties. It may or may not be correct if we try to generalize the finding. Such small employee groups may include but not limited to following types: This includes small pay analysis groups, PA G. Categories or job groups who are comparable for purposes of analyzing contractors pay practice.
Rare but not impossible. Sometimes following the general guideline, only a few employees are found in the pagePAG. It may be the CEO or a few top executives of a mid size company.

The second type of a small employee group may be gender or race categories. This happens a lot given the low numbers of minority works in management levels.

The third type is applicant pool for specific position of collective positions, for example, the computer programmer may be further divided in to positions with strong fluences like C plus plus, database or Internet systems.

So the applicants for the general programming positions are many, but there may be only a few applicants considered to articulate in each of the computer languages. This job posting of programmer will be advertised separately as C plus plus programmer, database programmer or Internet programmer jobs and consequently advocacy of these job titles will be few. We see this happen often with increased positions of jobs.

Another type is the multiple hiring reck question situations in hiring. Employees, especially seasonal employees, often hire through multiple requisitions spread over extended period of time, a year for example. And only offers a few positions depending on business needs. All basic quantities may be the same. Not all applicants are present if we treat each the same, the applicant may be small. A number of factors determines this. One is the effect size. Small group statistics do not provide reliable statistics as large groups do. Though it may be plausible, it does just as well for a number of reasons.

Recognizing the challenges of small numbers, a number of statistical approaches have been developed historically. This includes different methods, includes non Pam rah metric methods, nearest neighbor matching, or cock ran man tell method or BA Y-E SI A-N methods. We will discuss some of this later in the presentation.
There are also other simulation techniques. These techniques require broad information input, more research of variable structures and model assumptions. Consequently they may be constrained by turn around time. Nevertheless, these are useful techniques. We will explore this later in the presentation.
Now I turn it to my colleague, Ryan, to next present.
Thank you.

>> Good afternoon. Some tools are not reliable or usable. If you're trying to assess a group of 20 applicants or employees, you kind of can't be running a regression analysis. So I will talk about the alternatives briefly, but first I want to emphasize if small samples are in fact all you have, then there are other options if you glean information. Keep in mind by definition small sample tests are small, they only have a little bit of information.
Without further adieu, one thing you might do with a small sample is let's just skip statistics entirely. Let's look for people who are essentially carbon copies of each other, but they differ in only protected class. They have the same education, same experience, same time with the company and so on and so forth. And then see, hey, do they have a big difference in pay or, you know, or if there are lots of them in higher rates, look for -- or two people, one is retired, one is not. One is better, at least good or better in all cases and they do worse.
No math needed, no fuss. Thinking of an obvious problem here, one reason why you might be looking at using a small sample, we have a small sample, is maybe these are a group of individuals who are unusual in some way. Small samples, people don't always look that much alike. Maybe there are differing ways you can talk about numerically, this person has more years of experience, but they might not be -- they're more likely to be carbon copies of each other. Partly because they're different, but also simply you're unlikely to get carbon copies if you have very few people.

This is not deposited. You pay one person twice as much as the other and there is no way to tell them apart other than projected class, but not having failed that test is not necessarily -- doesn't necessarily convey information. You may have a problem or may not, but it doesn't show up in that big red flashing sign sort of way.

Some other examples. In certain circumstances you might use fisher's exact test. I don't want to spend a lot of time now speaking technically here, but I want you to understand the name. Fisher's exact test, yes, that's great, that should solve my problems. Maybe, but the word exact here is talking that can tell you about small samples, that it sort of -- if you think about trying -- intuitively I'm trying to tell the difference of if the coin is not -- if it is a fair coin or not. The only way to have to get any information about this coin is flipping it.
I am flipping it a few times. The only mathematical tool I have in my mental tool box is logistic regression. Well, I can't tell you anything there. What makes fisher

exact testing exact is that it can say something if you flip the coin multiple times, but that doesn't mean it can tell you for sure if it is a bias point or not. I guess nothing is ever certain in statistics, but it certainly can't tell you with the sort of degree of certainty one way or the other that would make you feel comfortable.

There is also rank sum tests. A great advantage here is the non parametric and principal, you can work whenever you're ranking outcomes. And again, they work in small samples. You don't have to make any sort of simplifying assumption that we normally can make with a larger group.

But we are losing some information. Ranking every one from one to 15, say, in terms, and you put them in two sub samples, male/female, black/white, whatever.

You would expect if it is just random, different totals would tell you -- similar groups should have similar totals. One has high numbers and one has unusually low numbers. The absence of a finding is not necessarily dispositive. What do you do with that information? We will talk about that later. I will turn it over to my colleague now.

>> Hello, good afternoon. How are you doing today? My name is Andy. Today I will introduce one method called a K-nearest neighbor for example, Andy has a salary of $100,000. My name is Andy, and Andy has the same qualification, education, whatever. That's how the K gets in the process.

K may be in good alternative method or predictions when dealing with a small sample size. It worked in 1970, but became popular when computation speed was greatly improved.

We try to explore that data set. We look at different attributes when we analyze this points. But the K nearest neighbors is a non athematic method. Based on the given data, to find the best K neighbor. For example, it can be assigned one point, 2 points, and then three points and so on and so forth.

These other points are how close distance wise they are from the interest point. The average value of the K nearest neighbor is then taken to be the final predictions. The future attributes are not limited to time involved in working with a company, experience, performance ratings, management label, job group, or type, or full-time or part-time status, exempt status, department category, staff status and locations. In order to capture that distance, the attributes between the interested or predictor objects and reference of observation, an algorithm, beauty lies in the distance. We use this here to look at this scale. It is to find the nearest number for a predicted value. When K is too small, the results were over 15. When the results were under. In general, there is a method that can obtain the root mean squared, we can compare this for different variety of K and then choose the K that best represents.

Let's look at the graph. We have two attributes here. One is a height, and we can predict this on given data, we can predict how much weight there is for the person.

We have eight and we have a given data ready. Right now we try to predict 38 years old with a 5.5 height. This shows on the graph with a question mark and big red circle. We do the average, 48 and 60 and we have 59. But if we want to include a three, that would probably be above or the one next to the circle, 58. So this is very easy to apply. We can continue to look at variables, but it can also be applied to categorical variable. For example, it can be used in hiring case, promotion cases and many more. This process has been applied in many industrial fields such as medical, news, banking, planning, and users.

In conclusion, there are several reasons to why K and N is beneficial and is a significant method to utilize K and N.

It is applicable to small samples. It does not require any assumptions. I will stop here and let Brett continue on other issues. Thank you.

>> Thank you, that was a lot of great information. David, go ahead and go two slides over.

We will switch gears now. I will present an idea regarding cohort analysis, which I had 7 or 8 years ago. This isn't something that O FCC P would use in its enforcement activities, but instead I am introducing it to the contractor community as an idea for their self audit of pay.

So just to briefly introduce the topic when I am talking about cohort analysis, it is a generally non statistical technique where individual employees are compared to determine whether pay disparity exists by some protected status or race. Classically this is done over time, but however the EEO analysis is often presented with cross sectional or snapshot data, so we will be talking specifically about doing cohort analysis in snapshot data.

And generally, there is a lot of different techniques to doing it, but you're basically just looking at a single spreadsheet of individuals. And just try to sort it and compare people to try to detect under paid individuals. Just make an eyeball comparison so establish the employees, the similarity in employees qualifications and pay.

If we go to the next slide, there are some problems with that approach. Trying to compare people using multiple factors simultaneously is pretty hard work from just a cognitive perspective. We're asking the analyst to essentially try to give different information just sort of in their head.

As a result, the analysis will only folk on the most direct comparator set which narrows the study to an equal pay act analysis.

And another big problem with doing cohort analysis is that there's really no expectation for what anyone's pay should be just eyeballing the spreadsheet. There is no real way to determine who is over paid or under paid and by how much.

So if we go to the next slide, the idea I had back in 2012 or whatever was to conduct cohort analysis using a technique called multi attribute utility analysis. And this is borrowed from decision theory. It is a non statistical method that produces a weighted sum utility score to compare people across multiple attributes simultaneously. This breaks the evaluation down in to simpler components which makes it easier on the analysts.

If we go to the next slide, now traditionally this technique is used in policy analysis to scale different policy options in terms of their preferability in data. It is a stretch on the face to think about this in terms of pay equity. But one way we could make that leap is to think about pay in terms of the employers expressed preference for each of its employees.

And then human capital factors associated with each employee or their pay related factors are then objective measures of each employee's preferability.

So then we can think about matching the employee's preference to the employee's preferability.

So there is an eight stop process, and I have an example I will walk through.

But we need to gather our data and we require a set of attributes along with each employee that can be at least ranked.

And there are some examples there and we can get that with a spreadsheet along with race, gender and pay.

The next step, or actually, this is the actual example data. This is data from a federal contractor. This is a cohort of 13 employees that we were looking at. And the first six columns are the data submitted by the contractor, obviously we would be identified, those employees. And then the three yellow columns are to calculate the vase your tenured variables that we could typically use in an analysis of pay, total years of service, and then years in current job which is based off the effective date of current job code. And then differences in columns G and H which would be years in other jobs.

So that is our basic data.

If we're looking in terms of ranking employees, we can rank them in ascending order where one is the lowest, and this is along each attribute. One is the employee scoring the lowest on the attribute and N would be the highest where N is the number of employees. We're not worrying about pay rate or gender at this point, we'll deal with those later in the process. Once data was ranked, we would refer to that as utilities using the following formula.

Zero would be the employee lowest on whatever attribute we're talking about, and one is the highest. Now, again, this is for strictly rank data, if you have data that are interval level, that means actually measuring amounts of things such as tenure variable in our sample, we can use more sophisticated formula that is on the next slide.

Then we can compute utilities using this formula.

I will just note here that one of the strengths of this process is that we don't have to have data that is all interval or all ranks, we can actually mix. We can have some attributes that are interval level data and some that are rank level data and we just apply the appropriate utility formula to each attribute and generate the utility values accordingly.

If we go to the data sample now, you recollect see in the yellow highlighted column that I have taken the data in columns H and I, years in current job and years in other jobs, and using the utility formula on the previous slide, I have actually calculated the utility values. You can see employee number one, they have a one because they have the largest number of years in current job.

Where as on the other hand, employee number seven has a one for other job because they have the highest number of years in other jobs.

This is a simple example. We could be using more factors than this, but we wanted to keep the examples simple for this presentation.

If we go to the next slide, once we have computed the utilities, we would then calculate the weight of the attributes.

And the attributes range from zero to one and reflect the proportion of pay believed to be influenced by a particular attribute.

The most simple scheme could be to weigh them equally which is done by the formula on the screen there, but other weighting systems are possible and we'll talk about different weighting systems later.

But if we go to the next slide, we will show you an example of equal weighting. In this example, we only have three attributes on which we're weighting these

employees.
We will do an equal weighting scheme, so each weight will be one divided by two which is the number of attributes and so we get 0.5 for each attribute. That's what you see above the yellow utility. Again, other systems are possible and we will talk about that a little later.
So we have the attributes -- I'm sorry, we have the utility and we have the weight, and now what you want to do is put that information together. So you would take each employee's utility, multiply it by the weight of the associated attribute, and then some, the weighted scores, to produce the final M-A-U T-score for each employee.
Again, they range from zero to one, where zero would be an employee who was ranked the lowest or scored lowest on every attribute, and one who was ranked highest on every attribute. If we go to the next slide, we will see what that looks like in the example data.

It sort of cut off on my screen, but it should be column L, you should see the M-A-U T-scores for each employee.
You can see employee two is the highest score. Where as the lowest M-A-U T-score, we did get a score, it is .12. That's how you combine all of this information in to a final M-A-U T-score.

At this point you should be going, okay, what? That is not really telling us anything.
So this is how we apply the scores. We use them to actually compute a predicted pay value, which you can see on your screen there. Once we generate a predicted pay, we can then subtract the actual pay from the predicted pay and get what we call in statistical terms a residual. It is just the difference between predicted and actual pay.
And then if we take that residual and actually divide it by the actual pay and multiply it by one hundred, we can convert it in to a percentage term. So it is a little easier to understand and you can see that in the example here. We have shown the percent, under pay and over pay.

And this is where we can start looking at how race and gender kind of fit in to the analysis. You can see that the lone white male, employee number one, according to the MAUT procedure seems to be over paid by about 12 percent.
However if you look at the Hispanic employees, they show a negative residual, so eight out of the 12 Hispanics are under paid. And we also have two females in this mix and one out of the two females are under paid. So we can kind of see that perhaps there might be an issue here with Hispanic employees being under paid in this set of data.
If we move to the next slide, I see a question in the Excel spreadsheet. I mean, you will see images of Excel spreadsheet in the slide deck. We had not intended to share the actual Excel, but talk about the after the fact. You will see the slides.

The 7th step is that we will examine our under paid employees.

What we want to do here, and again, when I say we, O-S cc P is not involved in this. The analyst would look for the percent under paid. So some sort of cut off. As a note here there is no legal consensus on what a practical threshold should be.

If we go to the next slide, step eight, if we decide based on our analysis that we need to distribute remedy to employees within the cohort, we can then allocate it. Obviously O FCC P position would be to make all the class members whole, but within self audit you might decide to remedy just up to practice threshold. And then prospectively, you would want to talk about doing pay adjustments for those class

members. And so again, because prospective pay would be a future fixed expense, if someone is under paid one merit cycle, we would like to remedy that person up to within one merit cycle. Of where they should be. Column 0 is our percent under paid, I bolted the ones under paid three and a half percent. Then in the next column I have taken that number and subtracted three and a half percent to show you what the adjustment should be to get them within one merit cycle.
It is actually a really simple procedure to enact in Excel. You can also add a loop in here to get it within a dollar budget if management wants you to stay within a certain target, that is easy to implement in this procedure also.

But if we move on to the next slide, the amount procedure is going to involve a lot of decision making between the EEO analyst and their management. You have to decide who is in the cohort. You have to decide what attributes to include. You have to decide how should attributes be weighted.
So because those are judgment calls, you may actually want to do this analysis under multiple assumptions and study the effect and that way you can see what the effect for decisions are around that might bring any issues relating -- regarding the calculations.
And I was going to go through an example of this, but in the interest of time let's just skip two more slides now, you can get more sophisticated and add on additional tests based on the residuals, but those tests will be fairly under powered and give you a false sense of security. Also, the point is we're looking at small cohorts in the first place, so presumably we're not necessarily interested in doing large statistical tests at this point.

Any way, to wrap up here -- I have a lot more to say here, but unfortunately we have to cut it short. To wrap up here, again, this is not something O FCC P will do in the course of an audit. This is just an idea I had that might potentially be useful for contractors. You need to review this and see if it works for your organization. So that concludes my portion of the panel. Hopefully you had this informative, but now I will turn the microphone to the next presenter, Dave.

>> Thank you for that very interesting presentation. I think everybody or most in this audience knows that we spent a lot of time talking about pay analysis groups, S-S-E-Gs, two belongs in them, who doesn't, and all those related side conversations. I am going to invite everybody here on to my robot, that's me on the screen, and we will row towards the horizon beyond the PA G.

I guess the theme here is if we're really interested in a systemic analysis, we're missing something by focusing on individual PA Gs.

So the motivation for this is the round table we had back in January, point number seven expressed concerns about isolated indicators, and this is not word for word I don't think from the slides from that round table, but isolated indicators of pay disparities that are not reflective of systemic issues. And -- hold on for a second.

Yes, suggestions for minimum threshold for find consistency in findings. For example, if there are 20 PA Gs, is there a certain number of them that have to show bias, significantly bias indicators to make it a finding.

We share those concerns from this perspective, but also from a different perspective when perhaps a lack of statistical significance for any single PA G that

has been identified, particularly when the small group sizes, should we really be interpreting the over all results as a situation of pay equity?
Let's explore this question.
We will start with two scenarios. I tried to make this as untechnical as possible. Here we have this made up purpose, here we have a contractor XY Z incorporated. On the left is some regression results for 15 different groups, and then you have a test from Excel that also shows the results. So what we see in the chart are the standard deviation results from our typical regression analysis. Many are familiar with this. You can see there is a single PA G with a statistical significant pay disparity in size of four percent. But we don't get any statistically significant indicators elsewhere. This is what I call, and elsewhere that we see, a non statistically significant disparity favoring men and some favoring women. That's why I call this scenario one. Does this look like systemic bias? What I call the imperfect answer is yes because there is a disparity impacting females of four percent. I have seen examples of this imperfect answer coming out of contractor world and certainly out of our own analyses. A more thorough answer, might be, however, that there is perhaps something undesirable happening in PA G No. 5, but it doesn't seem to be part of a larger pattern. And this very simple example represents the concern that was brought up by an expert in point number seven of the computation round table.

There is also another side of this issue, this handsome gentleman on the screen is Ronald fisher of fisher's exact test as we heard earlier. And from his work of 1932, this famous statistician alluded to the issue that we're talking about here. It sometimes happens, quoting on the screen, that although few or none can be claimed significantly individually, yet the aggregate gives an impression that the probabilities on the whole are lower than what would have been obtained by chance. So even back then there was concern about focusing too much on individual tests with no look at all of the overall.
This is the second scenario. Here we have a contractor. There have 15 regressions for 15 PA Gs and there is nothing statistically significant. Does this look like systemic bias to you? The imperfect answer here is no, because the pay system seems to be equitable. However, a more thorough answer is we're not so sure but it doesn't look that good. Why do females earn less than their male counterparts in 14 of 15 PA Gs, which is represented by the 14 blue horizontal columns going to the left. But there is a pattern that is consistent and it is consistently against one under here.

Also some of these PA Gs contain too few employees for analysis. If you look at the right, it is how many employees meet PA G, and certainly No. 14 doesn't have enough for regression and certainly No. 8. And one can argue back and forth whether or not some of these others do.
So the problem here is we have got isolated statistically significant bias indicator that may not be part of a larger pattern. Or a lack of any statistical bias indicator being incorrectly interpreted as an equitable situation. Both of these are not -- both are flawed.

The problem here is we're generally -- we're only assessing within, the within PA G. We're neglecting the aspect of the pay analysis.
And I am not talking about the process, but just generally. Again, we're not seeing the forest before the tree. The current PA G frame work ignores the forest for the trees. A comprehensive pay equity, would require both types of assessment. And this ongoing debate between S-S-E-G and sample size can be partially addressed by a more comprehensive approach. There is a variety of summary tests or nominee bus test. Some are extremely simple to execute that can fill in some of the blanks of a

comprehensive assessment.

So what we will do now, and it will be a bridge because of the amount of time we have, is a demonstration. We will simulate a simple database that represents an employer that discriminates on the basis of gender. And then we are going to run standard regression analysis on this stimulated database and show that in certain situations regression are unable to detect discrimination.
And finally, and this is the part I am afraid I have to skip but it will be in the slide, is we're going to introduce a formal summary or combination test that does detect the overall bias pattern or alternatively I should have said, detect lack of a bias pattern in other cases.

So I want to give fair credit here to Joseph whose 2011 paper gave me the idea of the parameters of the simulation itself.
We have a discriminating employee against females lawyers incorporated. I don't want to go through the details here, but we know that the database is simulated in such that in reality, this employee pays a lower starting salary to females and pays less to females that return in tern years, that is lower than that of males and that is the form of bias. We are going to assume they are similarly situated and the same job title. We will independently generate a thousand groups, otherwise PA G, up to 30 employees each.
The magic number is 30. And let's see what happens after we run regression on these thousand groups of 30 employees.
The chart on the left shows the standard deviation results of the thousand regressions. That's what most of us are familiar with. Less than negative two, it indicates bias against females. S-T greater than two would demonstrate significantly bias against males. These summary tests, most of them are in terms of one sided probability value. I will not go in to detail on what those mean, but let's just say the chart on the right is equivalent to the chart on the left. When the probability is left than 0.25, this indicates significantly bias against females.
These are the results from this simulation. What we notice is that even though this is employer is discriminating in all thousand PA Gs because that's how the simulation was designed, typical regression is only able to pick up 30.6 percent of the bias against females when it is occurring in 100 percent.
In addition, I want you to notice on the right chart, that felon statistically significant pay disparity has been picked up in 630 additional groups and a pay disparity non significance against males has been picked up in 64 groups. So what is going on here? We know the employer is discriminating and regression is only detecting 31 percent.
So you have heard about some of these issues already. The group sizes are small at 30. There is five control variables in these regressions. I didn't go in to those but you will have to trust me, there are five control variables so there are only six observations per variable which is fairly small.
The size of disparity is 2.3 percent. There is a gender imbalance, males is making up 80 percent. So all of these leads to an analysis of low power. The flip side is we receive a high rate of false negatives. Some of you know what type two error, so we have a 69.1 percent of rate of type two error. You could interpret the 30.6 percent on the left as the power of the regression analysis, the rate of detecting the true behavior.

What we see on the right is let's change the simulation so that we increase the PA G to one hundred and that's all we're changing. In that case we end up with the distribution of results as we do on the right, 87 percent of the regression detect the bias against female. So the power of the regression has gone up from 31 percent to 87 percent just because the sample size is greater, or the PA G side.

So the solution to this, and given our time I am not actually going to show you how the solution works, thank you for your patience. You don't see this but I'm controlling all the slides and also speaking.
We will introduce a Omni bust test. We know this contractor is biassed against females. We can only detect this in 31 percent of the group in our typical group analysis. But we also see females are disfavored in nearly 97 percent of the group, whether it is statistically significant or not.
What is important to understand is that the pay system -- in an equitable pay system, we would expect around 50 percent of groups to favor males and 50 percent to favor females on average.
What the chart on the right shows is an assimilation again for the only thing that has changed is that I have taken out the bias. That's the only thing that has changed and we run the regression and we get the P values that are on the right instead of the left.
What is on the right generally in an equitable situation what is called a uniform distribution of P value. This means a straight line of columns on the right. It is not there, but you can see it is very much approaching there.

What an Omni bust test will do is compare these two charts and tell us if the skewed outcome on the left chart is significantly different than the right chart. If it is, it may mean that chart on the left is highly unlikely that it comes from an equitable situation.
Give me two minutes of your time and then I will stop.

So this is one of the simple tests. Quick disclaimer. This particular test and any of these tests gives consideration to these methods, we're experimenting at the moment. You shouldn't interpret this as any official policy. If it does become official policy eventually, it will be after a long period of exploration and discussion with various folks.

So this is fisher's not exact test, this is his combined test. What we're doing here, I can't get in to details because of time, but I want to show this can be done simply in Excel. This is easy. Follow the steps here and you can do it.
If the P value is less than 0.025, then this distribution that you got from the thousand regression or the 20 regression is highly unlikely that comes from an equitable situation.
Disclaimer 2, this particular test I am showing because it is very simple but it shows specific things on the data that are often not true, so we will probably in reality not end up using this particular test, but it can still give you -- it is not a useless test to do at home. It can still give you -- it is a good start for you to assess your over all Workforce in addition to the typical PA G analysis.
There is a second test, the combined test but we won't look at that now. The bonus example, look at this at home, this is the situation of using these tests in the regression of this chart where we have 5.7 percent of the regression showing

statistically significant bias sometimes against men and sometimes against women. And these summary tests can actually assess whether or not across all of these tests is this representative of a pattern across the Workforce, or not? And in this particular example, it shows it is not representative of a pattern.
So the outcome can show you whether there's a bias being hidden by the way you're structuring the PA G definition, or whether a statistically significant individual PA G result is maybe just a one off.

The main takeaways from perhaps consider expanding a myopic analysis of individual pay analysis group to prevent comprehensive assessment. There are more beyond the two I showed as examples to do this. There is a short list of them on a previous slide that I would invite you to take a closer look at. Now we're done. Thank you for your patience. We will go to questions with the time we have left.

>> Thank you David, I appreciate it. We have a ton of questions and are trying to process them. I will tee up some questions to your simulation first. So stick with me. Kind of the vein of the questions are does O-S cc P do the utility calculations you presented on, and if so what do they do with them? Do we, meaning the contractor community, need to be using those methods?

>> If you want the read all the questions then I will speak to them.

>> In the MAUT example not giving any prior experience prior to coming to the company, education, that type of thing?
That's it for you for right now. Do you want to speak to those questions in the utility example?

>> Sure. First of all, in terms of whether they are calculating utilities, like I stated in the beginning and end, this is not something they're doing. Again, this is just an idea I had and I am just putting it out there because I think contractors might find it useful in doing their own analyses.
Second consideration, we're limited in these presentations to how specific these examples can get. I provided a very simplistic example. Obviously when you're doing this, you need to decide what factors to include and then how you want to weight them. I didn't do it in this analysis because A, we didn't have it for this contractor, and B we didn't really have time to go through a more complicated example.

>> All right. I will skip around a little bit. David, I will pick on you for a second. There were a couple of questions that came in at the end for you about the factors and the control variables you may have used in your simulation. I think a better question related to your simulation was related to can O FCC P really state that the employer is discriminating based on statistics alone? Isn't more information needed? Are you able to speak to that?

>> Well, I am not going to speak directly to the current law and regulations because I am not an attorney. But I will say that even the statistician very strongly urged investigators to make sure they come to the table with anecdotal information. Specifically because it helps us construct our statistical models, first of all. The

models are based on what we know, knowledge, based on well accepted theory, not just made up information. So we're not happy when we don't have the anecdotal information brought up, is that good?

>> Yes, this is Bob. I will just under score that David's example of simulation. The point he is making is that a disparity was baked in. He generated that in to the outcomes. Not in the context of an investigation where we're trying to identify discrimination, we can -- you can use whatever term you want, disparity or discrimination there.
Also, should we be using these methods? Your only obligation is to review your pay and engage in a self audit. Anything we have expressed here today are just ideas. This is a brainstorming process, if you will. We are laying out ideas and areas for you, it is very difficult to come up with an example that speaks to your particular company. That includes which variables to use. You will have to make a determination about which methods are best for you, which risk tolerance you want to assume when you apply those methods, which attributes, which characteristics and all of those are contractor specific.

>> I think the final question that I would have in the whole chat box feature everybody sent was what is the message from O FCC P? And it sounds like you just hit the nail on the head, Bob. The answer, you know, that was a great answer, you gave ideas for folks to consider as they're doing their self assessment.

>> Yes. We didn't highlight them all. I think David's slides mentions nine different techniques for global test. We know there are many other metric tests out there. And there was even the suggestion of compound or multiple parametric examples, tests. So there are opportunities or possibilities that we just want to keep the conversation going about many of those.

>> I know we have a couple other questions, but to be mindful of time we will take those questions and anything unanswered today, I will send them to Bob via e-mail. Thank you so much to the branch of expert services for an informative Webinar. Don't forget next year, 2021 we will be in Nashville in early August at the hotel. Hopefully things are better during that time so hopefully we see you next year in Nashville. And on the next slide, one final take away is again thank you to all the sponsors for this session. If you need credits, there is another slide on this presentation that will kind of give you access to that information or you'll get it after the fact. Thank you again to our sponsors. Thank you to everybody that joined us today. We appreciate it.