American Bar Association EEO Committee
Mid-Winter Meeting
New Orleans, LA

The Numbers Game: Demystifying the Use of Data in Class Actions
March 30, 2017

Christine E. Webber
Cohen Milstein Sellers & Toll PLLC
1100 New York Ave., NW Suite 500
Washington, DC 20005
(202) 408-4600
cwebber@cohenmilstein.com

I. OBTAINING DATA TO ANALYZE

A. Data Sources

1. Human Resources databases are the most commonly sought in EEO litigation. These will typically include, in addition to names, addresses, and other contact information, demographic information about each employee (i.e. date of birth, gender, race), perhaps some educational background, and typically their job history with the employer. Job history will generally show not only the dates during which the employee held different titles, but whether a move was lateral or a promotion. Such databases also generally include salary history, and may include performance evaluation scores.

2. Job posting and bidding data are frequently stored in separate databases often referred to as Applicant Tracking Systems ("ATS"). Such databases include job vacancies that have been posted (including the dates during which they were posted) and identify which employees applied to be considered for each vacancy. Some ATS services use algorithms to automate how resumes are processed and prioritized. As a result, the primary evidence of a discriminatory hiring scheme might be the algorithms themselves, a separate topic for discovery.

B. Selecting Subset to Extract

There are three major ways in which requests for production from databases are commonly limited, as well as an infinite number of variations. The three major areas of limitation to consider are: limiting the fields of data to be produced, limiting the records produced by individuals/groups of individuals, and limiting the records by temporal criteria.

1. <u>Selecting Fields</u>:

Often HR databases will contain information that you do not need, for example, information about emergency contacts or the number of dependents covered by the employer's insurance. You can approach this from two directions simultaneously.

a. First, identify all the categories of information that you hope that one or more databases includes and that you would like to have. For example: job history (positions held, date and reason for change in position); job performance; education/training/certifications; demographics (race/gender/age); name and contact information; applications for positions; compensation history.

b. Second, ask for production of a data dictionary or equivalent information in which defendant identifies the specific fields included in its database and what sort of information is stored in each field. The data dictionary will serve many purposes once you receive the data as well, but getting information on what fields of information the employer tracks is valuable, and may give you ideas about data you can request that you would not have had on your original wish list.

2. <u>Limiting Individuals Included</u>

For example, you might limit your request to data on individuals who held specified positions, or who worked in a particular geographic area or organizational unit. *See, e.g.*, *Johnson v. Kraft Foods N. Am., Inc.*, 238 F.R.D. 648, 657 (D. Kan. 2006) (imposing such limitations when ordering discovery).

3. <u>Limiting Time Period Covered</u>

While it is common for discovery to be limited to the time period encompassed by the claim, with perhaps some time immediately preceding the claim, beware of how time limits are implemented. For example, while you may be content to have personnel data regarding employees who worked for an employer in a particular position from 2010 to the present, for any such persons, you will likely need information about those individuals' experience with the employer *prior* to 2010. In a promotion case, for example, there may be a big difference between an employee hired on December 2, 2009 and one hired on March 3, 2005 when it comes to competing for a promotion in May 2010. Thus, be careful in how date limitations are framed.

C. <u>Information About the Database</u>

Crucial to your ability to work with the database is obtaining discovery of how the database is organized, what types of data are stored in which fields, what codes are used to record the data, whether a field is set up specifically to record dates in a particular format, etc, etc. A good description of databases and the information you will need to obtain about them is included in *NAACP v. Acusport Corp.*, 210 F.R.D. 268, 278-282 (E.D.N.Y. 2002). In particular, there is often a "data dictionary" which summarizes all of this information, and if available, is a key document to request. *NAACP*, 210 F.R.D. at 280-82. You are entitled to this discovery:

1. *Fautek v. Montgomery Ward & Co., Inc.*, 96 F.R.D. 141, 144-45 (N.D. Ill. 1982) (granting sanctions for, among other conduct, failing to provide complete information responsive to plaintiffs' requests for codes necessary to understand the HR database).

2. *In re Seroquel Prods. Liab. Litig.*, 244 F.R.D. 650, 660-61 (M.D. Fla. 2007) (finding sanctionable party's failure to produce IT employees for informal discovery of how to understand various databases, as agreed to in the parties case management order).

3. *Zamora v. D'Arrigo Bros. Co. of Cal.*, No. C04-00047 JW (HRL), 2007 WL 806518, at *3 (N.D. Cal. Mar. 15, 2007) (granting motion to compel where 30(b)(6) deponent did not have complete information about the database codes that were the subject of the deposition).

The question then arises how best to obtain information on database objects, fields, processes, and architecture. Certainly, requests for production are valuable for obtaining existing documents setting forth or explaining the structure of the database in general and the definitions of fields, objects, and processes. A 30(b)(6) deposition can also be extremely useful to obtain information on employer databases, but both of these formal discovery devices have limitations in the context of database discovery. In our experience, the best "tool" for obtaining the information you need to frame your requests for production of information from databases is informal, cooperative ongoing conversations with opposing counsel, IT personnel, and, as needed, experts. If your opposing counsel does not want to cooperate with you, then we recommend sending a letter with case cites and other authority on the need for cooperation.[1] If this fails, then a conference with the judge or a motion to compel may be in order. If you do pursue a deposition on this topic, seek to ensure that the witness will have access to the database during the deposition, which will often be the most effective way to refresh the witness' memory on important information.

---

[1] The Third Sedona Principle states: "Parties should confer early in discovery regarding the . . . production of electronic data and documents when these matters are at issue in the litigation, and seek to agree on the scope of each party's rights and responsibilities." The Sedona Conference, *The (2004) Sedona Principles: Best Practices, Recommendations, & Principles for Addressing Electronic Document Production*, 5 Sedona Conf. J. 151, 162 (2004). *See also* Fed. R. Civ. P. 26(f) advisory committee's note to 2006 amendment ("Rule 26(f) is amended to direct the parties to discuss discovery of electronically stored information during their discovery-planning conference."); *In re Seroquel Prods. Liab. Litig.*, 244 F.R.D. at 658 (finding that a party's "refusal to allow contact between individuals with appropriate technical backgrounds as part of the effort to resolve technical issues is an inexplicable departure from the requirements of Rule 26, the Sedona Principles and this Court's expressed expectations"); *id.* at 664 (finding that Rules 26 and 34 require "dialogue to discuss the search terms").

D. Proportionality in Discovery of Databases

The court in *Chen-Oster v. Goldman, Sachs & Co.*, 285 F.R.D. 294 (S.D.N.Y. 2012), addressed the proportionality argument with a very sophisticated understanding of databases. In that putative class case, the defendant claimed that extracting the requested information from its HR database would require 90 to 150 hours of staff time, plus another 40 to 80 for quality control checks.[2] Defendant argued this was too great a burden to ask them to bear in proportion to the needs of the case.

The court first considered two different approaches for reducing the cost to defendant: (a) extracting a sample from the database, which both parties opposed for different reasons, and which the court conceded would not reduce the time required to program the query to extract the data, though it might decrease the time spent on quality control reviews (*id.* at 304); (b) a data dump – simply copying the database, instead of running a query to extract only certain information, which plaintiffs proposed but the court rejected because whatever savings in time was made on the front end would be lost in the end requiring defendant to provide all of the information needed to work with the data (*id*. at 305).

Having concluded there was not a significantly cheaper, equally effective alternative, the court turned to weighing the need for the data and found the burden imposed on defendant was proportionate to the needs of the case. Specifically, the court held "[t]here is little doubt that the needs of this case justify the discovery sought by the plaintiffs. The information in the databases is central to the plaintiffs' claims of gender discrimination in compensation, promotion, and evaluation. The amount in controversy, while not specifically quantified, is surely substantial." The court also found the defendant had ample resources to provide the discovery. *Id.* at 305. Moreover, the court recognized that "the importance of this litigation is not measured in dollars alone; the plaintiffs seek to vindicate the civil rights of the class members, and thus further an important public interest." *Id.* at 306.

Finally, the court understood that Goldman Sachs had somewhat inflated the burden by building in so much quality control, noting that their estimate "which is rather conclusory, appears to be based on a goal of providing a pristine set of data." *Id.* The court pointed to the Sedona Conference to support its conclusion that "the standard for the production of ESI is not perfection. Rather, '[a] responding party must use *reasonable* measures to validate ESI collected from database systems to ensure completeness and accuracy of the data acquisition.' The Sedona Conference, *The Sedona Conference Database Principles: Addressing the Preservation and Production of Databases and Database Information in Civil Litigation*, March 2011 Public Comment Version, at 32 (emphasis added)." The court held that defendant could sample the data extracted to identify any systematic errors rather than conduct the comprehensive quality review defendant proposed. *Id.*

As *Chen-Oster* demonstrates, it is important to press for detail underlying defendant's estimates, so that the "padding" in the estimate can be identified. It is also useful to find out how often defendant engages in similar searches and extraction of data for its own internal purposes, to show how routine such work is. Finally, on the benefit side, it is important to ensure that the

[2] Other databases were also discussed, each with varying numbers of hours estimated.

court considers the non-economic value of civil rights litigation, and does not merely consider the lost wages or other damages sought.

The other end of the scale, an example of a case finding a lack of proportionality with a relatively cursory discussion is *EEOC v. Supervalu, Inc.*, No. 09 CV 5637, 2010 WL 5071196, at *8 (N.D. Ill. Dec. 7, 2010). In that case, the EEOC sought data showing employee histories that they hoped to use to identify instances when there were vacancies that could have been filled, a request defendant estimated would take a week of staff time to comply with. However, the court found that "as plaintiff appears to acknowledge by virtue of its 'piecing together' argument, the information it seeks would not definitively prove the existence or number of open or filled positions at defendants' retail stores at any particular time. Instead, that information would require significant analysis as well as the inference that each time one type of employee position ended—whether by termination, resignation, or otherwise—a corresponding position became open and ready for hire." *Id.* Given the lack of evidence supporting that inference, and contrary evidence from defendant, the court refused to order defendant to produce the requested data. The court also noted that discovery closed in one week, and found that the EEOC's explanation for why it was seeking the database had shifted over the course of the litigation, making it less inclined to grant the request.

II. LEGAL FRAMEWORK FOR STATISTICAL EVIDENCE

The Supreme Court has long recognized the utility of statistical evidence in establishing the existence of a pattern or practice of discrimination, specifically because intent can be inferred from statistical evidence. See, e.g., *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 335 n.15, 339‐40 n.20 (1977) ("Statistics showing racial or ethnic imbalance are probative . . . because such imbalance is often a telltale sign of purposeful discrimination."); *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307-8 (1977); *Kilgo v. Bowman Transp., Inc.*, 789 F.2d 859, 874 (11th Cir. 1986); *Griffin v. Carlin*, 755 F.2d 1516, 1525 (11th Cir. 1985); Ramona L. Paetzold & Steven L. Willborn, *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases* 69-72 (West 2012-2013 ed.). In *Dukes*, the Supreme Court cited *Teamsters* as the standard for establishing a pattern or practice of discrimination on the merits. *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2552, 2556, 2565 n.7 (2011). The Court specifically referred to the "substantial statistical evidence of company-wide discrimination" present in *Teamsters*. *Id.* at 2556.

Courts have agreed that, "Statistics alone can make out a prima facie case of discrimination if the statistics reveal 'a gross disparity in [the] treatment of workers based on race.'" *Robinson v. Metro-North Commuter R.R.*, 267 F.3d 147, 158-59 (2d Cir. 2001), *quoting Lopez v. Laborers Int'l Union Local No. 18,* 987 F.2d 1210, 1214 (5th Cir. 1993); *see also Ardrey v. United Parcel Serv.,* 798 F.2d 679, 684 (4th Cir. 1986) ("Since strong statistical evidence, without anecdotal evidence, may in some cases form a prima facie case, a defendant's successful rebuttal of each alleged instance of discrimination weakens, but does not defeat, a plaintiff's class claim. Neither statistical nor anecdotal evidence is automatically entitled to reverence to the exclusion of the other.").

While anecdotal evidence is also important, the burden of establishing that discrimination was defendant's standard operating procedure" rests heavily on statistical evidence. *Teamsters*, 431 U.S. at 336. As *Teamsters* explained, the employer then generally seeks to show that such statistical evidence is "inaccurate or insignificant." *Id.*, at 360. The "employer's defense must, of course, be designed to meet the prima facie case … . The point is that at the liability stage of a pattern-or-practice trial the focus often will not be on individual hiring decisions, but on a pattern of discriminatory decisionmaking." *Id.*, at 360 & n.46.

In disparate impact claims, plaintiffs are not required to prove that an employer intended to discriminate, thus statistical evidence does not need to serve that function. Instead, plaintiffs must show that the challenged, facially neutral practice had an adverse impact on the protected class. *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

III. COMMON TYPES OF STATISTICAL ANALYSES

In showing the disparate impact of a test or similar selection criterion, a simple t-test may be all that is required.

However, for disparate treatment claims, forms of analysis which permit consideration of many factors which may legitimately impact the decision are preferred. Multiple regression analysis is the most common method for analyzing claims of pay discrimination. Probit or logit regression methods permit similar analysis of promotion, termination or other "yes/no" events. Pools analysis, such as Mantel-Haenszel, is also used with promotions.

IV. COMMON DISPUTES

A. Statistical Significance

Generally, two standard deviations is the accepted level of statistical significance for statistical evidence in employment discrimination cases. *See, e.g., Castaneda v. Partida*, 430 U.S. 482, 496-97 n.17 (1977); *Hazelwood Sch. Dist.*, 433 U.S. at 309 n.14; *Kilgo v. Bowman Transp., Inc.*, 789 F.2d 859, 874 (11th Cir. 1986); *Page v. United States Indus., Inc.*, 726 F.2d 1038, 1047 (5th Cir. 1984); *Segar v. Smith*, 738 F.2d 1249, 1283 (D.C. Cir. 1984).

However, courts have also accepted evidence that is less than two standard deviations in some circumstances. *Chin v. Port Auth. of N.Y. & N.J.*, 685 F.3d 135, 153 (2d Cir. 2012); *Kadas v. MCI Systemhouse Corp.*, 255 F.3d 359, 362 (7th Cir. 2001); *Waisome v. Port Auth. of N.Y.& N.J.*, 948 F.2d 1370, 1378-79 (2d Cir. 1991); *Palmer v. Shultz*, 815 F.2d 84, 96, 97 n.10 (D.C. Cir. 1987) ("[p]laintiffs are in no way foreclosed from establishing an inference of discrimination simply because the contested disparity falls short of the 1.96 standard deviations mark when analyzed statistically.").

B. Omission of Tainted Variables

In *Bazemore*, the Supreme Court held that a multiple regression analysis does not have to include "all measurable variables," however, the analysis model must account for the "major factors." *Bazemore v. Friday*, 478 U.S. 385, 400 (1986). There will likely be disputes over what the "major factors" are. Ensure that your experts have access to applicable written policies and practices, and relevant testimony from Rule 30(b)(6) witnesses and other managers, so that they can craft regression models in light of the variables that decision-makers actually consider when making employment decisions, models that can actually be defended as encompassing the "major factors." Be aware that what decision-makers say they consider and what data show is a factor do not always match.

While there will likely be disputes over whether a factor is a "major factor" which should have been included, case authority supports the exclusion of "tainted" variables. Tainted variables are ones which are either a product of prior discrimination or a pretext for current discrimination. For example, in a case complaining of discrimination in compensation, starting salary might be a relevant variable to consider when evaluating whether current disparities in compensation are due to discrimination in starting pay, discrimination in pay increases, or neutral factors. However, if starting salary was itself assigned in a discriminatory manner, then it cannot serve as a "neutral" variable – it is tainted. *See e.g. Greenspan v. Auto. Club of Mich.*, 495 F. Supp. 1021, 1061-64 (E.D. Mich. 1980).

Likewise, in *Chen-Oster*, the court found that several variables were tainted because they were likely affected by gender bias. *Chen-Oster v. Goldman, Sachs & Co.*, 114 F. Supp. 3d 110, 116-121 (S.D.N.Y. 2015). There, plaintiffs' expert created a regression analysis based on a number of factors, including the division in which the employee worked, geographic office, education, prior related work experience, experience with the defendant company, and lateral or direct hiring status. *Id*. at 116. Based on this methodology, the expert found that female associates earned 7.6% less than male associates after accounting for these factors. *Id*. Defendant argued that the expert's model was flawed because it failed to take into account two specific evaluation processes used by the company. *Id*. at 117-18. The court disagreed, finding that the expert "reasonably suspected that these variables were tainted" because "they themselves [we]re affected by gender bias." *Id*. at 118. Specifically, the expert "demonstrated that women were graded lower in both [evaluations] even after controlling for measurable and consistently available predictors of performance. Thus, there was a legitimate basis for him to omit these performance measures from his analysis on the ground that they were themselves subject to bias and would therefore mask discrimination." *Id*.

Absent such evidence a variable is tainted, however, the failure to take neutral variables into account can be fatal to an expert report. For example, plaintiffs' expert in *Summy-Long* compared the mean salary between black and white employees in an attempt to provide evidence of intentional discrimination, but paid no regard to the level of skill, education and training of these employees. *Summy–Long v. Pennsylvania State Univ*., No. 1:06-CV-01117, 2016 WL 7448710, at *17 (M.D. Pa. Dec. 27, 2016). The court found that because the expert "fail[ed] to take into account the fact that a number of factors operate simultaneously to influence the amount of salary an employee receives," *id*. (internal quotation marks omitted), the expert report, without more, was insufficient to make out a prima facie case.

One variable which has often been challenged as tainted is grade level or similar measures of position. While grade level may be an important factor in explaining compensation, in cases where plaintiffs have also alleged discrimination in promotions, the grade level itself may be tainted by discrimination. *See Coward v. ADT Sec. Sys.*, 140 F.3d 271, 274 (D.C. Cir. 1998) ("Where plaintiffs allege discriminatory promotion practices, for example, this court considers inclusion of grade variables [in salary regression model] 'inappropriate' because an employee's grade may itself reflect discrimination."); *Valentino v. USPS*, 674 F.2d 56, 70-71, n.30 (D.C. Cir. 1982) ("Grade level, in a case such as this one, is an 'inappropriate variable.' Absent clear, affirmative evidence that promotions were made in accordance with neutral, objective standards consistently applied, there is no assurance that level or rank is an appropriate variable, untainted by discrimination." (citation omitted)); *James v. Stockham Valves and Fittings Co.*, 559 F.2d 310, 332 (5th Cir. 1977) (Wisdom, J.) (rejecting defendant's regression analysis because it included measures of "skill level," performance rating, factors that were "defined in such a way as to incorporate discrimination. ... The systematic exclusion of blacks from promotion and training opportunities for such [high grade level] jobs, as is alleged here, will automatically produce no black employees with 'skill level.' .... If there is racial bias in the subjective evaluation of white supervisors, then that bias will be injected into [defendant's] earnings analysis"); *Greenspan*, 495 F. Supp. at 1061-64 (approving use of "productivity-related" variables and omission of all "company variables" (job assignment, grade, salary, promotions, disciplinary actions) as potentially tainted). While the *Valentino* court placed the burden on defendant to establish that the variable was *not* tainted, other courts have placed the burden on plaintiffs to support exclusion of a variable with evidence that the variable is tainted by discrimination. *Coates v. Johnson & Johnson*, 756 F.2d 524, 544 (7th Cir. 1985).

C. Defendants Cannot Rely on Theoretical Critiques

Employers cannot simply provide theoretical claims that the statistical results might be different if a particular factor were accounted for, or some other methodological change were implemented. For example, if defendant criticizes plaintiffs' analysis for omitting a variable, or using a particular method of calculation, then defendant must come forward with evidence that doing the calculation differently would change the outcome of the analysis in a way favorable to defendant.

Bazemore, 478 U.S. at 399-400, 403 n.14 (rejecting defendant's argument that plaintiffs' analysis was unsound because certain factors were omitted, finding that defendant "made no attempt . . . to demonstrate that when these factors were properly organized and accounted for there was no significant disparity between the salaries of blacks and whites"); *EEOC v. Gen. Tel. Co. of NW, Inc.*, 885 F.2d 575, 581-83 (9th Cir. 1989) ("[T]he defendant cannot rebut an inference of discrimination by merely pointing to flaws in the plaintiff's statistics," rather, the defendant "had to produce credible evidence that curing the alleged flaws would also cure the statistical disparity."); *Capaci v. Katz & Besthoff, Inc.*, 711 F.2d 647, 653-654 (5th Cir. 1983) ("The defendant must do more than raise theoretical objections to the data or statistical approach taken; instead, the defendant should demonstrate how the errors affect the results."); *Segar v. Smith*, 738 F.2d 1249, 1287 (D.C. Cir. 1984); *Catlett v. Mo. Highway & Transp. Comm'n.*, 828 F.2d 1260 (8th Cir. 1987).

D. Level of Aggregation

A perennial dispute in the use of statistical analyses in employment class cases is whether aggregated or disaggregated analyses are more appropriate. Specifically, the parties often dispute what level of aggregation is appropriate.

Numerous cases pre-*Dukes* rejected disaggregated analyses both because broader groups were subject to the same challenged practice and because dividing the data into small sample sizes made it likely that a real difference in treatment would not appear statistically significant. *Paige v. California*, 291 F.3d 1141, 1148 (9th Cir. 2002), *as amended by*, 2002 U.S. App. LEXIS 14463 (9th Cir. 2002); *Eldredge v. Carpenters 46 N. Cal. Ctys. Joint Apprenticeship & Training Comm.*, 833 F.2d 1334, 1340 n.8 (9th Cir. 1987) ("In general, 'the plaintiff should not be required to disaggregate the data into subgroups which are smaller than the groups which may be presumed to have been similarly situated and affected by common policies.' Id. § 7.1 (1986 Supp.).") (citation omitted); *Capaci*, 711 F.2d at 654; *Segar,* 738 F.2d at 1286 (disapproving statistical analysis that "repeatedly disaggregat[ed] until groups were too small to generate any statistically significant evidence of discrimination"); *McReynolds v. Sodexho Marriott Servs., Inc.*, 349 F. Supp. 2d 1, 15 (D.D.C. 2004) (recognizing that disaggregation of data made it more difficult to detect statistical significance). Thus, plaintiffs commonly completed class-wide or company-wide analyses, in which various locations or units might be controlled for as factors in the analysis, but separate regressions were not run for each location or unit.

In *Dukes*, the Supreme Court expressed the view that national or regional statistics were not appropriate vehicles for examining decisions made at the store level, where outcomes might vary by store. 131 S. Ct. at 2555 ("A regional pay disparity, for example, may be attributable to only a small set of Wal-Mart stores."). The Court went on to express concern about comparing results at an individual store to "nationwide figures or the regional figures" because of the possibility that "the availability of women, or qualified women, or interested women, in the[] stores' area does not mirror the national or regional statistics." *Id.* The focus of the Court's concern was not that plaintiffs had not appropriately considered the availability of qualified, interested women in the variables included in the analysis, but that doing so at the national or regional level could mis-state the availability in the relevant geographic area. *Id.*

There are now commonly three options pursued to address this concern.

1. Establish central decisionmaking. If there is a common decisionmaker, then one aggregated analysis is appropriate.

2. Continue to use a single regression analysis, but incorporate interaction terms, i.e. controlling not only for store but also for the interaction between gender and store. In this fashion, a small set of stores cannot cause an overall disparity in the way the Court was concerned about.

3. Run many separate regression analyses.

As an example of the first option, in *Chen-Oster*, the court found that because the allegedly discriminatory evaluation process was used in each division and every business unit at the defendant company, "it [was] appropriate to examine these policies across the entire population, even if the effects of those policies may vary in different business units, and even if those effects may not be statistically significant in many individual units." *Chen-Oster*, 114 F. Supp. 3d at 120. The court went on to note the problems with disaggregating data to the point that it can no longer provide statistically significant results: "The difficulty with analyzing data on the business unit level, as [defendant] advocates, is that such disaggregation tends to mask common mechanisms because the sample size in each unit is so small." *Id.*

Aggregation across periods of time and across positions is still commonly permitted. In *Moore v. Napolitano*, 925 F. Supp. 2d 8 (D.D.C. 2013), the court found that plaintiffs' expert properly aggregated data across the years at issue, as opposed to disaggregating year-by-year, because the allegedly discriminatory promotion process "did not substantially change over the class period" and because "disaggregated annual data may have only yielded small numbers … not powerful enough to detect a disparity." *Id.*, at 24-25 (internal quotation marks omitted). *But see EEOC v. Mavis Disc. Tire, Inc.*, 129 F. Supp. 3d 90, 99-100, 115 (S.D.N.Y. 2015) (correctness of aggregating over time is a question for the jury, where plaintiffs' expert looked at hiring across 5 years combined, while defendant's expert argued that "hiring decisions from one year can affect the demand for employees the next year," making such aggregation inappropriate).

The court in *Moore* similarly found that aggregation across vacant positions was reasonable because disaggregation "may mask whether the overall decision-making process produces a discriminatory result, whereas analyzing an entire group will indicate whether the identified employment practice was the cause of the disparity." *Moore*, 926 F. Supp. 2d at 25 (internal quotation marks omitted).

Aggregation is also available when discriminatory employment decisions in one department affect the employment opportunities in other departments. For example, *Brown v. Nucor Corp.*, 785 F.3d 895 (4th Cir. 2015), concerned a single plant where workers "shared common spaces, were in regular physical contact with other departments, could apply for promotions in other departments, and were subject to hostile plant-wide policies and practices." *Id.*, at 910. Thus, even though the statistical evidence of discrimination disproportionately involved one department at the plant, evidence existed that the departments were not "autonomous operations," and "racial bias in one … department itself diminished the promotional opportunities … in all the departments." *Id*. at 911.

E. <u>How to Analyze Disaggregated Results</u>

When many separate analyses are run, for sub-units of the class (i.e. by geographic area, division, store), there is nonetheless the need to evaluate what overall conclusion can be drawn from the data. The professional literature gives explicit direction that when a data set is broken down, or "stratified," so that a large number of separate analyses are completed, that a summary statistic for the overall data set must be calculated, in addition to presenting the results of the individual analyses. Ramona L. Paetzold and Steven L. Willborn, *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases* 169-71 (West, 2012-2013 ed.); Joseph L. Gastwirth et al., *Some Important Statistical Issues Courts Should Consider in Their Assessment of Statistical Analyses Submitted in Class Certification Motions: Implications for Dukes v. Wal-Mart*, 10 Law, Probability & Risk 225, 228, 234-35 (2011); Michael O. Finklestein & Bruce Levin, *Statistics for Lawyers* 241-249 (Springer, 1990) ("having disaggregated the data to reduce bias and increase validity, we then seek a statistic that sums up the situation in an appropriate way.").

Common options are:

1. Look for a pattern of non-significant but adverse results. See, e.g., *Ellis v. Costco Wholesale Corp.*, 285 F.R.D. 492, 523-24 (N.D. Cal. 2012).

2. Test distribution of results against expected distribution. See *Gastwirth, supra*, at 228, 234-35.

3. Test whether results form a bell curve or other distribution. See Statisticians and Employment Analysts' Amicus Brief in Support of Plaintiffs' 23(f) Petition, *Dukes v. Wal-Mart Stores, Inc.*, No. 13-90184, dkt. 2 (9th Cir. Aug. 16, 2013).

4. Complete a companywide analysis in addition to the sub-unit analyses. See *Paetzold and Willborn, supra*, at 169-71.

5. Majority rule, counting only individually statistically significant results (followed in Dukes on remand, but neither case law nor statisticians support, *Dukes v. Wal-Mart Stores, Inc.*, No. CV 01-02252 CRB, 2013 U.S. Dist. LEXIS 109106, at *14 (N.D. Cal. Aug. 2, 2013)).

6. Unanimous rule, counting the evidence only if every single sub-unit is individually statistically significant.

Many courts have considered whether there appears to be a pattern in the multiple results, considering results that are not individually statistically significant (the first option noted above). *McReynolds*, 349 F. Supp. 2d at 14-16 (Plaintiffs' expert found commonality where "73.7% (84 out of 114) of the [sub-units] show[ed] a disparity (although not a statistically significant one)"); *Anderson v. Boeing Co.*, 222 F.R.D. 521, 532-34 (N.D. Okla. 2004) (finding commonality based on analysis that showed non-statistically significant disparities in some subpopulations and, when disaggregated to certain job groupings, showed variation including compensation favoring women in some subunits). In *Ellis*, the court addressed the disaggregated data, and found a pattern of non-significant results "telling":

> Even if the data were examined strictly on a region by region basis, the fact that *all seven non-Texas regions* show a raw gender disparity in promotions is telling. Examining the data by region, the data supports Plaintiffs' contention that gender disparities extend across all regions, and the absence of statistical significance within each individual region is of limited value for the reasons discussed above.

Ellis, 285 F.R.D. at 523-24 (record citations omitted).

However, others have advocated for more formal statistical analyses of whether a pattern exists, whether it be simply conducting an aggregated analysis (option 4), or through calculations based on the disaggregated results such as in the Gastwirth article and Lieder brief cited above as options 2 and 3.

There is less support in either case law or academic literature for the other options – requiring all sub-units or a majority of sub-units to be independently statistically significant to support a finding of discrimination.

Professor Gastwirth criticizes analyses which consider only significant strata, especially when some subunits could not generate a statistically significant result, given the size of the sample and other relevant variables. *Gastwirth, supra*, at 234. He notes where the vast majority of strata showed the protected group being disadvantaged, the data should be combined and a summary estimate obtained, concluding "from a statistical viewpoint, the data show a statistical and meaningful difference in the hiring rates of older and younger employees." *Id*. at 235.

Employers often wrongly assert that only statistically significant store-level results are probative. Instead, when there are numerous subgroups that are analyzed "the configuration of the individual groups in stratified data may not allow statistical significance to be found in any single group; however, the data as a whole provide very convincing evidence of a system-wide disparity." *Id.* at 258. Indeed, Gastwirth does not only find a summary statistic to be permissible, he urges that it be required. *Id.* at 259 (recommending that "when parties submit data stratified into subgroups, courts require that a summary statistic be computed and interpreted.")

F. <u>Varying Magnitude of Disparity in Disaggregated Results</u>

Another issue raised by many defendants when there have been multiple separate analyses, aside from whether the results are statistically significant, as referenced above, is whether the magnitude of the disparity in the gender coefficient is the same across sub-units. Does it matter that in some divisions women are underpaid by 5% and in other divisions they are underpaid by 10%? Courts have held that the size of the disparity need not be the same for each sub-group. This argument was explicitlye rejected in *Velez v. Novartis Pharm. Corp*., 244 F.R.D. 243 (S.D.N.Y. 2007):

> To show commonality, however, it is not necessary to find the *same common difference* in each group. In other words, plaintiffs need not show that they each suffer the same degree of pay disparity. The asserted common question is whether there was discrimination; the degree of damage presumably differs in most class-action discrimination cases.

Id. at 262 n.14; *see also Eldredge*, 833 F.2d at 1339-41 & n.7 (summary judgment was entered for the plaintiff class on a pattern or practice claim where disparities in subpopulations ranged from 16.09% to 147.27%, and the overall disparities were statistically significant). The claim that the gender coefficients must all be identical has been rejected for statistical, as well as legal, reasons. *Gastwirth*, *supra*, at 248 (rejecting the use of the Chow test and the requirement that the gender effect be the same in all stores before combining the data); Efstathia Bura et al, *The Use of Peters-Belson Regression in Legal Cases* 12-13 (2011) (requiring identical gender coefficients before combining data is too strict a requirement); *Chen-Oster*, 114 F. Supp. 3d at 122 ([T]he case law does not support the contention that passing the Chow test is an absolute prerequisite for including different data sets in the same regression …. There are also sound statistical reasons for being cautious about using the Chow test alone to reject a model.").

G. Availability Pools vs. Applicant Flow

Often employers may not have a record of who applied for a promotion, or there may not have been any application process. In such circumstances, availability pools – which identify candidates available for selection who are similar to the candidate actually selected – are regularly used in promotion and hiring analyses in discrimination cases. *Hazelwood Sch. Dist.*, 433 U.S. at 308-309. See also *Van v. Plant & Field Serv. Corp.*, 672 F. Supp. 1306, 1316 (C.D. Cal. 1987), *aff'd*, 872 F.2d 432 (9th Cir. 1989) (statistical comparison of workforce and general population may establish pattern or practice); *Forehand v. Fla. State Hosp.*, 89 F.3d 1562, 1574 (11th Cir. 1996) (no per se rule that applicant flow is best). Indeed, availability pools may be preferred. *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 650 (1989); *Phillips v. Cohen*, 400 F.3d 388, 399 (6th Cir. 2005) (directing comparison between those hired and those in the qualified local labor force, and only if labor force data is unavailable, using the applicant pool).

Proxy pools may also be used when the defendant cannot provide complete data. In *Texas Roadhouse*, the EEOC's expert applied statistical tests for each restaurant position for each store-year for which he had sufficient data against three different benchmarks: census data, paper applications and electronic applications. *EEOC v. Texas Roadhouse, Inc.*, No. CV 11-11732-DJC, 2016 WL 6134123, at *4 (D. Mass. Oct. 19, 2016). The expert used the census data because the actual data from the defendant was incomplete. *Id*. The defendant moved to strike the expert's testimony, arguing that the census data could not be used as a proxy for the positions at the defendant's restaurant given the breadth and variety of service jobs reflected in the census data. *Id*. at *6. The court denied the motion, finding that the subset of census data the expert used was more particularized than general population census information, and need not be a perfect match for the positions at issue. *Id*. The court also found that "failing to use a perfect set of variables that incorporates all relevant factors or excludes all potentially irrelevant variables is not a means for rejecting an expert's analysis." *Id*. Any use of overbroad census data goes to weight, not admissibility. *Id*. *See also Brown*, 785 F.3d at 903-04 (where defendant failed to retain actual bidding records, proper for expert to assume that the pool of candidates for promotions in the year of lost records had the same average racial composition as the pools for similar jobs in the years with records).

H. One-tailed vs. Two-tailed Tests

Generally, a one-tail test measures the likelihood that the discrimination against the protected class is the product of chance. A two-tail test includes the likelihood that there is "reverse discrimination," that the protected class is being favored. The two tests have different uses:

> Statisticians can employ either one or two-tailed tests in measuring significance levels. The terms one-tailed and two-tailed indicate whether the significance levels are calculated from one or two tails of a sampling distribution. Two-tailed tests are appropriate when there is a possibility of both overselection and underselection in the populations that are being compared. One-tailed tests are most appropriate when one population is consistently overselected over another. The practical difference between one and two tailed tests is that the P-value

> produced by a two-tailed test is usually twice as great as that produced by a one-tailed test.

Stender v. Lucky Stores, Inc., 803 F. Supp. 259, 323 (N.D. Cal. 1992). While *Stender* found use of the one-tailed test appropriate, other courts have required use of two-tailed tests. *See, e.g., Palmer v. Shultz*, 815 F.2d 84, 95-96 (D.C. Cir. 1987).

Recent cases continue to illustrate differing views.

In *Moore*, the court approved the expert's use of a one-tailed test instead of a two-tailed test. *Moore*, 926 F. Supp. 2d at 23 n.10 ("The terms 'one-tailed' and 'two-tailed' refer to the 'tails' or ends of the bell-shape curve, which represents in graph form a 'random normal distribution.'" (citation omitted)). The expert explained that a one-tailed test, unlike a two-tailed test, is sufficient to determine whether an observed difference is adverse or favorable to African-Americans. *Id*. at 23.

In *Smith v. City of Boston*, 144 F. Supp. 3d 177, 195-198 (D. Mass. 2015), the court collected cases on both sides of the issue, and expressed sympathy with the logic of the one-tail test, before settling on the two-tailed test as most appropriate.