2010 Decisions

Opinions of the United
States Court of Appeals
for the Third Circuit

8-16-2010

Sharyn Stagi v. Natl RR Passenger

Follow this and additional works at: https://digitalcommons.law.villanova.edu/thirdcircuit_2010

Recommended Citation

UNITED STATES COURT OF APPEALS
FOR THE THIRD CIRCUIT

No. 09-3512

SHARYN STAGI, individually and on
behalf of all others similarly situated;
WINIFRED LADD,

Appellants

v.

NATIONAL RAILROAD PASSENGER
CORPORATION, t/d/b/a AMTRAK

Appeal from the United States District Court
for the Eastern District of Pennsylvania
(D.C. Civil No. 2-03-cv-05702)
District Judge: Honorable Anita B. Brody

Argued May 28, 2010

Before: McKEE, Chief Judge,
RENDELL and STAPLETON, Circuit Judges.

(Filed : August 16, 2010)

Ari R. Karpf, Esq.
Karpf, Karpf & Virant
3070 Bristol Pike
Building 2, Suite 231
Bensalem, PA 19020

Timothy M. Kolman, Esq.
Michael F. Mirarchi, Esq.
Timothy M. Kolman & Associates
414 Hulmesville Avenue
Penndel, PA 19047

Scott M. Lempert, Esq.
Alan M. Sandals, Esq.   **[ARGUED]**
Sandals & Associates
One South Broad Street
Suite 1850
Philadelphia, PA 19107
    Counsel for Appellants

Sarah Andrews, Esq.
Morgan, Lewis & Bockius
301 Grant Street
One Oxford Centre, Suite 3200
Pittsburgh, PA 15219

James E. Bayles, Jr., Esq.
Morgan, Lewis & Bockius
77 West Wacker Drive
6th Floor
Chicago, IL 60601

William J. Delany, Esq.   **[ARGUED]**
Morgan, Lewis & Bockius
1701 Market Street
Philadelphia, PA 19103
    Counsel for Defendant

OPINION OF THE COURT

RENDELL, <u>Circuit Judge</u>.

Plaintiffs Sharyn Stagi and Winifred Ladd brought a class action against the

National Railroad Passenger Corporation ("Amtrak"), asserting that a company policy requiring all union employees to have one year of service in their current position before they could be considered for promotion has a disparate impact on female union employees in violation of Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e, and the Equal Protection component of the Due Process Clause of the Fifth Amendment. The District Court, presented with motions for class certification and for summary judgment, granted summary judgment in favor of Amtrak, finding that "the plaintiffs' evidence of disparate impact lack[ed] both statistical and practical significance," thus concluding that "the plaintiffs have failed to make out a *prima facie* case of discrimination under Title VII." *Stagi v. Nat'l R.R. Passenger Corp.*, Civ. No. 03-5702, 2009 WL 2461892, at *1 (E.D. Pa. Aug. 12, 2009) (*Stagi II*).

Although it is a close call, we will reverse and remand for further proceedings consistent with this opinion.

I.

At issue in this case is Amtrak's policy referred to as the "one-year blocking rule." Under that rule, a union member must be in her current union position for at least one year in order to be eligible for promotion into a management position. The policy states, "[a]n agreement covered employee may not apply for a posted non-agreement covered

position unless he or she has been in his or her current union for one year." App. 299.[1]

The rule has no exceptions. The rule was first promulgated on May 1, 1994 and was

revised in September 2000, which revision was in force during the time period relevant

for this case.

Plaintiffs Stagi and Ladd are long-time Amtrak employees who have been

employed in both its union and management ranks during their careers. Stagi began her

career at Amtrak in 1973 as a reservation and information clerk, and eventually worked

her way up to various union positions until the early 1990s, when she was promoted to a

management position. She was in a management position in April 2002 when she was

laid off as a result of a corporate-wide management restructuring effort. Ladd was

promoted to management in 1986 and continued to be promoted through management

until April 2002, when her job was similarly eliminated. Because they had previously

worked in Amtrak's union ranks, they were both entitled to "bump down" into a union

position based on their retained union seniority. In the year following their layoffs, both

applied for management vacancies, some of which they had previously held or

---

[1] Although the policy says "in his or her current union[,]" the parties agree that the policy has been interpreted and applied by Amtrak as blocking an employee who has not been in his or her current union *position* for at least one year. *See Stagi II*, 2009 WL 2461892, at *1 n.4 ("Although the way the rule is written appears to prevent consideration of agreement-covered employees based on time-in-current-*union*, since at least 1999 or 2000, the Policy has been applied consistently to consider time-in-*position*, not time-in-union . . . . The language of the policy [sic] was changed in 2004 (after the commencement of this litigation).").

supervised. They were both blocked by the one-year rule from being considered for those positions. Stagi remains in her union position. Ladd was not able to return to management before 2004, when she left on long-term disability and retired with benefits inferior to those she would have enjoyed had she been permitted to access a management position.

In October 2003, Stagi filed a class complaint, and later amended it to add Ladd. Plaintiffs' complaint alleges that Amtrak violated Title VII, 42 U.S.C. § 2000e *et seq.*, and the Equal Protection component of the Due Process Clause of the Fifth Amendment by adopting and applying the one-year rule to plaintiffs.

In May 2005, Amtrak moved for judgment on the pleadings under Rule 12(c) of the Federal Rules of Civil Procedure. The District Court denied Amtrak's motion holding that plaintiffs had "made out a prima facie case" of disparate impact by the blocking rule at issue here. *Stagi v. Amtrak*, 407 F. Supp. 2d 671, 676 (E.D. Pa. 2005) (*Stagi I*).

The District Court held a discovery conference on January 2, 2006, and plaintiffs moved to compel production of discovery material related to the qualifications of the various management positions as well as the work histories and other qualifications of union employees who might have been qualified for management positions (although they might be blocked by the one-year rule). The court held additional discovery conferences on April 4, 2007 and May 4, 2007. One of the issues discussed at each conference was the use and availability of qualifications data. Amtrak subsequently produced certain

employee data in July 2007. Based in part on this data, plaintiffs submitted an expert

report by Mark R. Killingsworth on October 23, 2007. Amtrak submitted a responsive

expert report by David W. Griffin on January 25, 2008.

Plaintiffs filed a motion for class certification under Rule 23 on February 29, 2008.

Before that motion was fully briefed, Amtrak moved for summary judgment on April 21,

2008. Briefing was complete for the class certification motion on June 6, 2008 and for

the summary judgment motion on November 17, 2008. A hearing was held on July 21,

2009, at which each party's expert testified. By memorandum and order dated August 12,

2009, the District Court granted Amtrak's summary judgment motion.[2] *Stagi II*, 2009 WL

2461892, at *13. Plaintiffs timely appealed.[3]

II.

A. Title VII and Disparate Impact

Under Title VII of the Civil Rights Act of 1965, it is unlawful for an employer to

"limit, segregate, or classify his employees or applicants for employment in any way

which would deprive or tend to deprive any individual of employment opportunities or

otherwise adversely affect his status as an employee, because of such individual's race,

---

[2] On appeal, plaintiffs argue that the District Court erred in ruling on the summary judgment motion when it did because the District Court had informed the parties that the July 21 hearing would be limited to questions relating to class certification. Because we will reverse the District Courts's order granting summary judgment on other grounds, we need not decide this issue.

[3] The District Court had subject matter jurisdiction under 28 U.S.C. §§ 1331 and 1343(a)(4). We have jurisdiction under 28 U.S.C. § 1291.

color, religion, sex, or national origin." 42 U.S.C. § 2000e-2(a)(2). This prohibition

against disparate impact is distinct from disparate treatment by an employer, which

requires a showing of discriminatory intent. Under Section 2000e-2(a)(2), an otherwise

facially neutral business practice that disproportionately affects or impacts a protected

group may be unlawful. *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971); *see also*

Lanning v. SEPTA, 181 F.3d 478, 485 (3d Cir. 1999). "Title VII strives to achieve

equality of opportunity by rooting out artificial, arbitrary, and unnecessary

employer-created barriers to professional development that have a discriminatory impact

upon individuals." *Connecticut v. Teal*, 457 U.S. 440, 451 (1982) (internal quotation

marks omitted). Accordingly, the Supreme Court has noted that "[i]n considering claims

of disparate impact . . . this Court has consistently focused on employment and promotion

requirements that create a discriminatory bar to *opportunities*. This Court has never . . .

requir[ed] the focus to be placed . . . on the overall number of minority or female

applicants actually hired or promoted." *Id*. at 450.

A prima facie case of disparate impact discrimination has two components. First,

a plaintiff must identify "the specific employment practice that is challenged." *Watson v.*

Ft. Worth Bank & Trust, 487 U.S. 977, 994 (1988). Second, the plaintiff must show that

the employment practice "causes a disparate impact on the basis of race, color, religion,

sex, or national origin." 42 U.S.C. § 2000e-2(k)(1)(A)(i). To show causation, the

plaintiff must present "statistical evidence of a kind and degree sufficient to show that the

7

practice in question has caused exclusion of applicants for jobs or promotions because of their membership in a protected group." *Watson*, 487 U.S. at 994; *see also EEOC v. Greyhound Lines*, 635 F.2d 188, 193 (3d Cir. 1980).

If a plaintiff makes out a prima facie case, the burden shifts to the employer to show that the employment practice at issue is job related for the position in question and is consistent with business necessity.[4] *Watson*, 487 U.S. at 994; 42 U.S.C. § 2000e-2(k)(1) (clarifying that to maintain a claim, plaintiff must make out a prima facie case and the employer must then "fail[] to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity").[5]

B. The Prima Facie Case

As the District Court noted, there is no "rigid mathematical formula" courts can mandate or apply to determine whether plaintiffs have established a prima facie case. *Stagi II*, 2009 WL 2461892, at *3. If statistical evidence is used, as it typically will be in disparate impact cases, it must be "sufficiently substantial" to raise "an inference of causation." *Id*. (quoting *Watson*, 487 U.S. at 994-95). The Supreme Court has not provided any definitive guidance about when statistical evidence is sufficiently

---

[4] The District Court did not reach the issue of business necessity because it held that plaintiff failed to establish a prima facie case and ended its inquiry.

[5] The statute also allows plaintiff to show that an alternative employment practice exists that has a less disparate impact and would also serve the business's legitimate interest and the employer refuses to adopt it. 42 U.S.C. § 2000e-2(k)(1)(A)(ii); *Lanning*, 181 F.3d at 489-90. This alternative is not relevant here.

substantial, but a leading treatise notes that "[t]he most widely used means of showing

that an observed disparity in outcomes is sufficiently substantial to satisfy the plaintiff's

burden of proving adverse impact is to show that the disparity is sufficiently large that it

is highly unlikely to have occurred at random." 1 B. Lindemann & P. Grossman,

Employment Discrimination Law 124 (4th ed. 2007) (hereinafter "Lindemann &

Grossman"). This is typically done by the use of tests of statistical significance, which

determine the probability of the observed disparity obtaining by chance.

There are two related concepts associated with statistical significance: measures of

probability levels and standard deviation. Probability levels (also called "p-values") are

simply the probability that the observed disparity is random—the result of chance

fluctuation or distribution. For example, a 0.05 probability level means that one would

expect to see the observed disparity occur by chance only one time in twenty cases—there

is only a five percent chance that the disparity is random. A standard deviation is a unit

of measurement that allows statisticians to measure all types of disparities in common

terms.[6] In this context, the greater the number of standard deviations from the mean, the

greater the likelihood that the observed result is not due to chance. To offer some sense

---

[6] Technically, a standard deviation is defined as "a measure of spread, dispersion, or variability of a group of numbers equal to the square root of the variance of that group of numbers." D. Baldus & J. Cole, *Statistical Proof of Discrimination* 359 (1980). The "variance" of the group of numbers is computed by subtracting the "mean," or average, of all the numbers, "squaring the resulting difference, and computing the mean of these squared differences." *Id.* at 361.

of the relationship between these two measures, two standard deviations corresponds roughly to a probability level of 0.05; three standard deviations correspond to a probability level of 0.0027. *See* Lindemann & Grossman 126 n.85 and accompanying text.

As a legal matter, the Supreme Court has stated that "[a]s a general rule for . . . large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the [result] was random would be suspect to a social scientist." *Castaneda v. Partida*, 430 U.S. 482, 496 n.17 (1977). Additionally, many courts accept a 0.05 probability level (or below) as sufficient to rule out the possibility that the disparity occurred at random. *See, e.g.*, *Waisome v. Port Auth.*, 948 F.2d 1370, 1376 (2d Cir. 1991) ("Social scientists consider a finding of two standard deviations significant, meaning there is about one chance in 20 that the explanation for a deviation could be random and the deviation must be accounted for by some factor other than chance." (citation omitted)); *Palmer v. Shultz*, 815 F.2d 84, 92-96 (D.C. Cir. 1987) (noting that "statistical evidence meeting the .05 level of significance . . . [is] certainly sufficient to support an inference of discrimination" (citation and internal quotation marks omitted, alterations in original)).

In addition to using formal measures of statistical significance, some courts have also relied upon the "80 percent rule" from the Equal Employment Opportunity Commission's (EEOC) Uniform Guidelines on Employee Selection Procedures to assess

whether a plaintiff has established a prima facie disparate impact case. *See, e.g.*, *Stout v.*

Potter, 276 F.3d 1118, 1124 (9th Cir. 2002) (applying "four-fifths rule" and calling it

"rule of thumb" courts use when considering adverse impact of selection procedures);

Boston Police Superior Officers Fed'n v. City of Boston, 147 F.3d 13, 21 (1st Cir. 1998)

(affirming district court's use of four-fifths rule in context of consent decree, holding that,

although "violation of the four-fifths rule, standing alone, is not conclusive evidence of

discrimination," it nonetheless serves as an "appropriate benchmark"); *Smith v. Xerox*

Corp., 196 F.3d 358, 365 (2d Cir. 1999) (finding EEOC Guidelines "persuasive"). These

Guidelines are codified at 29 C.F.R. § 1607.4(D), entitled "Adverse impact and the 'four-

fifths rule,'" and they state, in relevant part,

> A selection rate for any race, sex, or ethnic group which is
> less than four-fifths (4/5) (or eighty percent) of the rate for the
> group with the highest rate will generally be regarded by the
> Federal enforcement agencies as evidence of adverse impact,
> while a greater than four-fifths rate will generally not be
> regarded by Federal enforcement agencies as evidence of
> adverse impact.

29 C.F.R. § 1607.4(D).

EEOC Guidelines are entitled only to *Skidmore* deference, *Skidmore v. Swift &*

Co., 323 U.S. 134, 140 (1944), under which EEOC Guidelines "get[] deference in

accordance with the thoroughness of [their] research and the persuasiveness of [their]

reasoning." *El v. SEPTA*, 479 F.3d 232, 244 (3d Cir. 2007) (citing *EEOC v. Arab*

11

American Oil Co., 499 U.S. 244, 257 (1991)).[7] The "80 percent rule" or the "four-fifths

rule" has come under substantial criticism, and has not been particularly persuasive, at

least as a prerequisite for making out a prima facie disparate impact case. The Supreme

Court has noted that "[t]his enforcement standard has been criticized on technical grounds

. . . and it has not provided more than a rule of thumb for the courts." *Watson*, 487 at 995

n.3. *See also* Lindemann & Grossman 130 (noting that the 80 percent rule "is inherently

less probative than standard deviation analysis"); E. Shoben, *Differential Pass-Fail Rates

in Employment Testing: Statistical Proof Under Title VII*, 91 Harv. L. Rev. 793, 806

(1978) (arguing that the "four-fifths rule should be abandoned altogether" and that "flaws

in the four-fifths rule can be eliminated by replacing it with a test of . . . statistical

significance").

Another non-statistical standard that has been discussed in the context of assessing

whether a plaintiff has made out a prima facie case is the requirement that the disparity

have "practical significance."[8] For example, Lindemann and Grossman write that "[t]o

_____

[7] It is worth noting that although the Supreme Court initially said that EEOC
Guidelines were entitled to "great deference," the Supreme Court itself has made it clear
that this is not the case. As we noted in *El v. SEPTA*: "It does not appear that the
EEOC's Guidelines are entitled to great deference. While some early cases so held in
interpreting Title VII, *Griggs*, 401 U.S. at 434 . . . more recent cases have held that the
EEOC is entitled only to *Skidmore* deference." 479 F.3d at 244 (citing *Arab American
Oil*, 499 U.S. at 257).

[8] A related concern, that the statistical disparity be "substantial," has been held out
as an additional requirement for a plaintiff's prima facie case. *See, e.g.*, *Thomas v.
Metroflight, Inc.*, 814 F.2d 1506, 1511 n.4 (10th Cir.1987) (suggesting that courts may
require, in addition to statistical significance, that the observed disparity be substantial).

guard against the possibility that a finding of adverse impact could result from the

statistical significance of a trivial disparity or a meaningless difference in results, the

Uniform Guidelines on Employee Selection Procedures and some courts have adopted an

additional test for adverse impact: that a statistically significant disparity also has

practical significance." Lindemann & Grossman 131 (citations omitted).

We can identify no Court of Appeals that has found "practical significance" to be a

requirement for a plaintiff's prima facie case of disparate impact, including the Third

---

This requirement, however, appears to be derived from the Supreme Court's early
disparate impact cases that were decided prior to the use of formal notions of statistical
significance as the means by which causation was to be demonstrated. In these early
formulations of the causation requirement, rather than requiring a particular level of
statistical significance, the Supreme Court required that the relevant rule had a
"substantially" disproportionate effect. *See, e.g.*, *Griggs*, 401 U.S. at 426 (examining
"requirements [that] operate[d] to disqualify Negroes at a substantially higher rate than
white applicants"); *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975) (plaintiffs
are required to show "that the tests in question select applicants for hire or promotion in a
racial pattern significantly different from that of the pool of applicants"); *Washington v.
Davis*, 426 U.S. 229, 246-47 (1976) ("hiring and promotion practices disqualifying
substantially disproportionate number of blacks"); *Dothard v. Rawlinson*, 433 U.S. 321,
329 (1977) (employment standards that "select applicants for hire in a significantly
discriminatory pattern"). The Supreme Court has made it clear that the "substantial"
language was meant to address the plaintiff's burden to demonstrate causation. As the
Supreme Court noted in *Watson*, the Supreme Court's "formulations . . . have consistently
stressed that statistical disparities must be sufficiently substantial that they raise . . . an
inference of causation," in other words, that the statistical disparities are adequate to
"show that the practice in question has *caused* the exclusion of applicants for jobs or
promotions because of their membership in a protected group." 487 U.S. at 994-95
(O'Connor, J., plurality opinion) (emphasis added). The requirement of "substantiality"
was not meant to introduce an additional burden on the plaintiff above that of offering
evidence of causation.

Circuit. The "practical significance" language stems from the EEOC Uniform Guidelines on Employee Selection Procedures, which note that "[s]maller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and *practical* terms." 29 C.F.R. § 1607.4(D) (emphasis added). However, even the non-binding EEOC Guidelines only suggest that "practical significance" might be a requirement when differences in the selection rate were greater than eighty percent. *Id*. The one case identified by Lindemann and Grossman, *Waisome*, noted that the EEOC Guidelines, including the aforementioned one, "provide no more than a rule of thumb to aid in determining whether an employment practice has a disparate impact." 948 F.2d at 1376 (internal quotation marks and citation omitted), *cited in* Lindemann & Grossman 131 n.98. The Second Circuit Court of Appeals in *Waisome* did disregard a finding of statistical significance (2.68 standard deviations), but on the grounds that the African-American pass rate for a written examination was 87% of the white pass rate, and that the statistical significance of the disparity would disappear if just two additional African-American candidates, out of a total of 64 African-American candidates, had passed the written examination. 948 F.2d at 1376-77. Other courts have also found that, in cases where the "statistical significance" of the results would disappear if the numbers were altered very slightly, the plaintiff failed to make out a prima facie case. *See, e.g.*, *Apsley v. Boeing Co.*, --- F. Supp.2d ---, No. 05-1368, 2010 WL 2670880, at *18 (D. Kan. June 30, 2010) (noting that "[s]tatistical significance does not tell us whether the disparity we

14

are observing is meaningful in a practical sense nor what may have caused the disparity," and finding that because of the fact that if "forty-eight more people over the age of 40 would have been hired, Plaintiffs' hiring statistics would not have been statistically significant," plaintiffs failed to establish a prima facie case). As "practical" significance has not been adopted by our Court, and no other Court of Appeals requires a showing of practical significance, we decline to require such a showing as part of a plaintiff's prima facie case.

In sum, to establish a prima facie case of disparate impact in a Title VII case, a plaintiff must (1) identify a specific employment policy or practice of the employer and (2) proffer evidence, typically statistical evidence, (3) of a kind and degree sufficient to show that the practice in question has caused exclusion of applicants for jobs or promotions (4) because of their membership in a protected group. *See Watson*, 487 U.S. at 994. With respect to meeting her burden with respect to (3), a plaintiff will typically have to demonstrate that the disparity in impact is sufficiently large that it is highly unlikely to have occurred at random, and to do so by using one of several tests of statistical significance. There is no precise threshold that must be met in every case, but a finding of statistical significance with a probability level at or below 0.05, or at 2 to 3 standard deviations or greater, will typically be sufficient. *See Castaneda*, 430 U.S. at 496 n.17.

III. The District Court Decision

As noted above, the District Court granted Amtrak's summary judgment motion on the grounds that Plaintiffs failed to carry their burden of presenting a prima facie case of disparate impact. This decision was based on two main considerations: (1) that "the applicant pool plaintiffs analyzed to demonstrate the disparate impact of Amtrak's policy erroneously compares employees who may not have the minimal qualifications for the particular jobs at issue," and (2) that "when viewed in context, plaintiffs' evidence of discrimination lacks practical significance." *Stagi II*, 2009 WL 2461892, at *13. The District Court's reasoning behind these conclusions is nuanced and worth considering in some detail.

The District Court, in laying out the standard for a prima facie disparate impact case, correctly noted that the plaintiff does not need to offer proof of the employer's subjective intent to discriminate, but that, instead, she must "first identify the specific employment practice that is challenged" and then she must "show causation" by offering "statistical evidence of a kind and degree sufficient to show that the practice in question has caused the exclusion of applicants for jobs or promotions because of their membership in a protected group." *Stagi II*, 2009 WL 2461892, at *3 (internal quotation marks and citations omitted). The District Court also noted that the "statistical disparities must be sufficiently substantial such that they raise an inference of causation." *Id.* (internal quotation marks and citation omitted).

The District Court then stated that there is no "rigid mathematical formula that

16

satisfies the sufficiently substantial standard in the disparate impact analysis." *Id*.

(internal quotation marks and citation omitted). But rather than discuss the importance of

various measures of statistical significance, particularly with respect to demonstrating that

the disparity is unlikely to have been the product of chance, the District Court instead

referenced the EEOC Guidelines "eighty percent" rule. The District Court stated that "the

Supreme Court has indicated that the guidance of this administrative body should be

considered with 'great deference,' and no consensus has developed around any alternative

standard." *Id*. (quoting *Griggs*, 401 U.S. at 433-34). The District Court did note that this

rule "is not intended to be an absolute requirement." *Id*.

Applying its statement of the law to the facts of the case before it, the District

Court noted that Plaintiffs satisfied the first part of their prima facie case by identifying

the one-year rule as the specific employment practice being challenged. *Id*. at *4. The

District Court then conducted an extended discussion of the statistical evidence of

disparate impact offered by Plaintiffs in the form of the expert report of Dr.

Killingsworth, and the criticism of that report by Amtrak's expert, Dr. Griffin.

The District Court found that the one-year rule makes this situation equivalent to

an "entrance requirement" case, which means that the pool of *actual* applicants for the

position will under-represent those who would otherwise qualify, because the requirement

itself would discourage the people who are claiming that the requirement has a disparate

impact from applying. *Id*. at *5. The District Court noted that "[i]n such cases, it is

proper to establish disparate impact through reference to a reasonable proxy for the pool

of individuals actually affected by the alleged discrimination." *Id*. (internal quotation

marks and citation omitted).

The District Court then discussed Dr. Killingsworth's method for creating proxy

pools. The key part of Dr. Killingsworth's method of creating the proxy pools is this

multi-step process:

> (1) Identify each management vacancy occurring during the time at issue (between March 8, 2002 and June 30, 2007).

> (2) Of that full set of vacancies, isolate the vacancies that were filled by a union employee (which we will refer to as a "job fill").

> (3) For each successful union employee, identify the job title that the union employee had *prior* to getting the management job (which we will refer to as a "feeder job").

> (4) Define a "Feeder Pool" for a particular management vacancy as the set of people who had the same job title as the successful candidate for that vacancy on the date just before the vacancy was filled.

Dr. Killingsworth's model, using the above approach, identified 716 separate "Feeder

Pools," each tied to a specific management vacancy, at a specific point in time. Each

entry in a pool is called a "candidacy," rather than a candidate or person because the same

potential applicants (or people) could be in more than one Feeder Pool. After discussing

Dr. Killingsworth's method of creating the Feeder Pools, the District Court found that

"[b]ased on the information provided to Dr. Killingsworth by Amtrak, plaintiffs' method

is a reasonable one." *Id*. at *6.

The District Court objected, however, to Dr. Killingsworth's decision to "aggregate" all of the individual Feeder Pools into "one giant pool" (the "Aggregated Pool") in order to analyze "the degree to which the Policy disqualified women in the Aggregated Pool relative to men." *Id*. Specifically, Dr. Killingsworth combined all 716 individual Feeder Pools into one large pool in order to conduct his statistical analysis. The District Court noted that when Dr. Killingsworth analyzed the data using a "corrected probit analysis" (which corrects for the fact that the same individual might appear in more than one pool), the results yielded a standard deviation of 3.855, with a p-value of less than 0.001—results which the District Court acknowledged were "unlikely to have occurred as a result of chance alone."[9] *Id*.

Despite the statistical significance of this result, however, the District Court found that Plaintiffs had not done enough to carry their prima facie burden. First, the District Court was convinced by Amtrak's argument that Dr. Killingsworth's analysis was flawed, and that the statistical significance of his result was thus irrelevant. Amtrak's expert, Dr. Griffin, offered a report demonstrating that if one does not combine the 716 Feeder Pools into one large Aggregated Pool, and if, instead, one just examines whether women in each individual Feeder Pool were ineligible at a greater than expected level (given the

---

[9] The District Court also noted that using an "uncorrected" conventional chi-square test to analyze the data, Dr. Killingsworth's results were even more statistically significant (in terms of being unlikely to have occurred at random), with a standard deviation measure of 8.42.

ineligibility rate of that particular pool), one does not find that women were disadvantaged relative to men at a statistically significant level.

Dr. Griffin determined this by first determining the percentage of ineligible men and women in a particular Feeder Pool (i.e., if 50 out of 500 people are blocked, the total ineligibility rate would be 10%). Next, Dr. Griffin multiplied that percentage by the total number of women in the pool to determine the number of "expected" ineligibles (i.e., if there were 300 women in the pool, multiplied by 10%, one would expect 30 women in the pool to be ineligible). Finally, Dr. Griffin compared the "expected" number of ineligible women with the actual number of ineligible women in the pool, to assess whether there was a shortfall or a surplus of ineligible women in that particular pool, relative to what was expected (i.e., if 20 women were actually ineligible, then there would be a shortfall of 10 women—10 fewer women were ineligible than would be expected given the Feeder Pool's particular ineligibility rate as a whole).

Having conducted this analysis for approximately 600 "job fills," Dr. Griffin then summed the surpluses and shortfalls of ineligible women across those approximately 600 "job fills." This resulted in a net surplus of 6.2 ineligible women, meaning that 6.2 fewer women were promotion eligible than would have been if there were perfect gender parity across all 600 job fills. As the District Court noted, "[s]ix fewer promotion eligible females across 600 plus 'job fills' is *not* statistically significant by any measure, and does not support an inference of discrimination." *Id*. at *8 (emphasis in original).

At this point, the District Court noted that "the parties have merely presented two different statistical models that produce opposite results," and that "[s]imply demonstrating that an alternative analysis leads to alternative results is not sufficient to defeat a plaintiff's prima facie case—the defendant must also show that there is no genuine issue of material fact that plaintiffs' model is fundamentally flawed for the purpose of demonstrating disparate impact in the case at issue." *Id*. (citation omitted). The District Court continued:

> The key difference between the experts can be boiled down to this: Dr. Griffin looks at whether women applying to job X are disadvantaged relative to men applying to job X, whereas Dr. Killingsworth analyzes whether women applying to jobs X and Y are disadvantaged relative to men applying for jobs X and Y, combined. When seen in those terms, the difference between the expert analysis presented in this case is simply a question of whether the plaintiffs have analyzed the appropriate relevant labor pool for purposes of comparison. This question can be decided as a matter of law.

Id. at *9. Essentially, the District Court saw itself as forced to decide whether Dr. Killingsworth's decision to aggregate the 716 Feeder Pools into one Aggregate Pool was appropriate, and considered this to be a question of law.

The District Court noted that "[a]ggregated statistical data may be properly used to prove disparate impact where it is more probative than subdivided data," *id*. (citing *Paige v. California*, 291 F.3d 1141, 1148 (9th Cir. 2002)), but that "'[w]hen special qualifications are required to fill particular jobs, comparisons to the general population (rather than to the smaller group of individuals who possess the necessary qualifications)

21

may have little probative value.'" *Id*. (quoting *Hazelwood Sch. Dist. v. U.S.*, 433 U.S. 299, 308 n. 13 (1977)). The District Court then stated that Dr. Killingsworth acknowledged that every union employee was not fungible for purposes of promotion, since he created the 716 Feeder Pools, "otherwise he would have simply compared all union employees across the board." *Id*. The District Court contended that because Dr. Killingsworth takes the "distinctions between job categories [to be] important . . . then the defendant's argument that these distinctions should be maintained throughout the analysis rings true." *Id*. Accordingly, the District Court found that "because plaintiffs' analysis is focused on an overbroad and incomparable pool of employees, it lacks the statistical significance necessary to make out a prima facie case of discrimination." *Id*. at *11.

In the alternative, the District Court found that "[e]ven if Dr. Killingsworth's methodology was sound and his results recognized as having 'statistical significance,' the results of his analysis are undermined by a lack of practical significance." *Id*. at *12. To reach this conclusion, the District Court credited Dr. Griffin's calculation that if female candidates in the Aggregated Pool had the same eligibility rate as male candidates, this would have translated to a "gender gap" of only 726 additional female promotion-eligible candidacies (not necessarily equal to the number of affected individual people or candidates) overall. The District Court also noted that, under the EEOC Guidelines' "80 percent rule," the adverse impact ratio's "practical significance is of limited magnitude," since the ratio here was 96.8 percent—well over the 80 percent baseline. *Id*.

22

In conclusion, the District Court found that "the applicant pool plaintiffs analyzed to demonstrate the disparate impact of Amtrak's policy erroneously compares employees who may not have the minimal qualifications for the particular jobs at issue," and that "plaintiffs' evidence of discrimination lacks practical significance." *Id*. at *13. The Court therefore granted Amtrak's motion for summary judgment.

IV.

We review a district court's grant of summary judgment *de novo*. *See, e.g.*, *Slagle v. County of Clarion*, 435 F.3d 262, 263 (3d Cir. 2006). Under Rule 56(c) of the Federal Rules of Civil Procedure, summary judgment is appropriate when "there is no genuine issue as to any material fact." The moving party "bears the initial responsibility of informing the district court of the basis for its motion, and identifying those portions of the pleadings, depositions, answers to interrogatories, and admissions on file, together with the affidavits, if any, which it believes demonstrate the absence of a genuine issue of material fact." *El*, 479 F.3d at 237 (quoting *Celotex Corp. v. Catrett*, 477 U.S. 317, 323 (1986)). The court must draw all reasonable inferences against the moving party. *Id*. at 238. "If the moving party successfully points to evidence of all of the facts needed to decide the case on the law short of trial, the non-moving party can defeat summary judgment if it nonetheless produces or points to evidence in the record that creates a genuine issue of material fact." *Id*. "Thus, if there is a chance that a reasonable factfinder would not accept a moving party's necessary propositions of fact, pre-trial

judgment cannot be granted." *Id*.

We find that there is a genuine issue of material fact as to whether the one-year rule caused a disparate impact on female employees. Accordingly, although it is a close case, we find that the District Court should not have granted Amtrak's motion for summary judgment based on this record.

As noted above, to establish a prima facie case of disparate impact in a Title VII case, a plaintiff must (1) identify a specific employment policy or practice of the employer and (2) proffer evidence, typically statistical evidence, (3) of a kind and degree sufficient to show that the practice in question has caused exclusion of applicants for jobs or promotions (4) because of their membership in a protected group. To establish (3), a plaintiff will typically have to demonstrate that the disparity in impact is sufficiently large that it is highly unlikely to have occurred at random, and to do so by using one of several tests of statistical significance. A plaintiff need not demonstrate that the disparate impact ratio satisfies the EEOC's 80 percent rule (the figure at which or below the EEOC will presume the existence of disparate impact). As noted above, the EEOC Guidelines are not entitled to great deference, but to *Skidmore* deference, under which EEOC Guidelines "get[] deference in accordance with the thoroughness of [their] research and the persuasiveness of [their] reasoning." *El*, 479 F.3d at 244 (citing *EEOC v. Arab American Oil Co.*, 499 U.S. at 257). The 80 percent rule has come under significant criticism and we do not find the reasoning that might support its application here persuasive in light of

the statistical significance of Dr. Killingsworth's results.

Similarly, this Court has never established "practical significance" as an independent requirement for a plaintiff's prima facie disparate impact case, and we decline to do so here. The EEOC Guidelines themselves do not set out "practical" significance as an independent requirement, and we find that in a case in which the statistical significance of some set of results is clear, there is no need to probe for additional "practical" significance. Statistical significance is relevant because it allows a fact-finder to be confident that the relationship between some rule or policy and some set of disparate impact results was not the product of chance. This goes to the plaintiff's burden of introducing statistical evidence that is "sufficiently substantial" to raise "an inference of causation." *Watson*, 487 U.S. at 994-95. There is no additional requirement that the disparate impact caused be above some threshold level of practical significance. Accordingly, the District Court erred in ruling "in the alternative" that the absence of practical significance was fatal to Plaintiffs' case.

There is no question that Dr. Killingsworth's results, if the product of a relevant and otherwise compelling statistical analysis, are statistically significant above the threshold that courts have required.[10] As noted above, when Dr. Killingsworth analyzed the data using a corrected probit analysis, the results yielded a standard deviation of

---

[10] Even Amtrak concedes that the results, if they stand, meet the threshold requirement for statistical significance. Oral Arg. Tr., 47-48.

3.855, with a p-value of less than 0.001—meaning the results are incredibly unlikely to have occurred as a result of chance alone. The Supreme Court has suggested that a standard deviation between 2 and 3 would be sufficient, and Dr. Killingsworth's results are considerably above that. *See, e.g.*, *Castaneda*, 430 U.S. at 496 n.17.

Thus, the only issue is whether the District Court was correct in finding that Dr. Killingsworth's statistical analysis was, in effect, legally irrelevant to satisfying Plaintiffs' burden with respect to their prima facie case because his analysis used aggregation, and in particular the Aggregated Pool, in conducting his statistical analysis. We find that Dr. Killingsworth's decision to aggregate the data, although not obviously correct, is also not obviously incorrect, and so there remains a genuine issue of material fact—whether the one-year rule caused a disparate impact on Amtrak's female employees.

The one-year rule applies to all union employees. However, including all union employees in the statistical sample would have been inappropriate, since many of them may not have been even remote candidates for *any* management position. To identify all those union employees who might reasonably be thought to be candidates for a management position, Dr. Killingsworth identified those candidates who obtained a management position during the relevant five-year span, and then identified the previous union positions held by those candidates. At that point, Dr. Killingsworth assumed, and the District Court found this assumption reasonable, that all those individuals who were in the same union position as the position that the successful candidate had previously

26

occupied might reasonably be thought to have been a possible candidate for the

management position that the successful candidate actually obtained. Thus, if Smith was

hired into Management Position One, and Smith had previously been in Union Position

One, Dr. Killingsworth assumed that all other individuals—Jones, Williams, Johnson,

etc.—who had been in Union Position One were possible candidates for Management

Position One. This is not a perfect proxy, as all parties concede. For example, Smith

might have had much more experience than Jones and Williams, or he might have

educational degrees that they lack. But, given that the one-year rule operates as an initial

bar from even becoming a candidate for a job, the only way to measure its effect is to

devise some way of identifying those who might reasonably be thought to have been

possible candidates were it not for the existence of the one-year rule. We agree with the

District Court that Dr. Killingsworth's method here was reasonable.

It is true that while "the population selected for statistical analysis need not

perfectly match the pool of qualified persons," without "a close fit between the population

used to measure disparate impact and the population of those qualified for a benefit, the

statistical results cannot be persuasive." *Carpenter v. Boeing Co.*, 456 F.3d 1183, 1196

(10th Cir. 2006). One must have the proper pool of people in view before performing

statistical analysis, or that analysis will be irrelevant. This, however, goes to the issue of

whether Dr. Killingsworth's use of the individual Feeder Pools was reasonable or not. In

discussing this issue, the District Court stated:

> In the absence of explicit measures of qualifications and job
> interest, Dr. Killingsworth assumed that information about the
> position held prior to promotion could reasonably serve as an
> indicator of qualifications and job interest. Based on the
> information provided to Dr. Killingsworth by Amtrak,
> plaintiffs' method is a reasonable one.

Stagi II, 2009 WL 2461892, at *6. We agree.

Where the District Court identified a problem was with the combining of the individual Feeder Pools into one Aggregated Pool. The District Court stated that because Dr. Killingsworth takes the "distinctions between job categories [to be] important" in creating the individual Feeder Pools, "then the defendant's argument that these distinctions should be maintained throughout the analysis rings true." *Id*. at *9. Amtrak's counsel made this same point repeatedly at oral argument, stating that "if you're going to live in a stratified world, you have to follow that stratified world through to your analysis" and that "the problem is that we're aggregating after we stratify, that's the heart of the matter." Oral Arg. Tr., at 41, 45.

However, neither the District Court nor Amtrak's counsel has offered a convincing explanation of why the use of aggregated data in this case is improper. The District Court reintroduces the "qualifications" issue, asserting that "[t]he single aggregated statistic Dr. Killingsworth relies on compares individuals who may never actually be in competition for the same jobs, and does not accurately account for what job the employee in question is coming from, where they are looking to go, and what the relevant qualifications are." *Stagi II*, 2009 WL 2461892, at *9. But this criticism misses its target. Creating the

28

Aggregated Pool out of the individual Feeder Pools does not erroneously imply that a

person from Feeder Pool A (created based on Management Position A) is a possible

candidate, along with the members of Feeder Pool B, for Management Position B.

Rather, it just puts together all of those people (or candidacies, more precisely) who are in

union positions currently, and who are reasonably thought of as possible candidates for

some management position or other. All of these people are susceptible to the one-year

rule, and thus all of them are potentially "blocked" by its uniform application if they have

served less than one year in their respective union positions. Aggregating the individual

Feeder Pools in this way appears to be no more problematic, at least with respect to the

issue of qualifications, than doing what Dr. Griffin did when he simply "added up" the

difference between the expected ineligibility rate and the actual ineligibility rate for each

of the 600 plus individual Feeder Pools.

At various points, Amtrak's counsel at oral argument appeared to be arguing that,

as a matter of *consistency*, once one has subdivided the pool into categories, one ought

not to recombine those categories into an aggregate pool. The District Court appeared to

accept a similar line of thought when it noted that because Dr. Killingsworth took the

"distinctions between job categories [to be] important" in creating the individual Feeder

Pools, "then the defendant's argument that these distinctions should be maintained

throughout the analysis rings true." *Id*. at *9. But there has been no argument made that

somehow the statistical analysis is corrupted if one "changes horses" from a stratified to

an aggregated analysis midstream. Indeed, Amtrak's counsel explicitly stated that "the

actual manner in which [Dr. Killingsworth] performs the numbers is not incorrect, it's the

underlying numbers that are the problem." Oral Arg. Tr., at 44. Finally, Plaintiffs'

counsel stresses that they never were doing a "stratification" analysis in the first place, but

that they were simply attempting to "define what is the subset of total union employees

who seemed to be in positions that made them eligible to seek promotion." *Id*. at 56.

A final possible reason to object to the use of aggregated data is presented by the

District Court when it notes that Dr. Griffin's report suggests that there are some Feeder

Pools in which fewer women than men were made ineligible by the one-year rule, and

some in which the reverse was true, and that the overall result of women doing worse

than men (at least under Dr. Killingsworth's model) obscures these facts. This would be

a reason against aggregating insofar as aggregating produces a misleading picture of the

overall situation for women. (As one court has noted, "[i]f Microsoft-founder Bill Gates

and nine monks are together in a room, it is accurate to say that on average the people in

the room are *extremely* well-to-do, but this kind of aggregate analysis obscures the fact

that ninety percent of the people in the room have taken a vow of poverty." *Abram v.*

United Parcel Serv. Inc., 200 F.R.D. 424, 431 (E.D. Wis. 2001).) For example, it might

be that in 400 of the 716 Feeder Pools, women are made ineligible at a rate significantly

greater than that of men, and that in 316 of the Feeder Pools, the reverse is true. In such a

situation, the one-year rule appears to have a disparate impact on women only in a subset

30

of the 716 Feeder Pools.

Plaintiffs' second expert, Ramona Paetzold, submitted an affidavit arguing that stratification is inappropriate in this case precisely because of this possibility. In particular, stratification is inappropriate because the numbers of women in each feeder job at any given point in time is determined, in part, by the existence of the one-year rule itself, "because the one-year rule at least partially affects how long men and women must remain in the feeder job before being eligible for promotion." Paetzold Aff. 3. The District Court contends that this is a problem for Plaintiffs, because "the gender composition of feeder jobs may very well be affected by additional factors such as wage levels, working conditions, movement prospects, layoffs, and the union's collectible bargaining agreement that allows unrestricted lateral job movements among union employees, none of which the plaintiffs have made any attempt to identify or control for in their analysis." *Stagi II*, 2009 WL 2461892, at *10. But this seems to be a problem only if the reasons against aggregation are compelling. There is no legal requirement to use the smallest possible unit of analysis. If there are additional factors (such as seniority rules)—apart from just the one-year rule—that are determining the composition of the individual Feeder Pools in a "gendered" way, these factors may aid Amtrak in mounting a business justification defense, but it is inappropriate to require Plaintiffs to control for every possible such factor in order to sustain their burden of proving a prima facie case. If the aggregated data yields a statistically significant finding, such as the one here, that

31

the one-year rule is having a disparate impact on women, and there is no compelling

reason to avoid use of aggregated data, that is enough for Plaintiffs to establish their

prima facie case.

Additionally, there may be good reasons to aggregate data in a case such as

this—reasons that have nothing to do with simply picking and choosing the model which

will generate the most favorable results for plaintiffs' case. Perhaps most significantly, as

the Fourth Circuit has observed, "by increasing the absolute numbers in the data, chance

will more readily be excluded as a cause of any disparities found." *Lilly v. Harris-Teeter

Supermarket*, 720 F.2d 326, 336 n.17 (4th Cir. 1983). This makes intuitive sense. "For

example, if a coin were tossed ten times . . . and came up heads four times, no one would

think the coin was biased (0.632 standard deviations), but if this same ratio occurred for a

total of 10,000 tosses, of which 4,000 were heads, the result could not be attributed to

chance (20 standard deviations)." *Id*. Here, by combining all of those candidacies in the

716 Feeder Pools into one Aggregated Pool, Dr. Killingsworth was better able to test

whether the difference in the ineligibility rate for men and women was merely the product

of chance. Many courts have found such a reason for aggregating compelling. *See, e.g.*,

*Eldredge v. Carpenters 46 N. California Counties Joint Apprenticeship and Training

Comm.*, 833 F.2d 1334, 1339 (9th Cir. 1987) ("Aggregated data presents a more complete

and reliable picture."); *Cook v. Boorstin*, 763 F.2d 1462, 1468-69 (D.C. Cir. 1985)

(rejecting defendant's argument to restrict statistical analysis to particular job categories);

Capaci v. Katz & Besthoff, 711 F.2d 647, 654 (5th Cir. 1983) (allowing a plan to aggregate data over several years because aggregation was necessary in order to accomplish a meaningful statistical analysis).

At a minimum, we find that there is a genuine issue of material fact as to whether the one-year rule caused a disparate impact on female employees. It is possible that there are reasons to prefer Dr. Griffin's methodology to Dr. Killingsworth's methodology, given that they yield conflicting conclusions regarding whether the one-year rule has an all-things-considered disparate impact on women. But we cannot so conclude on this record, and the reasons presented by the District Court for finding that Plaintiffs have failed to make out a prima facie case do not withstand scrutiny. Accordingly, we find that the District Court should not have granted Amtrak's motion for summary judgment based on this record.

V.

We will reverse the judgment of the District Court granting Amtrak's motion for summary judgment and will remand for further proceedings consistent with this opinion.