

PRECEDENTIAL

UNITED STATES COURT OF APPEALS
FOR THE THIRD CIRCUIT

No. 15-3435

RUDOLPH A. KARLO; MARK K. MCLURE;
WILLIAM S. CUNNINGHAM; JEFFREY MARIETTI;
DAVID MEIXELSBERGER,
Appellants

v.

PITTSBURGH GLASS WORKS, LLC

On Appeal from the United States District Court
for the Western District of Pennsylvania
District Court No. 2-10-cv-01283
District Judge: The Honorable Terrence F. McVerry

Argued November 9, 2016

Before: SMITH, *Chief Judge*, McKEE, and RESTREPO,
Circuit Judges

(Opinion Filed: January 10, 2017)

Samuel J. Cordes **[ARGUED]**
245 Fort Pitt Boulevard
Pittsburgh, PA 15222

Bruce C. Fox
Andrew J. Horowitz
Obermayer Rebmann Maxwell & Hippel
500 Grant Street
Suite 5240, One Mellon Center
Pittsburgh, PA 15219
Counsel for Appellant

Rachel E.A. Atterberry
David S. Becker **[ARGUED]**
Jennifer L. Fitzgerald
Tina C. Wills
Freeborn & Peters
311 South Wacker Drive
Suite 3000
Chicago, IL 60606

Robert B. Cottington
Cohen & Grigsby
625 Liberty Avenue
Pittsburgh, PA 15222
Counsel for Appellee

Neal D. Mollen **[ARGUED]**
Paul Hastings
875 15th Street, N.W.
Suite 1000
Washington, DC 20005
*Counsel for Amicus Appellee Chamber of
Commerce of the United States of America*

Michael P. Bracken
NT Lakis
1501 M Street, N.W.
Suite 400
Washington, DC 20005
*Counsel for Amicus Appellee Equal Employment
Advisory Council*

Anne N. Occhialino **[ARGUED]**
Equal Employment Opportunity Commission
5th Floor
131 M Street, N.E.
Washington, DC 20507
*Counsel for Amicus Appellant Equal Employment
Opportunity Commission*

OPINION

SMITH, *Chief Judge*.

The Age Discrimination in Employment Act (“ADEA”) protects only those individuals who are at least forty years of age. The question in this case is whether a disparate-impact claim is cognizable where a “subgroup” of employees at the upper end of that range—in this case, employees aged fifty and older—were alleged to have been disfavored relative to younger employees.

We answer in the affirmative. Our decision is dictated by the plain text of the statute as interpreted by the Supreme Court. In particular, the ADEA prohibits disparate impacts based on *age*, not forty-and-older identity. A rule that disallowed subgroups would ignore genuine statistical disparities that could otherwise be actionable through application of the plain text of the statute. Although several of our sister circuits have ruled to the contrary, their reasoning relies primarily on policy arguments that we do not find persuasive.

We will therefore reverse the judgment of the District Court based on its interpretation of the ADEA. We will also vacate the District Court’s order excluding the testimony of plaintiffs’ statistics expert and remand for further *Daubert* proceedings. We will affirm in all other respects.

I

Defendant Pittsburgh Glass Works, LLC (“PGW”) manufactures automotive glass in Harmarville, Pennsylvania. PGW also owns (1) GTS Services, a software business, (2) PGW Auto Glass, an automotive replacement-glass distribution business, (3) LYNX Services, an insurance claims administrator, and (4) Aquapel, a glass treatment supplier.

In 2008, the automobile industry began to falter. PGW engaged in several reductions in force (“RIFs”) to offset deteriorating sales. The RIF of relevance to this case occurred on March 31, 2009, and terminated the employment of approximately one hundred salaried employees in over forty locations or divisions. Individual unit directors had broad discretion in selecting whom to terminate. PGW did not train those directors in how to implement the RIF. Nor did PGW employ any written guidelines or policies, conduct any disparate-impact analysis, review prospective RIF termines with counsel, or document why any particular employee was selected for inclusion in the RIF.

Plaintiffs Rudolph A. Karlo, William S. Cunningham, Jeffrey Marietti, David Meixelberger, Mark K. McLure, Benjamin D. Thompson, and Richard

Csukas¹ worked in PGW's Manufacturing Technology division. They were terminated as part of the March 2009 RIF by their supervisor, Gary Cannon. Each was over fifty years old at the time.

In January 2010, plaintiffs filed charges of employment discrimination with the Equal Employment Opportunity Commission ("EEOC"). Thereafter, they received a Dismissal and Notice of Rights from the EEOC, and this lawsuit followed. Plaintiffs brought a putative ADEA collective action, asserting three claims: (1) disparate treatment, (2) disparate impact, and (3) retaliation as to only Karlo and McLure.

On plaintiffs' motion for conditional certification, the District Court ruled that ADEA subgroups are cognizable, and conditionally certified a collective action to be comprised of employees terminated by the RIF who were at least fifty years old at the time. *See Karlo v. Pittsburgh Glass Works, LLC*, 880 F. Supp. 2d 629 (W.D. Pa. 2012). In addition to the named plaintiffs, eleven individuals opted in. Three voluntarily dismissed their claims and four settled. Four opt-ins remained: Michael Breen, a former production supervisor at a plant in Crestline, Ohio; Matthew Clawson, a former Project Engineer in Evansville, Indiana; Stephen Shaw, a former

¹ McLure, Thompson, and Csukas settled their claims prior to this appeal.

marketing manager in Pittsburgh, Pennsylvania; and John Titus, a former Area Services Manager in Irving, Texas.

On June 26, 2013, the case was transferred to another district judge. PGW filed a motion to decertify the collective action. On March 31, 2014, the District Court granted the motion, concluding that the collective action should be decertified because the opt-in plaintiffs' claims are factually dissimilar from those of the named plaintiffs. *See Karlo*, 2014 WL 1317595.

PGW then filed motions to exclude plaintiffs' experts. Of relevance to this appeal, PGW sought to exclude three areas of expert testimony. First, Dr. Michael Campion was prepared to offer statistical evidence in favor of plaintiffs' disparate-impact claim. Second, Dr. Campion intended to offer his expert opinion on "reasonable" human-resources practices during a RIF. And third, Dr. Anthony G. Greenwald proposed to testify as to age-related implicit-bias studies. By Order dated July 13, 2015, the District Court excluded the testimony of each. *See Karlo*, 2015 WL 4232600.

PGW moved for summary judgment on each claim. On September 3, 2015, the District Court ruled on the motions, granting them in part and denying them in part. *See Karlo*, 2015 WL 5156913. As to plaintiffs' disparate-impact claims, the District Court granted summary judgment on two grounds: (1) plaintiffs' fifty-and-older disparate-impact claim is not cognizable under

the ADEA; and (2) plaintiffs' lack of evidence to support their claim of disparate impact following the exclusion of Dr. Champion's statistics-related testimony. The District Court also granted summary judgment as to plaintiffs' disparate-treatment claims. That ruling has not been appealed. Finally, the District Court denied summary judgment as to Karlo's and McLure's individual retaliation claims.

On October 2, 2015, the District Court certified the disparate-impact and disparate-treatment claims for final judgment pursuant to Rule 54(b) of the Federal Rules of Civil Procedure. *See Karlo*, 2015 WL 5782062. This appeal followed. Plaintiffs seek reversal of the District Court's summary judgment decision and statistics-related *Daubert* ruling regarding their disparate-impact claims. Plaintiffs also appeal the District Court's other *Daubert* rulings and its order decertifying the collective action.

II

The District Court had jurisdiction pursuant to 28 U.S.C. § 1331. We have jurisdiction pursuant to 28 U.S.C. § 1291.

The parties dispute whether our jurisdiction extends to one or all named plaintiffs. PGW concedes that Karlo perfected an appeal, but argues that the other remaining named plaintiffs—Cunningham, Marietti, and Meixelberger—were not identified in the Notice of

Appeal, and therefore did not preserve their appellate rights under Rule 3(c) of the Federal Rules of Appellate Procedure. *See Torres v. Oakland Scavenger Co.*, 487 U.S. 312, 317 (1988).² We conclude that plaintiffs complied with Rule 3(c) with respect to all named plaintiffs.

Rule 3(c)(1)(A) requires a notice of appeal to “specify the party or parties taking the appeal by naming each one in the caption or body of the notice,” but that rule is relaxed where “an attorney [is] representing more than one party.” Fed. R. App. P. 3(c)(1)(A). The attorney “may describe those parties with such terms as ‘all plaintiffs,’ ‘the defendants,’ ‘the plaintiffs A, B, et al.,’ or ‘all defendants except X.’” *Id.*

The Notice of Appeal here states, “*Plaintiffs* in the above-captioned case hereby appeal . . . an order . . . entering judgment against *Plaintiffs* . . . on *Plaintiffs’* discrimination claims” A.1 (emphases added). The

² Although it does not influence our analysis, the Notice of Appeal’s caption has a perfectly innocent explanation. Following Rule 54(b) certification, the District Court amended the caption by order to identify only Karlo and McLure as plaintiffs because their retaliation claims were the only claims remaining after summary judgment. The Notice of Appeal used the District Court’s updated caption.

use of “Plaintiffs” is equivalent to “the defendants” in the example provided by the Rule.³ We have observed that “[t]he purpose of Rule 3(c)’s identification requirement is to provide notice to the court and the opposing parties of the identity of the appellants.” *In re Cont’l Airlines*, 125 F.3d 120, 129 (3d Cir. 1997). Because all of the remaining named plaintiffs were identically situated as to this appeal, were represented by the same counsel, and were each identified by name in the District Court’s “order . . . entering judgment against [all named] Plaintiffs,” as referenced on the face of the Notice, Rule 3(c)’s purpose is amply served, and “the intent to appeal is otherwise clear from the notice.” Fed. R. App. P. 3(c)(4); see *United States v. Carelock*, 459 F.3d 437, 441 (3d Cir. 2006) (“The Supreme Court has stated that courts should ‘liberally construe the requirements of Rule 3.’” (quoting *Smith v. Barry*, 502 U.S. 244, 248 (1992))).

III

The central question in this case is whether so-called “subgroup” disparate-impact claims are cognizable under the ADEA. We hold that they are.

³ The phrase “in the above captioned case” does not change our interpretation. We read the Notice to mean “Plaintiffs in [Civil Action No. 10-1283] hereby appeal.”

Disparate-impact claims in ADEA cases ordinarily evaluate the effect of a facially neutral policy on all employees who are at least forty years old—that is, all employees covered by the ADEA. In this case, plaintiffs claim to have identified a policy that disproportionately impacted a subgroup of that population: employees older than fifty. But because the policy favored younger members of the protected class, adding those individuals to the comparison group washes out the statistical evidence of a disparity.

Plaintiffs' claim is cognizable under the ADEA. Specifically, we hold that an ADEA disparate-impact claim may proceed when a plaintiff offers evidence that a specific, facially neutral employment practice caused a significantly disproportionate adverse impact based on age. Plaintiffs can demonstrate such impact with various forms of evidence, including forty-and-older comparisons, subgroup comparisons, or more sophisticated statistical modeling, so long as that evidence meets the usual standards for admissibility. A contrary rule would ignore significant age-based disparities. Where such disparities exist, they must be justified pursuant to the ADEA's relatively broad defenses.

A

We begin with an overview of the statutory scheme. The Age Discrimination in Employment Act of

1967, 81 Stat. 602, as amended, 29 U.S.C. § 621 *et seq.*, makes it unlawful for an employer:

(1) to fail or refuse to hire or to discharge any individual or otherwise discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's age;

(2) to limit, segregate, or classify his employees 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 age; or

(3) to reduce the wage rate of any employee in order to comply with this chapter.

29 U.S.C. § 623(a). "Except for substitution of the word 'age,' for the words 'race, color, religion, sex, or national origin,' the language of that provision in the ADEA is identical to that found in § 703(a)(2) of the Civil Rights Act of 1964 (Title VII)." *Smith v. City of Jackson*, 544 U.S. 228, 233 (2005). But unlike Title VII, which protects individuals of every race, color, religion, sex, and national origin, the ADEA's protection is "limited to individuals who are at least 40 years of age." 29 U.S.C. § 631(a).

ADEA claims may proceed under a disparate-impact or disparate-treatment theory. *See Smith*, 544 U.S. at 231–32. Disparate treatment is governed by § 623(a)(1); disparate impact is governed by § 623(a)(2). *Id.* at 235 (plurality opinion); *cf. Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 991 (1988); *Connecticut v. Teal*, 457 U.S. 440, 446–47 (1982).

The disparate-impact theory of recovery was first recognized in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), a Title VII case. Unlike claims of disparate treatment, disparate-impact claims do not require proof of discriminatory intent. Disparate impact redresses policies that are “fair in form, but discriminatory in operation.” *Id.* at 431. To that end, disparate-impact claims “usually focus[] on statistical disparities” *Watson*, 487 U.S. at 987.

To state a *prima facie* case for disparate impact under the ADEA, a plaintiff must (1) identify a specific, facially neutral policy, and (2) proffer statistical evidence that the policy caused a significant age-based disparity. *Cf. NAACP v. N. Hudson Reg’l Fire & Rescue*, 665 F.3d 464, 476–77 (3d Cir. 2011). Once a plaintiff establishes a *prima facie* case, an employer can defend by arguing that the challenged practice was based on “reasonable factors other than age”—commonly referred to as the “RFOA” defense. 29 U.S.C. § 623(f)(1); 29 C.F.R. § 1625.7.

“[T]he scope of disparate-impact liability under the ADEA is narrower than under Title VII” because of “[t]wo textual differences” between the statutes. *Smith*, 544 U.S. at 240. First, the RFOA defense imposes a lighter burden on the employer than its Title VII counterpart, the “business necessity” defense. Under the ADEA, the employer only needs to show that it relied on a “reasonable” factor, not that “there are [no] other ways for the employer to achieve its goals” *Smith*, 544 U.S. at 243. Congress’s decision to impose a relatively light burden on employers “is consistent with the fact that age, unlike race or other classifications protected by Title VII, not uncommonly has relevance to an individual’s capacity to engage in certain types of employment.” *Id.* at 240. The second textual difference requires ADEA plaintiffs to “isolat[e] and identify[] the *specific* employment practices that are allegedly responsible for any observed statistical disparities.” *Id.* at 241 (quoting *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 656 (1989)). Congress stripped that requirement from Title VII when it amended the statute in 1991, but it remains operative under the ADEA. *Id.* at 240; *see* 42 U.S.C. § 2000e–2(k).

B

The ADEA’s disparate-impact provision makes it unlawful for an employer “to adversely affect [an employee’s] status . . . because of such individual’s age.” 29 U.S.C. § 623(a)(2). This plain text supports the

viability of subgroup claims. *See Hardt v. Reliance Standard Life Ins. Co.*, 560 U.S. 242, 251 (2010) (“We must enforce plain and unambiguous statutory language according to its terms.”). Two aspects of the text guide our decision in this case: (1) the focus on *age* as the relevant protected trait, as interpreted by *O’Connor v. Consolidated Coin Caterers Corp.*, 517 U.S. 308 (1996), and (2) the focus on the rights of *individuals*, as interpreted by *Connecticut v. Teal*, 457 U.S. 440 (1982). Our interpretation is further supported by the ADEA’s remedial purpose.

1

We begin with the Supreme Court’s unanimous opinion in *O’Connor v. Consolidated Coin Caterers Corp.*, 517 U.S. 308 (1996), an ADEA disparate-treatment case. *O’Connor* clarified that the ADEA proscribes *age* discrimination, not forty-and-over discrimination. The same interpretation applies to identical operative language in the ADEA’s disparate-impact provision.

The plaintiff in *O’Connor* was fifty-six years old when he was fired and replaced with a younger worker. 517 U.S. at 309. The plaintiff’s replacement, however, was over the age of forty, and therefore within the class of individuals protected by the ADEA. *Id.* The Fourth Circuit held that the ADEA *prima facie* case requires the

replacement to be younger than forty years old. *Id.* at 310. The Supreme Court reversed.

The Supreme Court began its analysis with the plain text of the statute: “The discrimination prohibited by the ADEA is discrimination ‘because of [an] individual’s age,’ though the prohibition is ‘limited to individuals who are at least 40 years of age.’” 517 U.S. at 312 (alteration in original) (citations omitted). On the basis of that text, the Court held that the ADEA

does not ban discrimination against employees because they are aged 40 or older; it bans discrimination against employees because of their age, but limits the protected class to those who are 40 or older. The fact that one person in the protected class has lost out to another person in the protected class is thus irrelevant, so long as he has lost out *because of his age*.

Id. Although the ADEA protects a class of individuals at least forty years old, it “prohibits discrimination on the basis of age and not class membership” *Id.* at 313. It is therefore “utterly irrelevant” that the beneficiary of age discrimination was also over the age of forty. *Id.* at 312. Accordingly, the proposed limitation on the *prima facie* case—replacement by an employee younger than forty—lacked a “logical connection” to the plain text of the ADEA. *Id.* at 311–12. As the Supreme Court later

reaffirmed, “[it] is beyond reasonable doubt[] that the ADEA was concerned to protect a relatively old worker from discrimination that works to the advantage of the relatively young.” *Gen. Dynamics Land Sys., Inc. v. Cline*, 540 U.S. 581, 590–91 (2004).

The Supreme Court’s reasoning ineluctably leads to our conclusion that subgroup claims are cognizable. Simply put, evidence that a policy disfavors employees older than fifty is probative of the relevant statutory question: whether the policy creates a disparate impact “because of such individual[s]’ age.” 29 U.S.C. § 623(a)(2). Requiring the comparison group to include employees in their forties has no “logical connection” to that prohibition. *O’Connor*, 517 U.S. at 311.

The key insight from *O’Connor* is that the forty-and-older line drawn by § 631(a) constrains the ADEA’s general scope; it does not modify or define the ADEA’s substantive prohibition against “discriminat[ion] . . . because of such individual’s age.” § 623(a)(1). The ADEA protects against “age discrimination []as opposed to ‘40 or over’ discrimination” *O’Connor*, 517 U.S. at 312.

The disparate-impact provision uses the same operative phrase, “because of such individual’s age.” 29 U.S.C. § 623(a)(2). Our interpretation of it, therefore, should be consistent with our interpretation of the disparate-treatment provision, § 623(a)(1). *See, e.g.*,

Dep't of Revenue of Or. v. ACF Indus., Inc., 510 U.S. 332, 342 (1994) (“[I]dential words used in different parts of the same act are intended to have the same meaning.”). Thus, “adversely affect . . . because of such individual’s age” must mean adversely affect based on *age*, not adversely affect based on forty-and-older status.⁴

O’Connor’s applicability is not diminished by the fact that it addressed a disparate-treatment claim. As demonstrated by the identical operative phrasing of § 623(a)(1) and § 623(a)(2), the two types of claims share the same “ultimate legal issue” *Watson*, 487 U.S. at 987; *see also Teal*, 457 U.S. at 455–56 (discussed *infra* Section III.B.2). Disparate-impact claims are primarily distinguished by “the *factual* issues that typically dominate”—namely, whether a facially neutral policy is

⁴ In *Cline*, the Supreme Court interpreted the word “age” to take different meanings in different parts of the ADEA. 540 U.S. at 596. The Court distinguished between two alternative definitions: “any number of years lived, or . . . the longer span and concurrent aches that make youth look good.” *Id.* The Court determined that “[t]he presumption of uniform usage thus relents” when comparing § 623(a)(1) and § 623(f). *Id.* at 595. In this case, the presumption holds because § 623(a)(1) and § 623(a)(2) employ “age” in virtually the same context. Both use “age” as in *Cline*’s second definition, to mean “old age.” 540 U.S. at 596.

discriminatory in operation. *Watson*, 487 U.S. at 987 (emphasis added); see *Tex. Dep't of Cmty. Affairs v. Burdine*, 450 U.S. 248, 252 n.5 (1981). A disparate impact “may in operation be functionally equivalent to intentional discrimination.” *Watson*, 487 U.S. at 987; see *Tex. Dep't of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc.*, 135 S. Ct. 2507, 2522 (2015) (“[D]isparate-impact liability . . . plays a role in uncovering discriminatory intent[.]”). Our holding restores the parity described in *Watson*. Under the ADEA, both disparate impact and disparate treatment address the same ultimate legal issue: *age* discrimination.

We conclude that the Supreme Court’s analysis in *O’Connor* answers the question now before us. A specific, facially neutral policy that significantly disfavors employees over fifty years old supports a claim of disparate impact under the plain text of § 623(a)(2). Although the employer’s policy might favor younger members of the forty-and-over cohort, that is an “utterly irrelevant factor,” *O’Connor*, 517 U.S. at 312, in evaluating whether a company’s oldest employees were disproportionately affected because of their age.

2

Our decision is further supported by the Supreme Court’s opinion in *Connecticut v. Teal*, 457 U.S. 440 (1982), a Title VII disparate-impact case. *Teal* confirms that, even under a disparate-impact theory, the plain text

of the statute is designed to protect the rights of individual employees, not the rights of a class.

In *Teal*, a Connecticut state agency used a two-step process to determine eligibility for promotions. First, Connecticut required applicants to take a written test. Second, Connecticut selected the employees for promotion out of the pool of candidates that passed the test. *Id.* at 443. Black applicants who failed the test sued, advancing evidence that black employees failed the written test at a significantly higher rate than white employees. In response, Connecticut argued that, at the second step of the process, the black employees who passed were given preferential treatment through an affirmative action program, counterbalancing the discriminatory effect of the written test. Connecticut argued that its two-step process promoted black employees at an overall higher rate than white employees. *Id.* at 444.

The Supreme Court rejected this so-called “bottom-line” defense and held that the purpose of Title VII “is the protection of the individual employee, rather than the protection of the minority group as a whole.” *Id.* at 453–54. “[F]avorable treatment of . . . members of these respondents’ racial group” did not justify discrimination against other members of the protected class. *Id.* at 454; see *El v. Se. Pa. Transp. Auth.*, 479 F.3d 232, 239–40 (3d Cir. 2007) (“Title VII operates not primarily to the benefit of racial or minority *groups*, but

to ensure that *individual applicants* receive the consideration they are due”).

This case presents a similar issue. The ADEA, like Title VII, protects individuals who are members of a protected class, not a class itself. *See* 29 U.S.C. § 623(a)(1) (proscribing forms of discrimination “because of such *individual’s* age”); *id.* § 623(a)(2) (same); *id.* § 631(a) (limiting the ADEA’s scope to “*individuals* who are at least 40 years of age”). Such protection under the statute does not disappear when a plaintiff advances a disparate-impact claim. *Teal* prohibits the use of a bottom-line statistic to justify ignoring a disproportionate impact against individuals that would otherwise be actionable under the plain text of the statute. That is precisely the problem subgroups are meant to address here.

As a result, *Teal* answers PGW’s argument that employees older than forty were, as a class, favored to keep their jobs. That is equivalent to Connecticut’s argument that black employees were collectively favored for promotions. The Supreme Court rejected that argument in *Teal*, and we reject it here.

Similar to the position of PGW and its *amici* in this case, the dissenting Justices in *Teal* accused the majority of “confus[ing] the distinction—uniformly recognized until today—between disparate *impact* and disparate

treatment.” 457 U.S. at 462 (Powell, J., dissenting). The majority responded as follows:

The fact remains . . . that irrespective of the form taken by the discriminatory practice, an employer’s treatment of other members of the plaintiffs’ group can be of little comfort to the victims of . . . discrimination. Title VII does not permit the victim of a facially discriminatory policy to be told that he has not been wronged because other persons of his or her race or sex were hired. That answer is no more satisfactory when it is given to victims of a policy that is facially neutral but practically discriminatory. Every *individual* employee is protected against both discriminatory treatment and practices that are fair in form, but discriminatory in operation.

Id. at 455–56 (internal citations and quotation marks omitted). The same reasoning applies to this case. The ADEA “does not permit the victim of a facially discriminatory policy to be told that he has not been wronged because other persons” aged forty or older were preferred. *Id.* at 455. “That answer is no more satisfactory when it is given to victims of a policy that is facially neutral but practically discriminatory.” *Id.*

PGW and its *amici* maintain that disparate-impact claims generally rely on comparisons between entire classes. Even in *Teal*, for example, plaintiffs' evidence showed that the written test caused a disparate impact on black employees as a class. 457 U.S. at 443. That general focus on groups, however, is explained by the fact that Title VII protects group identities like race and sex. The trait protected by the ADEA, age, is qualitatively different.

“The term ‘age’ employed by the ADEA is not . . . comparable to the terms ‘race’ or ‘sex’ employed by Title VII.” *Cline*, 540 U.S. at 597. Age is a continuous variable, whereas race and sex are treated categorically in the mine-run of Title VII cases. *See Bienkowski v. Am. Airlines, Inc.*, 851 F.2d 1503, 1506 (5th Cir. 1988) (“The ADEA does not lend itself to a bright-line age rule and in this respect differs from racial or sex discrimination cases”); *Goldstein v. Manhattan Indus., Inc.*, 758 F.2d 1435, 1442 (11th Cir. 1985) (observing that age discrimination is “qualitatively different from race or sex discrimination” because “the basis of the discrimination is not a discre[te] and immutable characteristic of an employee which separates the members of the protected group indelibly from persons outside the protected group”).

On account of that difference, the statistical techniques common in Title VII cases are not perfectly transferable to ADEA cases. If, for example, the comparison group in *Teal* omitted some black employees who took the written test, the statistics would likely have failed to address whether there was a disparate impact “because of . . . race” 42 U.S.C. § 2000e–2(a)(2); *see also id.* § 2000e-2(k)(1)(A)(i). It would be unclear whether the test’s effects fell more harshly on individuals of a particular race without looking at how the test affected all members. But with the ADEA, by contrast, a comparison group that omits employees in their forties is fully capable of demonstrating disparate impact “because of . . . age.” 29 U.S.C. § 623(a)(2).

The forty-and-older line established in § 631(a) does not convert age into a binary trait. By its own terms, it imposes a “limit[ation]” on the “individuals” covered by “[t]he prohibitions in this chapter” 29 U.S.C. § 631(a). It simply establishes “the age at which ADEA protection begins.” *Maxfield v. Sinclair Int’l*, 766 F.2d 788, 792 (3d Cir. 1985). The appropriate disparate-impact statistics should be guided by the trait protected by the statute, not the population of employees inside or outside the statute’s general scope. In fact, when the Supreme Court recognized ADEA disparate-impact liability in *Smith*, nothing in its reasoning turned on the

existence or purpose of § 631(a). That provision was not cited once.⁵

PGW and its *amici* would have us rewrite 29 U.S.C. § 623(a)(2) to proscribe “adverse[] effect[s] . . . because of such individual’s [membership in the 40-and-older class].” That interpretation would bring the ADEA closer to more familiar Title VII territory, but “[w]e have to read it the way Congress wrote it.” *Meacham v. Knolls Atomic Power Lab.*, 554 U.S. 84, 101–02 (2008). The continuous, non-categorical nature of age cannot be adequately addressed by simply aggregating forty-and-older employees. More exacting analysis may be needed in certain cases, and subgroups may answer that need.

4

Finally, our decision is supported by the ADEA’s remedial purpose. Refusing to recognize subgroup claims would deny redress for significantly discriminatory policies that affect employees most in need of the ADEA’s protection.

⁵ To be sure, the plaintiffs in *Smith* happened to rely on forty-and-older statistics. 544 U.S. at 242. Nothing in our opinion should be read to rule out such evidence. Nonetheless, *Smith*’s reasoning does not foreclose subgroup claims.

Mandating a forty-and-older comparison group “would allow an employer to adopt facially neutral policies which had a profoundly disparate impact on individuals over age 50 or 55,” so long as younger individuals within the protected class received sufficiently favorable treatment. *Finch v. Hercules Inc.*, 865 F. Supp. 1104, 1129 (D. Del. 1994). Such policies “reflect the specific type of arbitrary age discrimination Congress sought to prohibit,” but would nonetheless evade judicial scrutiny. *Id.*; see also *Graffam v. Scott Paper Co.*, 848 F. Supp. 1, 4 (D. Me. 1994).

We have also acknowledged in the disparate-treatment context that “[i]f no intra-age group protection were provided by the ADEA, it would be of virtually no use to persons at the upper ages of the protected class” *Maxfield*, 766 F.2d at 792. The same rationale applies to the disparate-impact context. The older the employees affected by a policy, the more confounding favoritism would be included in the rigid forty-and-older sample. Thus, an impact on employees in their seventies may be easier to average out of existence compared to an impact that also affects younger employees. Mandating forty-and-older comparisons would predominantly harm “those most in need of the statute’s protection.” *Lowe v. Commack Union Free Sch. Dist.*, 886 F.2d 1364, 1379 (2d Cir. 1989) (Pierce, J., dissenting in relevant part). “[I]t would indeed be strange, and even perverse, if the youngest members of the protected class were to be

accorded a greater degree of statutory protection than older members of the class.” *Id.*⁶

Accordingly, our interpretation of the ADEA is supported not only by the statute’s text and Supreme Court precedent, but also by the ADEA’s purpose.

C

Our holding in this case is at odds with decisions from three of our sister circuits. *See Lowe v. Commack Union Free Sch. Dist.*, 886 F.2d 1364 (2d Cir. 1989); *Smith v. Tenn. Valley Auth.*, 924 F.2d 1059, 1991 WL 11271 (6th Cir. 1991) (table opinion); *E.E.O.C. v. McDonnell Douglas Corp.*, 191 F.3d 948 (8th Cir.

⁶ Ironically, mandating a forty-and-older sample has the potential to harm employers in certain circumstances. For example, if a substantial disparate impact is experienced only by individuals sixty-five and older, the effect can show up in the forty-and-older aggregate statistic, creating the misimpression that forty-year-old plaintiffs were disparately impacted. *See* Ramona L. Paetzold & Steve L. Willborn, *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases* § 7:2, at 340 (2016–2017 ed. 2016) (noting that “[t]he errors can occur in either direction” when relying on forty-and-older comparisons).

1999).⁷ Those decisions have primarily relied on policy considerations that we do not find persuasive. In short, they are contradicted by *O'Connor* and *Teal*, confuse evidentiary concerns with statutory interpretation, and incorrectly assume that recognizing subgroups will proliferate liability for reasonable employment practices.

⁷ While we are generally reluctant to create circuit splits, we do so where a “compelling basis” exists. *Wagner v. PennWest Farm Credit, ACA*, 109 F.3d 909, 912 (3d Cir. 1997). For the reasons discussed in this opinion, we think a compelling basis exists in this case. Even so, we note that (1) the Second Circuit and Sixth Circuit cases predate the Supreme Court’s decisions in *O'Connor* and *Smith*; (2) the Sixth Circuit case is non-precedential; and (3) the Eighth Circuit case predates *Smith*. One circuit has noted the issue but declined to rule. *See Katz v. Regents of the Univ. of Calif.*, 229 F.3d 831, 835–36 (9th Cir. 2000).

PGW and its *amici* argue that we already decided this question in *Massarsky v. General Motors Corp.*, 706 F.2d 111, 121 (3d Cir. 1983). We did not. The plaintiff in *Massarsky* failed to advance *any* evidence of disparate impact. *Id.* Thus, its contemplation of a specific age group is dicta not binding on this panel. In any event, *Massarsky* predates the Supreme Court’s decisions in both *O'Connor* and *Smith*.

The United States Court of Appeals for the Second Circuit addressed disparate-impact subgroups in *Lowe v. Commack Union Free Sch. Dist.*, 886 F.2d 1364 (2d Cir. 1989). *See also Criley v. Delta Air Lines, Inc.*, 119 F.3d 102, 105 (2d Cir. 1997). Because *Lowe* predates *O'Connor*, it gives improper significance to the forty-and-older line drawn by § 631(a), and fails to compare the textual similarities between § 623(a)(1) and § 623(a)(2). *Lowe* also rejects subgroup claims because specific types of evidence could be misleading. We do not find *Lowe* persuasive.

The Second Circuit's legal analysis begins with the premise that disparate-treatment analysis in Title VII cases "generally has focused . . . on the protected group of which plaintiff is a member." *Lowe*, 886 F.2d at 1373. But *Lowe* does not address the text of § 623(a)(2). Divorced from that text, the Second Circuit allows the "general[] . . . focus[]" of a different statute to limit what *this* statute plainly permits. *Lowe* does not, and cannot, explain why forty-and-older group membership is "utterly irrelevant" to discrimination based on age,

O'Connor, 517 U.S. at 312, but is the *sine qua non* of an adverse effect based on age.⁸

Lowe is primarily concerned with the practical implications of subgroup claims.⁹ Its main objection is

⁸ *Lowe's* treatment of *Teal* is similarly unpersuasive. See *Lowe*, 886 F.2d at 1374. We addressed the same argument in Section III.B.3, *supra*.

⁹ For example, *Lowe* offers a flawed *reductio ad absurdum*: “an 85 year old plaintiff could seek to prove a discrimination claim by showing that a hiring practice caused a disparate impact on the ‘sub-group’ of those age 85 and above, even though all those hired were in their late seventies.” *Lowe*, 886 F.2d at 1373. This argument relies on the false assumption that an employer who favors 70-year-old employees could not possibly be liable under the ADEA. “If an 85-year old person . . . fails to attain that position for no reason other than age, s/he has suffered age discrimination under the Act” *Id.* at 1380 (Pierce, J., dissenting in relevant part). In any event, we can be reasonably assured that such a hypothetical would never arise due to the demographic characteristics of the workforce, which limit the statistical power to compare impacts on seventy- and eighty-year-old employees. See Sandra F. Sperino, *The Sky Remains Intact: Why Allowing Subgroup Evidence Is Consistent with the Age Discrimination in Employment Act*, 90 Marq. L. Rev. 227, 263 (2006).

evidentiary: “any plaintiff can take his or her own age as the lower end of a ‘sub-protected group’ and argue that said ‘sub-group’ is disparately impacted.” *Lowe*, 886 F.2d at 1373. Here, PGW and its *amici* similarly argue that plaintiffs will be able to “gerrymander” arbitrary age groups in order to manufacture a statistically significant effect. We disagree.

Essentially, PGW and its *amici* argue that a particular form of evidence carries such a high risk of manipulation that we should interpret the ADEA to preclude the entire claim. That is a thoroughly unsatisfactory justification for ignoring statutory text and Supreme Court precedent.¹⁰ Our interpretation of the ADEA is based on text, not evidentiary gatekeeping. That function is capably performed by district judges who routinely apply the Federal Rules of Evidence and *Daubert* jurisprudence. We consider that to be a sufficient safeguard against the menace of unscientific methods and manipulative statistics.

¹⁰ The Eighth Circuit, which ultimately agreed with *Lowe*’s outcome, did criticize the Second Circuit’s reasoning on this point. *See E.E.O.C. v. McDonnell Douglas Corp.*, 191 F.3d 948, 950–51 (8th Cir. 1999) (“The fact that a particular interpretation of a statute might spawn lawsuits is not a reason to reject that interpretation.”).

Preliminarily, PGW’s “gerrymandering” objection only applies to the kind of statistical studies that compare subgroups selected by an expert. Some scholars have proposed the use of statistical models that treat age as a continuous variable and thus avoid the need to draw “arbitrary” age groups. Options discussed in the literature include proportional hazards models and logistic regression. See Ramona L. Paetzold & Steve L. Willborn, *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases* § 7:11, at 372 (2016–2017 ed. 2016) [hereinafter Paetzold & Willborn]; see also, e.g., George Woodworth & Joseph Kadane, *Age- and Time-Varying Proportional Hazards Models for Employment Discrimination*, 4 *Annals Applied Statistics* 1139 (2010); Michael O. Finkelstein & Bruce Levin, *Proportional Hazard Models for Age Discrimination Cases*, 34 *Jurimetrics J.* 153 (1994).

We have no need today to bless any one approach. “Statistics ‘come in infinite variety and . . . their usefulness depends on all of the surrounding facts and circumstances.’” *Watson*, 487 U.S. at 995 n.3 (quoting *Teamsters v. United States*, 431 U.S. 324, 340 (1977)). Our purpose is rather to demonstrate that the gerrymandering objection exposes a weakness in one particular research method, not a cause of action. “The continuous nature of the age variable need not be a statistical problem under disparate-impact analysis; existing statistical procedures can be adapted to the

specific needs of disparate-impact analysis.” Paetzold & Willborn § 7:11, at 373.

Even if the statistical evidence in an ADEA case uses age groups selected by the expert, PGW and its *amici* overstate the risk of manipulation. “The claim can be analyzed, of course, to determine if the result is robust across various age breaks and whether the age breaks can be justified independently of the data” *Id.* § 7:3, at 344. In fact, some courts have long permitted statistical subgroup evidence in the context of disparate-treatment claims. *See, e.g., Barnes v. GenCorp Inc.*, 896 F.2d 1457, 1466–67 (6th Cir. 1990). We see no reason why that same evidence would be any less workable in a disparate-impact case. *See MacNamara v. Korean Air Lines*, 863 F.2d 1135, 1148 (3d Cir. 1988) (“[T]he statistical evidence supporting a claim of disparate impact often resembles that used to help establish disparate treatment.”).

So-called “age-break” analysis has well-understood limitations. *See Paetzold & Willborn* § 7:3, at 343–46. For example, if an expert does not devise the age breaks independently of the data, and instead cherry-picks groups to manufacture a particular result, that “may invalidate the usual tests of statistical significance.” *Id.* at 341. In addition, “the appropriate inference for plaintiffs near a selected age break is always likely to be problematic.” *Id.* at 344–45. Without more, this challenge may undermine the claims of plaintiffs who “take [their]

own age as the lower end of a ‘sub-protected group’ and argue that said ‘sub-group’ is disparately impacted.” *Lowe*, 886 F.2d at 1373; *see also Finch*, 865 F. Supp. at 1129–30 (“If a plaintiff attempts to define the subset too narrowly, he or she will not be able to obtain reliable statistics upon which to prove a prima facie case.”).

The EEOC and plaintiffs have only argued in favor of subgroups with “lower boundaries,” not “upper boundaries.” Oral Arg. Tr. 23:24–24:3. That rule would preclude, for example, a “banded” 50-to-55 subgroup. We think that limitation is well founded. A plaintiff would benefit from introducing an upper boundary if a policy favored employees older than that limit. But in *Cline*, the Supreme Court interpreted the term “age” in § 623(a)(1) to mean “old age.” 540 U.S. at 596. Under *Cline*, the ADEA protects only “relatively old worker[s] from discrimination that works to the advantage of the relatively young.” *Id.* at 590–91. If a facially neutral policy systematically favors a company’s oldest employees,¹¹ that fact may be fatal to a claim that

¹¹ We note that the ability to draw inferences about the treatment of a company’s oldest employees may be limited by sample size. In this case, for example, plaintiffs’ expert argues that there is no statistically significant effect on employees age sixty and older because “[t]here are only 14 terminations, which means the statistical power to detect a significant effect is very low.” A.244–45.

members of a younger subgroup were disparately impacted because of their “old age.” *Id.* at 596; *see Tex. Dep’t of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc.*, 135 S. Ct. 2507, 2523 (2015) (describing the importance of “[a] robust causality requirement” in disparate-impact cases); *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2009). Thus, a banded subgroup would be self-defeating under *Cline*, further limiting plaintiffs’ ability to gerrymander age groups.

We reject the notion that the risk of gerrymandered evidence is so great that it can override what the text of the statute otherwise permits. District courts should, as in any other case, ensure that plaintiffs’ evidence is reliable under *Daubert* and provides more than the “mere scintilla of evidence” needed to survive summary judgment. *S.H. ex rel. Durrell v. Lower Merion Sch. Dist.*, 729 F.3d 248, 256 (3d Cir. 2013) (internal quotation marks omitted) (quoting *Jakimas v. Hoffman-La Roche, Inc.*, 485 F.3d 770, 777 (3d Cir. 2007)). Accordingly, we are not persuaded by *Lowe*’s legal or practical groundings.

2

The United States Court of Appeals for the Sixth Circuit addressed disparate-impact subgroups in a non-precedential opinion, *Smith v. Tennessee Valley Authority*, 924 F.2d 1059, 1991 WL 11271 (6th Cir. 1991) (table opinion). This decision also predates *O’Connor*. Its reasoning contradicts both *O’Connor* and

Teal, and conflicts with a precedential Sixth Circuit opinion that allows subgroup analysis in disparate-treatment cases.

In *Smith*, the Sixth Circuit asserts by citation to *Lowe* that “[a] plaintiff cannot succeed under a disparate impact theory by showing that younger members of the protected class were preferred over older members of the protected class.” *Id.* at *4. As we have discussed, *Lowe*’s reasoning is explicitly rejected by *O’Connor* and *Teal*. *Teal* held that a plaintiff *can* succeed under a disparate-impact theory if other members of the protected class were preferred, 457 U.S. at 454, and *O’Connor* held that forty-and-older status is irrelevant to evaluating the application of a protection based on “age,” 517 U.S. at 312.

As we have also noted, the Sixth Circuit has long recognized statistical subgroup evidence in disparate-treatment claims. In a precedential opinion, *Barnes v. GenCorp Inc.*, 896 F.2d 1457 (6th Cir. 1990), the Sixth Circuit specifically rejected the defendant’s argument that “the only valid statistics would necessarily divide the employees into groups age 40-and-over and those under 40.” *Id.* at 1466. The Sixth Circuit suggests in a footnote, by citation to *Lowe*, that “[s]uch sub-group analysis may not apply to discriminatory impact cases[.]” *Id.* at 1467 n.12. With the exception of that speculative footnote, we find the Sixth Circuit’s decision in *Barnes* more persuasive than its decision in *Smith*.

Finally, the United States Court of Appeals for the Eighth Circuit addressed disparate-impact subgroups in *E.E.O.C. v. McDonnell Douglas Corp.*, 191 F.3d 948 (8th Cir. 1999). The Eighth Circuit's analysis is also unpersuasive because it contradicts *Teal* and ignores important limitations on the scope of disparate-impact claims.

First, the Eighth Circuit argued that if subgroup claims were cognizable,

a plaintiff could bring a disparate-impact claim despite the fact that the statistical evidence indicated that an employer's RIF criteria had a very favorable impact upon the entire protected group of employees aged 40 and older, compared to those employees outside the protected group. We do not believe that Congress could have intended such a result.

Id. at 951. This is no more than an endorsement of the bottom-line defense that the Supreme Court rejected in *Teal*. The State of Connecticut tried a similar argument by suggesting that black employees were favored for promotions as an overall class. But that bottom-line outcome concealed individual rights violations. Far from being a result "Congress could [not] have intended," *id.*,

the Supreme Court’s ruling in *Teal* vindicated Title VII’s plain text and purpose. The same applies to the ADEA.

Second, the Eighth Circuit panel wrote:

[T]he consequence would be to require an employer engaging in a RIF to attempt what might well be impossible: to achieve statistical parity among the virtually infinite number of age subgroups in its work force. Adoption of such a theory, moreover, might well have the anomalous result of forcing employers to take age into account in making layoff decisions, which is the very sort of age-based decision-making that the statute proscribes.

McDonnell Douglas Corp., 191 F.3d at 951.¹²

Even without the prospect of subgroups, it has always been the case that “a completely neutral practice will inevitably have some disproportionate impact on one group or another.” *DiBiase v. SmithKline Beecham Corp.*, 48 F.3d 719, 731–32 (3d Cir. 1995) (quoting *City of L.A., Dep’t of Water & Power v. Manhart*, 435 U.S.

¹² PGW’s *amici* make an opposite argument—that employers *already* fine-tune employment decisions to avoid creating a disparate impact, and our decision will make it more costly or complicated for them to do so.

702, 710 n.20 (1978)). That is precisely why deviating from statistical parity is not, by itself, enough to incur disparate-impact liability. Just last Term, the Supreme Court recognized that “disparate-impact liability has always been properly limited in key respects” so that it is not “imposed based solely on a showing of a statistical disparity.” *Inclusive Cmty. Project, Inc.*, 135 S. Ct. at 2512; *see also Wards Cove*, 490 U.S. at 657 (showing a statistical disparity alone “will not suffice to make out a prima facie case of disparate impact”); *Watson*, 487 U.S. at 994 (plurality opinion) (“[P]laintiff’s burden in establishing a prima facie case goes beyond the need to show that there are statistical disparities in the employer’s work force.”).

To make out a *prima facie* case, plaintiffs must first identify a specific employment practice that causes the disparity. *See Wards Cove*, 490 U.S. 642; *see also Smith*, 544 U.S. at 241 (noting that the *Wards Cove* holding remains in effect under the ADEA). The Supreme Court has recognized that this requirement guards against “the myriad of innocent causes that may lead to statistical imbalances.” *Smith*, 544 U.S. at 241 (quoting *Wards Cove*, 490 U.S. at 657). “Identifying a specific practice is not a trivial burden” *Meacham v. Knolls Atomic Power Lab.*, 554 U.S. 84, 101 (2008); *see Inclusive Communities Project*, 135 S. Ct. at 2523 (“[A] disparate-impact claim that relies on a statistical disparity

must fail if the plaintiff cannot point to a defendant's policy or policies causing that disparity.”).

Furthermore, not just any disparity will make out the *prima facie* case; the disparity must be significant. See *Watson*, 487 U.S. at 995 (“[S]tatistical disparities must be sufficiently substantial that they raise such an inference of causation.”); *Teal*, 457 U.S. at 446 (“[T]he facially neutral employment practice [must have] had a significantly discriminatory impact.”); *Wards Cove*, 490 U.S. at 657 (requiring a “significantly disparate impact”); *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307–08 (1977) (requiring “gross statistical disparities”). We have not adopted a uniform rule for what this requirement entails; it must be evaluated “on a case-by-case basis.” *Watson*, 487 U.S. at 995 n.3.

Finally, even if plaintiffs make out a *prima facie* case, the RFOA defense imposes a relatively light burden on employers. See *Smith*, 544 U.S. at 243. If a company's oldest employees are inadvertently disadvantaged by a merit-based policy, for example, the RFOA defense is designed to address just such a scenario. See *id.* at 229 (observing that Congress included the RFOA defense because age “not uncommonly has relevance to an individual's capacity to engage in certain types of employment”). But if an employer can provide no reasonable justification for a policy that creates a significant age-based disparity, the ADEA prohibits that policy.

In sum, the limitations applicable to any ADEA disparate-impact claim preclude liability for reasonable employment practices, regardless of subgroups. Nonetheless, as *amici* argue, our decision may very well require employers to be more vigilant about the effects of their employment practices. “But at the end of the day, *amici*’s concerns have to be directed at Congress, which set the balance where it is We have to read it the way Congress wrote it.” *Meacham*, 554 U.S. at 101–02; see *Watson*, 487 U.S. at 993–99 (plurality opinion) (explaining why “disparate impact theory need [not] have any chilling effect on legitimate business practices”).

* * *

We conclude that ADEA disparate-impact claims are not limited to forty-and-older comparisons. While claims based on subgroups present unique challenges, the limitations applicable to any other disparate-impact case—evidentiary gatekeeping, the *prima facie* case, and affirmative defenses—are adequate safeguards. Accordingly, we will reverse the District Court’s determination that PGW is entitled to summary judgment on this ground.

IV

We now address the District Court’s second ground for granting summary judgment in favor of PGW: the exclusion of plaintiffs’ statistics expert under *Daubert*

and Rule 702 of the Federal Rules of Evidence. For the reasons that follow, we will vacate and remand for further *Daubert* proceedings regarding plaintiffs' statistical evidence. We then turn to plaintiffs' other expert reports, concluding that the District Court did not err in excluding each.

A

Pursuant to *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), district courts perform a gatekeeping function to ensure that expert testimony meets the requirements of Federal Rule of Evidence 702. That function extends not only to scientific testimony, but also to other forms of “technical” or “specialized” knowledge. Fed. R. Evid. 702(a); *Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 141 (1999). “Rule 702 embodies three distinct substantive restrictions on the admission of expert testimony: qualifications, reliability, and fit.” *Elcock v. Kmart Corp.*, 233 F.3d 734, 741 (3d Cir. 2000). This case presents issues of reliability and fit.

“In order for expert testimony to meet *Daubert*'s reliability standard, it must be based on the methods and procedures of science, not on subjective belief and unsupported speculation.” *In re TMI Litig.*, 193 F.3d 613, 703–04 (3d Cir. 1999), *amended*, 199 F.3d 158 (3d Cir. 2000). “The test of admissibility is not whether a particular scientific opinion has the best foundation, or

even whether the opinion is supported by the best methodology or unassailable research.” *Id.* at 665. Instead, the court looks to whether the expert’s testimony is supported by “good grounds.” *Id.* at 665 (quoting *In re Paoli R.R. Yard PCB Litig. (Paoli II)*, 35 F.3d 717, 745 (3d Cir. 1994)). The standard for reliability is “not that high.” *Id.* It is “lower than the merits standard of correctness.” *Id.* Each aspect of the expert’s opinion “must be evaluated practically and flexibly without bright-line exclusionary (or inclusionary) rules.” *ZF Meritor, LLC v. Eaton Corp.*, 696 F.3d 254, 291 (3d Cir. 2012) (quoting *Heller v. Shaw Indus., Inc.*, 167 F.3d 146, 155 (3d Cir. 1999)).

The “fit” requirement ensures that the evidence or testimony “[helps] the trier of fact to understand the evidence or to determine a fact in issue.” *TMI*, 193 F.3d at 663 (quoting Fed. R. Evid. 702(a)). “This condition goes primarily to relevance.” *Id.* (quoting *Daubert*, 509 U.S. at 591).

“We review a district court’s decision to admit expert testimony for abuse of discretion and exercise plenary review over a district court’s legal interpretation of Rule 702 of the Federal Rules of Evidence.” *United States v. Walker*, 657 F.3d 160, 175–76 (3d Cir. 2011).

B

Plaintiffs' expert, Dr. Michael Campion, proposes to offer statistical evidence in support of the disparate-impact claims. Specifically, Dr. Campion would testify that employees older than forty-five, fifty, and fifty-five years old were likelier to be fired in the March 2009 RIF than were younger employees.

The District Court identified three grounds¹³ for exclusion: (1) Dr. Campion used facts or data that were not reliable; (2) he failed to use a statistical adjustment called the Bonferroni procedure; and (3) his testimony lacks "fit" to the case because subgroup claims are not cognizable. For the reasons that follow, we will vacate the District Court's order and remand for further *Daubert* proceedings.¹⁴

¹³ The District Court assumed that Dr. Campion is qualified. We need not, then, address that issue.

¹⁴ Dr. Campion based his opinion on two analyses: one using the EEOC's "four-fifths" test, *see* 29 C.F.R. § 1607.4 (1987), and another using a more traditional statistical method, a z-score test. The District Court noted that the four-fifths test "has been criticized," but may be "used in conjunction" with other statistical evidence. *Karlo*, 2015 WL 4232600, at *13 (citation omitted). That determination is not disputed on appeal. We therefore focus on the reliability of Dr. Campion's z-score test.

First, the District Court concluded that Dr. Campion's report should be excluded because it is not based on reliable data, contrary to Federal Rule of Evidence 702(b). Specifically, Dr. Campion's dataset included certain "Evert Terminees" who were not part of the "Agreed Data Set" to which the parties stipulated. We conclude that the District Court abused its discretion to the extent that it excluded Dr. Campion's testimony on this basis because the District Court ignored, without explanation, Dr. Campion's subsequent analysis.

Plaintiffs argue that Dr. Campion cured this deficiency, and the District Court's opinion provides no reason to doubt their argument. Specifically, plaintiffs claim that Dr. Campion excluded the Evert Terminees and determined that it did not affect his conclusions. At oral argument, plaintiffs explained that the Evert Terminees "skewed the data actually in favor of more of the defendants," Oral Arg. Tr. 5:6–7, whereas PGW insists that the Evert Terminees "skewed the data to favor Plaintiffs' theory of the case." Br. Appellee 29.

It is appropriate for the District Court to address this issue in the first instance. But the District Court noted plaintiffs' counterargument without addressing it. To the extent that the District Court excluded Dr. Campion's testimony based on problems that were cured by subsequent analysis, it abused its discretion. To the

extent that the subsequent analysis was deficient, the District Court also abused its discretion because it failed to provide any justification for discrediting that analysis. Because we will remand for further *Daubert* proceedings, as described below, the District Court will have the opportunity to revisit this issue.

2

Next, the District Court determined that Dr. Campion “does not apply any of the generally accepted statistical procedures (*i.e.*, the Bonferroni procedure) to correct his results for the likelihood of a false indication of significance. This sort of subgrouping ‘analysis’ is data-snooping, plain and simple.” *Karlo*, 2015 WL 4232600, at *13. We conclude that the District Court applied an incorrectly rigorous standard for reliability.

The Bonferroni procedure makes it more difficult to find statistical significance where a researcher tests multiple comparisons using the same data. In theory, a researcher who searches for statistical significance in multiple attempts raises the probability of discovering it purely by chance, committing Type I error (*i.e.*, finding a false positive). *See Ballew v. Georgia*, 435 U.S. 223, 234 (1978) (describing Type I and Type II errors). The Bonferroni procedure adjusts for that risk by dividing the “critical” significance level by the number of comparisons tested. In this case, PGW’s rebuttal expert, Dr. James L. Rosenberger, argues that the critical

significance level should be $p < 0.01$, rather than the typical $p < 0.05$, because Dr. Champion tested five age groups ($0.05 / 5 = 0.01$).¹⁵ Once the Bonferroni adjustment is applied, Dr. Champion's results are not statistically significant. Thus, Dr. Rosenberger argues that Dr. Champion cannot reject the null hypothesis and report evidence of disparate impact.¹⁶

Dr. Champion responds that adjusting the required significance level is generally required in a “data snooping” scenario where a researcher conducts “a huge number of analyses of all possibilities to try to find something significant.” A.239. In contrast to “data snooping,” Dr. Champion calls his methodology “hypothesis driven”; he evaluates the likelihood of termination on a small number of groups based on logical increments in age to discover “evidence that increasing age relates to increased likelihood of termination”

¹⁵ Dr. Champion notes that he tested only four age groups, not five. Dr. Rosenberger tests a subgroup of sixty-and-older employees, which Dr. Champion did not include in his analysis because “[t]here are only 14 terminations, which means the statistical power to detect a significant effect is very low.” A.244–45.

¹⁶ The relationship between statistical significance and admissibility is currently before this Court in *In re Zolof (Sertralinehydrochloride) Prod. Liab. Litig.*, No. 16-2247, *appealing* 176 F. Supp. 3d 483 (E.D. Pa. 2016).

A.240. He also points out that “nearly all the tests are significant,” which makes his method analogous to “cross-validating the relationship between age and termination at different cut-offs,” or “replication with different samples.” A.241. And finally, Dr. Campion includes supplemental results that he claims “control for the error rate by conducting only one analysis” A.242.

We conclude that the District Court erred by applying a “merits standard of correctness,” a higher bar than what Rule 702 demands. *TMI*, 193 F.3d at 665. After identifying a potential methodological flaw, the District Court did not proceed to evaluate whether Dr. Campion’s opinion nonetheless rests on good grounds. Instead, it applied a “bright-line exclusionary . . . rule[],” *ZF Meritor*, 696 F.3d at 291, based on Dr. Campion’s failure to perform a specific arithmetical adjustment. As we have observed, there could be good grounds for an expert’s conclusion “even if the judge thinks that . . . a scientist’s methodology has some flaws such that if they had been corrected, the scientist would have reached a different result.” *Paoli II*, 35 F.3d at 744.

In certain cases, failure to perform a statistical adjustment may simply diminish the weight of an expert’s finding. *See* Paetzold & Willborn § 6:7, at 308 n.2 (describing the Bonferroni adjustment as “good statistical practice,” but “not widely or consistently adopted” in the behavioral and social sciences); *E.E.O.C.*

v. Autozone, Inc., No. 00-2923, 2006 WL 2524093, at *4 (W.D. Tenn. Aug. 29, 2006) (“[T]he Court does not have a sufficient basis to find that . . . the non-utilization [of the Bonferroni adjustment] makes [the expert’s] results unreliable.”). The question of whether a study’s results were properly calculated or interpreted ordinarily goes to the weight of the evidence, not to its admissibility. See *Leonard v. Stemtech Int’l Inc.*, 834 F.3d 376, 391 (3d Cir. 2016). “Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.” *Daubert*, 509 U.S. at 596; cf. *Bazemore v. Friday*, 478 U.S. 385, 400 (1986) (“Normally, failure to include variables will affect the analysis’ probativeness, not its admissibility.”).

“That is not to say that a significant error in application will never go to the admissibility, as opposed to the weight, of the evidence.” *In re Scrap Metal Antitrust Litig.*, 527 F.3d 517, 530 (6th Cir. 2008). An expert’s failure to use a statistical adjustment may, in certain cases, present a “flaw . . . large enough that the expert lacks ‘good grounds’ for his or her conclusions.” *Paoli II*, 35 F.3d at 746; see *Erica P. John Fund, Inc. v. Halliburton Co.*, 309 F.R.D. 251, 266–67 (N.D. Tex. 2015) (applying a less conservative adjustment, Holm-Bonferroni, based on “the substantial number of comparisons” made by the expert, and citing an article explaining that the risk of finding false significance is

prevalent where at least twenty to forty comparisons are tested). Nonetheless, “[t]he grounds for the expert’s opinion merely have to be good, they do not have to be perfect.” *Paoli II.*, 35 F.3d at 744. “So long as the expert’s testimony rests upon ‘good grounds,’ it should be tested by the adversary process . . . rather than excluded from jurors[’] scrutiny for fear that they will not grasp its complexities or satisfactory [sic] weigh its inadequacies.” *TMI*, 193 F.3d at 692 (quoting *Ruiz–Troche v. Pepsi Cola of P. R. Bottling Co.*, 161 F.3d 77, 85 (1st Cir. 1998)). Accordingly, we will remand for further *Daubert* proceedings as to Dr. Champion’s statistics-related testimony to allow the District Court to apply the correct standard for reliability.

3

Finally, the District Court determined that Dr. Champion’s statistics lacked fit to the case. “[T]he subgrouping analysis would only be helpful to the factfinder if this Court held that Plaintiffs could maintain an over-fifty disparate impact claim.” *Karlo*, 2015 WL 4232600, at *13 n.16. Having held that plaintiffs’ over-fifty disparate-impact claim is cognizable, we conclude that this ground for exclusion fails as well. Because each ground fails, we will vacate the District Court’s order

excluding Dr. Campion's statistics-related testimony and remand for further *Daubert* proceedings.¹⁷

C

Dr. Campion offered a second expert report on a different subject: reasonable human-resources ("HR") practices. We conclude that the District Court did not abuse its discretion in excluding this testimony.

Dr. Campion intended to testify as to twenty "reasonable" HR practices that PGW could have, but did not, employ when conducting its RIFs. Plaintiffs aver that this testimony is necessary to rebut PGW's RFOA defense. The District Court disagreed. It concluded that Dr. Campion's HR testimony lacked relevance to the case because "plaintiffs c[ould] rebut Defendants' RFOA defense only by demonstrating that the factors offered by Defendants [we]re unreasonable." *Karlo*, 2015 WL 4232600, at *15 (quoting *Powell v. Dallas Morning News L.P.*, 776 F. Supp. 2d 240, 247 (N.D. Tex. 2011), *aff'd* 486 F. App'x 469 (5th Cir. 2012)) (alterations in original).

We agree. When a defendant proffers a RFOA, the plaintiff can rebut it by showing that *the factor relied*

¹⁷ We do not reach the issue of whether the District Court abused its discretion by declining to hold a *Daubert* hearing.

upon is unreasonable, not by identifying twenty *other* practices that would have been reasonable instead. *See Smith*, 544 U.S. at 243 (“While there may have been other reasonable ways for the City to achieve its goals, the one selected was not unreasonable.”).

Plaintiffs also argue that PGW’s proffered RFOA fails as a matter of law. If true, that would eliminate the need for Dr. Champion’s HR testimony under plaintiffs’ own explanation for its relevance. But because the District Court did not grant summary judgment on the basis of PGW’s RFOA defense, the question of that defense’s legal sufficiency is not before us.

D

Finally, the District Court excluded the testimony of Dr. Anthony G. Greenwald. Dr. Greenwald proposed to testify as to his experience with Implicit Association Tests (IAT), a type of test designed to measure “the strength of a mental association that links a social category (such as race, gender, or age group) with a trait (i.e., a stereotype)” A.405. Specifically, Dr. Greenwald reports that 80% of research participants hold an implicit bias based on age. He also evaluated the deposition transcripts of certain PGW employees and determined that their RIF procedures were susceptible to implicit biases.

The District Court concluded that Dr. Greenwald’s testimony lacks fit to this case because his population-wide statistics have only speculative application to PGW and its decision-makers. The District Court also observed that disparate-impact claims do not inquire into the employer’s state of mind. We agree. Plaintiffs are not required to prove that any particular psychological mechanism caused the disparity in question; they are only required to demonstrate that the disparity itself is “sufficiently substantial that [it] raise[s] such an inference of causation.” *Watson*, 487 U.S. at 995. That is not to say, however, that implicit-bias testimony is never admissible. Courts may, in their discretion, determine that such testimony elucidates the kind of headwind disparate-impact liability is meant to redress. We are simply unable here to conclude that the District Court abused its discretion in excluding this evidence.

V

The final issue presented in this appeal is whether the District Court committed clear error in decertifying the collective action.¹⁸ We hold that it did not.

¹⁸ Defendants argue that we should not reach this issue for two reasons. First, they argue that the four opt-in plaintiffs were not named in the notice of appeal. We have rejected that argument in Section II of this opinion. Second, they argue that the Notice of Appeal failed to

The collective action¹⁹ “is a form of group litigation in which a named employee plaintiff or plaintiffs file a complaint ‘in behalf of’ a group of other, initially unnamed employees who purport to be ‘similarly situated’ to the named plaintiff.” *Halle v. W. Penn Allegheny Health Sys. Inc.*, 842 F.3d 215, 223 (3d Cir. 2016). Courts in this circuit use a two-step certification

specify that plaintiffs sought review of the decertification order. We reject that argument as well. We exercise jurisdiction over orders not specified in a notice of appeal if: “[1] there is a connection between the specified and unspecified order, [2] the intention to appeal the unspecified order is apparent and [3] the opposing party is not prejudiced and has a full opportunity to brief the issues.” *Lusardi v. Xerox Corp.*, 975 F.2d 964, 972 (3d Cir. 1992). Each prong is met. The District Court’s 54(b) memorandum stated that “[T]he same reasons that warrant the certification of a final judgment on the disparate impact and disparate treatment claims also fully justify an immediate appeal on the decertification ruling.” *Karlo*, 2015 WL 5782062, at *4 n.2. Thus, defendants had full notice and opportunity to brief the issue.

¹⁹ “[T]he ADEA incorporates enforcement provisions of the [Fair Labor Standards Act], including the collective action provisions of 29 U.S.C. § 216(b).” *Halle v. W. Penn Allegheny Health Sys. Inc.*, 842 F.3d 215, 224 n.8 (3d Cir. 2016).

process. The first step, so-called conditional certification, requires the named plaintiffs to make a “modest factual showing” to demonstrate “a factual nexus between the manner in which the employer’s alleged policy affected him or her and the manner in which it affected the proposed collective action members.” *Id.* at 224.

The second step, final certification, is what is at issue here. “[T]he named plaintiffs bear the burden of showing that the opt-in plaintiffs are ‘similarly situated’ to them for FLSA purposes.” *Id.* at 226. “Being ‘similarly situated’ . . . means that one is subjected to some common employer practice that, if proved, would help demonstrate a violation of the FLSA.” *Id.* (quoting *Zavala v. Wal Mart Stores Inc.*, 691 F.3d 527, 538 (3d Cir. 2012)). In determining whether plaintiffs are similarly situated, relevant factors include:

whether the plaintiffs are *employed in the same corporate department, division, and location*; whether they advance similar claims; whether they seek substantially the same form of relief; and whether they have similar salaries and *circumstances of employment*. Plaintiffs may also be found dissimilar based on the *existence of individualized defenses*.

Zavala, 691 F.3d at 536–37 (emphases added). A district court’s determination as to whether plaintiffs are

similarly situated is a finding of fact that we review for clear error. *Zavala*, 691 F.3d at 535.

In this case, the District Court properly relied on *Zavala* in determining that plaintiffs did not meet their burden to show that they are similarly situated. Specifically, the District Court observed that the nine plaintiffs “held seven different titles with varied job duties in two separate divisions of PGW and across five locations in which no less than six decision-makers independently included them in the RIF.” *Karlo*, 2014 WL 1317595, at *18. The District Court also based its opinion on “[t]he existence of individualized defenses and procedural concerns” *Id.* at *19. Those considerations fall squarely within the factors listed in *Zavala*.

To be sure, the named plaintiffs and opt-in plaintiffs were each terminated in a single RIF that left full discretion in the hands of local managers. But the District Court did not clearly err when it concluded that “[t]he similarities among the proposed plaintiffs are too few, and the differences among the proposed plaintiffs are too many” for the case to proceed as a collective action. *Zavala*, 691 F.3d at 537–38. Such differences may undermine the “efficiencies for the judicial system through resolution in one proceeding of common issues” *Halle*, 842 F.3d at 223; *see Zavala*, 691 F.3d at 538 (“[T]hese common links are of minimal utility in streamlining resolution of these cases.”).

Plaintiffs essentially concede this point, but argue that the “small class size” makes the class “easily manageable even with the presence of potentially individualized defenses and damages evidence.” Br. Appellant 34. We decline to read the statutory phrase “similarly situated” differently depending on the size of the collective action.

Plaintiffs are correct that the existence of separate defenses or damage calculations “does not vitiate *automatically*” the collective action. *Lockhart v. Westinghouse Credit Corp.*, 879 F.2d 43, 52 (3d Cir. 1989) (emphasis added); *cf. Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1045 (2016) (quoting 7AA Charles Alan Wright, Arthur Miller & Mary Kay Kane, *Federal Practice and Procedure* §1778, at 123–24 (3d ed. 2005)). But under the guidance we have provided in *Zavala*, a district court may determine that such differences are too pronounced for the case to proceed as a collective action. Under our deferential standard of review, we are simply unable to conclude that the District Court committed clear error.

VI

We conclude that plaintiffs’ disparate-impact claims are cognizable under the ADEA. We will therefore vacate the District Court’s orders granting summary judgment in favor of PGW and excluding the statistics-related testimony of Dr. Champion. We will

remand for further *Daubert* proceedings consistent with this opinion. We will affirm the District Court in all other respects.