

USE OF STATISTICAL EVIDENCE IN COMPLEX WAGE LITIGATION

Adam T. Klein
Tarik F. Ajami
Outten & Golden LLP
3 Park Avenue 29th Floor
New York, New York 10016

Statistical evidence has taken its place as a core class of evidence in complex employment cases. Yet despite the centrality of statistics in the field, there is precious little caselaw addressing the use of expert statistical evidence in complex wage-and-hour litigation. However, as a growing number of practitioners have recognized, class and collective wage-and-hour litigation is as well-suited or better-suited for the use of statistical experts as are Title VII and other employment actions,

Statistical Evidence in Complex Employment Litigation

A report from an expert in statistical methods has become perhaps the *sine qua non* of modern pattern-or-practice complex employment discrimination litigation. *Robinson v. Metro-North Commuter R.R.*, 267 F.3d 147, 158 n.5 (2d Cir. 2001) (“The heavy reliance on statistical evidence in a pattern-or-practice disparate treatment claim distinguishes such a claim from an individual disparate treatment claim proceeding under the *McDonnell Douglas* framework.”)

(citing *Bell v. EPA*, 232 F.3d 546, 553 (7th Cir. 2000) ("In a pattern and practice disparate treatment case, statistical evidence constitutes the core of a plaintiff's prima facie case.")). Today, class litigation of discrimination claims turns heavily if not primarily on reports from statistical experts, especially during the class certification phase of the case. *See, e.g., Dukes v. Wal-Mart*, 222 F.R.D. 137 (N.D. Cal. 2004) (court devotes the bulk of the opinion resolving claims from competing expert statisticians).

As such, it is no surprise that courts have increasingly become expert themselves in reviewing and relying upon complex statistical analyses utilizing sophisticated techniques. Indeed, many courts now reject merely descriptive or otherwise simple analyses that show disparities suggestive of discriminatory bias, insisting on more sophisticated techniques such as regression analyses that can isolate and quantify the effect of multiple explanatory factors for observed phenomena, such as terminations or pay disparities. *See, e.g., Smith v. Xerox Corp.*, 196 F.3d 358, 363 (2^d Cir. 1999) (disregarding plaintiffs' statistical analyses and granting summary judgment against plaintiffs, in part, for failure to utilize multiple regression techniques); *Plair v. E.J. Brach & Sons, Inc.*, 105 F.3d 343, (7th Cir. 1997) (rejecting raw descriptive comparison of termination rates); *Doan v. Seagate Tech., Inc.*, 82 F.3d 974 (10th Cir. 1996) (raw descriptive comparison of eliminated positions to hires rejected); *Rea v. Martin Marietta Corp.*, 29 F.3d 1450 (10th Cir. 1994) (analysis disregarded due to failure to control for other explanatory variables); *Maniatis v. New York Hosp.-Cornell Med. Ctr.*, 58 F. Supp.

2d 221 (S.D.N.Y. 1999) (“Fisher’s exact” test cannot account for a range of possible explanatory variables and is insufficient).

The Long Tradition of Representative Evidence in Wage Actions

In a wage action under the FLSA, evidence of improperly compensated work is analyzed using a burden-shifting framework set out by the Supreme Court in 1946. *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680 (1946). There the Court, concerned that an employer could insulate itself from such suits by failing to maintain employment records that an employee could use to prove she was underpaid, ruled that in the absence of adequate employment records, an employee suing for lost wages under the FLSA must merely "submit sufficient evidence from which violations of the Act and the amount of an award may be reasonably inferred." *Martin v. Selker Bros., Inc.*, 949 F.2d 1286, 1296-97 (3d Cir. 1991).

As the Supreme Court stated:

“Where the employer's records are inaccurate or inadequate and the employee cannot offer convincing substitutes ... [t]he solution ... is not to penalize the employee by denying him any recovery on the ground that he is unable to prove the precise extent of uncompensated work. Such a result would place a premium on an employer's failure to keep proper records in conformity with his statutory duty; it would allow the employer to keep the benefits of an employee's labors without paying due compensation as contemplated by the Fair Labor Standards Act. In such a situation we hold that an employee has carried out his burden if he proves that he has in fact performed work for which he was improperly

compensated and if he produces sufficient evidence to show the amount and extent of that work as a matter of just and reasonable inference. The burden then shifts to the employer to come forward with evidence of the precise amount of work performed or with evidence to negative the reasonableness of the inference to be drawn from the employee's evidence. If the employer fails to produce such evidence, the court may then award damages to the employee, even though the result be only approximate.

Mt. Clemens, 328 U.S. at 687-88.

Representative Evidence in Wage Cases is Generally a Sampling of Testimony

To meet the burden under *Mt. Clemens*, a class of plaintiffs need not present testimony from each underpaid employee. Instead, the plaintiff class may and often does present the testimony of a representative sample of employees as part of the proof of the prima facie case under the FLSA. The adequacy of the sample relied upon or the testimony proffered, however, is often challenged in the courts.

In *Reich v. Southern New England Telecommunications Corp.*, 121 F.3d 58 (2d Cir. 1997), for example, the court held that the testimony of the representative sample of 2.5% of workers (or 39 of approximately 1,500 employees) was adequate evidence upon which to award back wages under the FLSA to the entire group of employees because (1) the testimony covered each clearly defined category of worker; (2) there was actual consistency among the workers' testimony, both within each category and overall; (3) the employer offered no

contradictory testimony; (4) the abuse arose from the admitted policy of employer that was consistently applied; and (5) the periods at issue were employees' lunch hours, which were predictable, daily-recurring periods of uniform and predetermined duration. Other courts have signed off on a wide range of sample sizes. *Donovan v. New Floridian Hotel, Inc.*, 676 F.2d 468, 472-73 (11th Cir. 1982) (23 employees testified; back wages awarded to 207 employees); *Herman v. Hector I. Nieves Transport, Inc.*, 91 F. Supp. 2d 435 (D. Puerto Rico 2000) (14 out of 100 testifying truck drivers provided adequate basis for determining average number of hours worked for testifying and non-testifying drivers); *McLaughlin v. DialAmerica Mktg., Inc.*, 716 F. Supp. 812, 824-25 (D.N.J. 1989) (testimony of 43 witnesses, both at trial and by deposition, confirms existence of violations for approximately 350 non-testifying employees); *Donovan v. Kaszycki & Sons Contractors, Inc.*, 599 F. Supp. 860, 868 (S.D.N.Y. 1984) (29 employees testified by deposition; back wages awarded to over 200 employees).

It appears that the only sure route to reversible error here is for a court to reject the concept of representative testimonial evidence altogether. In *Fegley v. Higgins*, 19 F.3d 1126 (6th Cir. 1994), the Sixth Circuit reversed the district court's outright refusal to "extrapolate" an approximation of damages and remanded the action for determination of damages in accordance with the *Mt. Clemens* standard. The circuit court held that the employee workers need not prove their damages with precision.

Representative Evidence, Outside the Wage Context, Is Often Presented as Expert Statistical Evidence

A classic example of representative evidence, outside the wage context, is the oft-cited *Hilao v. Estate of Marcos*, 103 F.3d 167 (9th Cir. 1996). There, a class of Philippine nationals and their survivors sued the estate of the deceased Philippine president Ferdinand Marcos for torture, disappearances, and summary executions. *Id.* After finding for the plaintiffs on the liability phase of the case, the court distributed notice to the class, informing them of their right to opt in to the case for purposes of damages; over 10,000 claims forms were filed. *Id.* at 772.

Faced with the unfathomable task of trying nearly 10,000 separate actions for compensatory damages¹ the court turned to expert methodology. *Id.* at 782. A statistical expert opined that a sample of 137 of the claimants, chosen randomly by computer, “would achieve ‘a 95 percent statistical probability that the same percentage determined to be valid among the examined claims would be applicable to the totality of claims filed.’” *Id.* After selection of a random sample of claimants, another expert, appointed by the court as special master, oversaw depositions of the selected claimants and their witnesses, reviewed the claim forms and testimony, and ultimately made a recommendation of the amount of damages to be awarded each of the sampled claimants. *Id.* at 782-83. Having invalidated six of the 137 claimants (or roughly 5%), the special master

¹ The court received a total of 10,059 claims; 518 were facially invalid. *Estate of Marcos*, 103 F.3d at 782.

devised a relatively simple analysis to reach an aggregated total amount of damages for the class. The analysis is summarized in the decision thusly:

	Torture	Summary Execution	Disappearance
Claims Filed	5,372	3,677	1,010
Facially Invalid Claims	-179	-273	- 66
Remaining Claims	5,193	3,404	944
Less 5% Invalidity Rate	-260	-170	- 47
Valid Claims	4,933	3,234	897
Valid Sample Claims	- 64	- 50	- 17
Valid Remaining Claims	4,869	3,184	880

[The Special Master] recommended that the award to the class be determined by multiplying the number of valid remaining claims in each subclass by the average award recommended for the randomly sampled claims in that subclass:

	Torture	Summary Execution	Disappearance
Valid Remaining Claims	4,869	3,184	880
x Average Awards	\$51,719	\$128,515	\$107,853
Class Awards	\$251,819,811	\$409,191,760	\$94,910,640

By adding the recommended awards in the randomly sampled cases, [the Special Master arrived at a recommendation for a total compensatory damage award in each subclass:

	Torture	Summary Execution	Disappearance
Class Awards	\$251,819,811	\$409,191,760	\$94,910,640
Sample Awards	\$ 3,310,000	\$ 6,425,767	\$ 1,833,515
TOTALS	\$255,129,811	\$415,617,527	\$96,744,155

Adding together the subclass awards, the [Special Master] recommended a total compensatory damage award of \$767,491,493.

Id. at 783-84.

This reasonably simple methodology was presented to the jury with instructions that they could accept or reject the special master's finding as to each claimant. *Id.* at 784. Though the jury disagreed with the Special Master in a number of specific instances, the total compensatory damages awarded in the case were more or less in line with the special master's recommendation: \$766 million. *Id.*

Use of surveys or sampling techniques has been allowed in a large number and wide variety of cases where the trial judge finds that he or she "face[s] crippling discovery and evidentiary costs." *In re Simon II Litigation*, 211 F.R.D. 86, 149 (E.D.N.Y. 2002) (J. Weinstein) (permitting statistical sampling and survey data to prove causation in complex tobacco litigation); *see also Harold's Stores, Inc. v. Dillard's Dept. Stores, Inc.*, 82 F.3d 1533 (10th Cir. 1996) ("In some cases, sampling techniques may provide the only practicable means to collect and present relevant data.") (quoting MANUAL FOR COMPLEX LITIGATION, THIRD, § 21.493 at 101 (1995)). Indeed, "[g]reater reliance on statistical methods is required by the profound evolution in our economic communication and data compilation and retrieval systems in recent decades." *In re Simon II Litigation*, 211 F.R.D. at 151.

The Use of Statistical Evidence in Class and Collective Wage Litigation

Given the bench's familiarity with statistical methods and the inherent suitability of wage cases to representative evidence, it is surprising that there are very few judicial opinions discussing the use of statistical methods in proving liability or damages in wage-and-hour cases under the FLSA or state wage-and-hour laws. Nevertheless, the Department of Labor and private practitioners have, for many years now, utilized statistical methods to present representative evidence in wage-and-hour cases. And with good reason. FLSA collective actions or class wage-and-hour actions under state laws, like many complex actions, present the court with the unappetizing prospect of resolving the claims of a large group of participating or covered plaintiffs who have suffered individualized harm pursuant to a common practice. Trying each claim separately would eradicate many of the efficiencies created by resorting to class or multi-plaintiff litigation in the first place.

Statistical sampling methods are particularly appropriate in large-scale wage-and-hour actions. These cases involve tremendous amounts of individualized data, namely the actual hours worked each day by each claimant during the relevant time period. Yet because records of the actual time worked often do not exist, but instead must be recreated based on recollection and estimation, the calculations are necessarily approximate. Statistical sampling enables courts to produce highly accurate aggregate determinations of hours worked using the same types of data required for individual claims and to

thereby efficiently resolve class-wide wage actions. Moreover, because the ultimate factual determination in wage-and-hour litigation concerns the actual minutes and hours of work performed, rather than questions such as discriminatory animus in the Title VII context that are a step removed from the statistical data, statistical sampling is highly determinative and therefore extremely useful. The downfall, however, comes from the requisite averaging that occurs, which while highly effective in producing aggregate damage amounts may result in a “rough justice” whereby each individual plaintiff recovers the average amount of unpaid wages rather than a more individualized determination of damages.

Learning the Hard Way: Dole v. Haulaway

Therefore, we should not be surprised that in 1989, the Department of Labor was all but *forced* by a district judge to introduce statistical evidence into a trial of an FLSA overtime action. *Dole v. Haulaway, Inc.*, 723 F. Supp. 274 (D. N.J. 1989). In the opinion, the judge describes, in almost epiphanic terms, his realization that statistical or expert information could simplify the trial:

A number of employees testified about various of the above factors which affected the hours of their work and there was stipulated testimony of an even greater number of employees about these factors. There was some variation in the employees' testimony about the same matters. During the first three days of trial the mass of undigested data referred to above was introduced into evidence by way of testimony and stipulated testimony. *By July 20, the third day of trial, it dawned upon me that the government had no intention of*

using its expertise in this field to pull together the vast array of data and by means of expert testimony or otherwise compute and present what it had concluded was the total overtime for which compensation had not been paid. As government counsel stated "after the Court [makes] a finding on ... how many hours did people work ... the government would then go back and present a schedule of the specific dollar amount due for each individual." (Tr. at 420, 421). The government correctly pointed out that such a procedure was contemplated by the pretrial order. However, awakening at last, I concluded that the procedure was "totally unacceptable" (Tr. at 421) and required the government to be prepared to present through a qualified witness or witnesses a schedule admissible in evidence showing exactly what it was that the government claimed was owing with respect to each employee.

Id. at 277 (emphasis added).

With the court literally crying out for use of an expert methodology to simplify the proceeding, the Department proffered the testimony of one its own compliance officers.² *Id.* at 278. The compliance officer summarized the hours of overtime claimed by each employee who had testified or for whom there had been a stipulation. *Id.* at 278-80. He then averaged the amount of per-employee overtime - excluding two obvious outliers - and applied the averaged figure to each non-testifying employee, reaching a total figure of a little over \$583,000 in unpaid overtime wages. *Id.* at 280.

² Though not noted in the *Haulaway* opinion, the government had, just a year earlier, also in the District of New Jersey, proffered the testimony of one of its compliance officers as an expert, presenting statistical analyses of piecework/homework data to show violations of the minimum wage law. *McLaughlin v. DialAmerica Mktg., Inc.*, 716 F. Supp. 812 (D. N.J. 1989).

Understandably, defendants objected to the late entry of this expert material and asked for an adjournment to prepare a statistical analysis of their own, which they did, determining at first that they had underpaid their employees by only \$200,875. *Id.* at 282. The government objected to defendant's methods, and later, *defendants* objected to defendant's methods, revising their estimate downwards to only \$47,736.³ *Id.* at 282-83.

The absurdity of the course the trial took and the difficulties caused by proceeding without experts and a plan for expert testimony was not lost on the court, which offered a telling *mea culpa* and a ringing endorsement of using sampling in future wage-and-hour cases:

It should be apparent that the tortured course of the trial of this case (which had to be tried in three segments during the July 18-August 24 period) was caused by the court's failure to require both the government and the defendants prior to the pretrial conference to submit detailed reports setting forth their computations of overtime hours and the entire basis for those computations. If that had been done the respective experts could have been deposed, the underlying bases for the computations could have been explored and verified, settlement discussions could have been conducted and failing settlement the trial could have been conducted expeditiously without periodic adjournments and without any party facing any surprise or prejudice.

Id. at 283.

³ For those keeping score, the court ultimately resolved the dispute by adopting the government's damages figure, but applying a flat discount of 25% to correct for the crudeness in the government's methodology. *Haulaway, Inc.*, 723 F. Supp. at 286.

Creating Efficiency With Experts: Bell v. Farmers

Perhaps the most detailed treatment of expert evidence in the wage-and-hour context comes from the California state courts, interpreting the California-law analogue to the FLSA. See *Bell v. Farmers Ins. Exch.*, 9 Cal. Rptr. 544. *Bell* was a “misclassification” case, that is, it sought unpaid overtime for insurance claims representatives who were wrongly treated as administrative employees exempt from California’s overtime coverage. *Id.* at 549. Plaintiffs, after securing certification of a class of about 2500 current and former claims representatives, moved for summary judgment on the exemption issue and prevailed. *Id.* at 550-51.

With liability determined, it was left for the parties to litigate damages. Defendant sought twice to decertify the class and have damages determined on an employee-by-employee basis; the court denied these attempts. *Id.* at 550-51. Both parties retained statistical experts who, not surprisingly, differed on the number of employees needed to be sampled to achieve a statistically suitable “confidence interval” of 95% or better to insure that the estimates derived from the sample would be within a one-hour margin of error: plaintiffs’ statistician felt sampling 95 employees would be sufficient; defendant’s, 1325. *Id.* at 551.

Both experts ultimately agreed to test their assumptions about the heterogeneity that sampling would yield by taking an initial pilot sample of 50 employees. *Id.* After reviewing the results of this sample, and being directed by the court to attempt to achieve a one-hour-per-week margin of error, the experts

agreed that a sample of 286 employee deponents would be sufficient.⁴ *Id.* at 551-52.

The quick jury trial on damages consisted of three witnesses: the two experts and an accounting expert offered by plaintiffs to apply materials from payroll records to the statistician's findings. *Id.* at 552-53. After a few hours of deliberation, the jury mostly adopted plaintiffs' accountant's findings, returning a verdict of a little more than \$90 million in unpaid wages. *Id.* at 553.

On appeal, defendant challenged the use of statistical evidence to determine aggregate class-wide damages. *Id.* at 571. Noting that defendant had not challenged the methodology or the qualifications of plaintiffs' expert, the appellate court restricted its review to two questions: whether (1) the trial court abused its trial management discretion in setting and approving the methodology of, the process for presenting, and the ultimate use of the statistical material in the case; and (2) the use of statistical inference violated due process concerns. *Id.* at 572.

The answer? No, on both counts. Noting the use of sampling in mass tort and other contexts, in addition to the use of sampling in other wage cases,⁵ the

⁴ Litigation vagaries being what they are, the parties ultimately ended up relying on a sample of 295 employee depositions. *Bell*, 9 Cal. Rptr. 552.

⁵ In addition to *DialAmerica*, discussed *supra* n.2, the *Bell* court cited to *Donovan v. Hudson Stations, Inc.*, 1983 WL 2110 (D. Kan. Oct. 14, 1983), a minimum wage and overtime case brought by the government on behalf of gas station workers; the court allowed the use of sampling and averaging methods that were "not . . . scientific" to award damages classwide. The *Bell* court also cited *Reich v. Walbaum, Inc.*, 833 F. Supp. 1037 (S.D.N.Y. 1993), wherein a DOL employee's

court ultimately determined that, while even the best sampling will yield averages that may overestimate or underestimate the damages due each individual employee, it creates judicial efficiencies and can be strikingly accurate in aggregate calculation of damages owed to the class as a whole:

[t]o the extent that the use of statistical sampling led to rough or expedient justice, it was in the adjudication of the relative entitlements of individual claimants to share in the aggregate award. [Defendant] itself was not prejudiced by this process.

Id. at 571-81.

* * *

Applying the lessons learned in previous cases, we can see that whatever danger of “rough justice” is posed by the use of statistical sampling in wage cases is probably only outweighed by the danger of unmanageability posed by *not* using sampling methods. The hallmarks of class and complex litigation – efficiency, aggregation of small claims that might never be otherwise heard – are advanced through sampling methods just as surely as they would be undermined by a requirement that individual damages be proven with exacting precision.

averaging to determine the damages due non-testifying employees were determined “reasonable,” if “not ideal.” Not discussed by the court, but also worthy of note here is *Reich v. IBP, Inc.*, 1996 WL 137817 (D. Kan. March 21, 1996), wherein both parties proffered expert evidence in the damages phase of a bifurcated trial of FLSA off-the-clock claims: the government relied on the type of sampling discussed at length here; IBP conducted time studies, wherein industrial engineers timed employees at their tasks and extrapolated off-the-clock time based on their estimates of the amount of time reasonably needed to complete certain tasks.