

Southern Methodist University

SMU Scholar

Faculty Journal Articles and Book Chapters

Faculty Scholarship

2018

The Admissibility of Sampling Evidence to Prove Individual Damages in Class Actions

Hillel J. Bavli

Southern Methodist University, Dedman School of Law

John Kenneth Felter

Author ORCID Identifier:

Hillel J. Bavli:  <https://orcid.org/0000-0003-4888-2842>

Recommended Citation

Hillel J. Bavli & John K. Felter, The Admissibility of Sampling Evidence to Prove Individual Damages in Class Actions, 59 B.C.L. Rev. 655 (2018)

This document is brought to you for free and open access by the Faculty Scholarship at SMU Scholar. It has been accepted for inclusion in Faculty Journal Articles and Book Chapters by an authorized administrator of SMU Scholar. For more information, please visit <http://digitalrepository.smu.edu>.

**THE ADMISSIBILITY OF SAMPLING EVIDENCE
TO PROVE INDIVIDUAL DAMAGES
IN CLASS ACTIONS**

HILLEL J. BAVLI
JOHN KENNETH FELTER

INTRODUCTION 657

I. DEMISE AND REVIVAL OF SAMPLING EVIDENCE IN FEDERAL COURT CLASS ACTIONS..... 659

A. Wal-Mart v. Dukes 661

B. Tyson v. Bouaphakeo 663

C. Harmonizing Wal-Mart and Tyson 667

D. Brief Bleak History of Sampling Evidence in Federal Court Class Actions 669

 1. *Cimino v. Raymark Industries* 669

 2. *Hilao v. Estate of Marcos* 671

II. SAMPLING EVIDENCE AND CLASS CERTIFICATION DISPUTES 674

A. Admissibility of Expert Evidence 674

B. Admissibility of Sampling Evidence 678

C. Generally Accepted Statistical Standards 680

D. Daubert and Class Certification Proceedings 682

III. SAMPLING EVIDENCE AND CLASS ACTION PREREQUISITES: COMMONALITY, PREDOMINANCE, AND SUPERIORITY 684

A. Commonality 685

B. Predominance 686

C. Superiority 690

D. Aggregate Damages and Individual Damages 691

E. Bifurcation 697

F. Sampling and Extrapolation 699

IV. ADMISSIBILITY OF SAMPLING EVIDENCE TO PROVE INDIVIDUAL DAMAGES..... 701

A. Revisiting Tyson..... 701

B. Threshold for Homogeneity..... 704

C. Claim Variability and Judgment Variability 704

D. Variability, Accuracy, and Evidentiary Reliability 704

E. Implications of Sampling for Accuracy and Reliability 707

F. Threshold for Admissibility 710

G. Hierarchical Model..... 710

H. Correlation 715

I. Estimating Claim Variability and Judgment Variability..... 717

J. Claim Variability and Samples..... 717

K. Judgment Variability and “Test Cases” 719

CONCLUSION 722

THE ADMISSIBILITY OF SAMPLING EVIDENCE TO PROVE INDIVIDUAL DAMAGES IN CLASS ACTIONS

HILLEL J. BAVLI*
JOHN KENNETH FELTER**

Abstract: The 2016 Supreme Court decision in *Tyson Foods, Inc. v. Bouaphakeo* revived the use of “representative” or sampling evidence in class actions. Federal courts are now more receptive to class plaintiffs’ efforts to prove classwide liability and, occasionally, aggregate damages, with sampling evidence. However, federal courts still routinely deny motions for class certification because they find that calculations of class members’ individual damages defeat the predominance prerequisite of Rule 23(b)(3). As a result, meritorious classwide claims founder. In this paper, we combine legal and statistical analyses and propose a novel solution to this dilemma that adheres to the *Tyson* decision while satisfying *Daubert*, the standards of Federal Rule of Evidence 702, and the prerequisites for Rule 23(b)(3) classes. We develop a method and derive a threshold to determine whether class damages claims are sufficiently homogeneous to justify the admissibility of sampling evidence to prove individual damages. Relying on *Daubert* and its progeny, as well as other well-recognized authority, we argue that accuracy is an appropriate standard for evidentiary reliability. Then, using generally accepted statistical methods and standards, we show that, when judgment variability exceeds claim variability (terms we define), sampling evidence improves accuracy and evidentiary reliability and is, therefore, admissible in Rule 23(b)(3) class certification proceedings. We also recommend several procedures to evaluate whether damages claims of a putative class satisfy the derived threshold. We conclude that our proposed method to prove individual damages achieves the Supreme Court’s stated goals of Rule 23(b)(3) class actions, “economies of time, effort, and expense” and the promotion of “uniformity of decision as to persons similarly situated, without sacrificing procedural fairness or bringing about other undesirable results.”

© 2018 Hillel J. Bavli & John Kenneth Felter. All rights reserved.

* Fellow, Harvard Institute for Quantitative Social Science; Visiting Fellow, Yale Law School Center for Private Law.

** Counsel, Ropes & Gray LLP (Boston); Fellow, American College of Trial Lawyers; Senior Fellow, Harvard University Advanced Leadership Initiative.

The authors thank Yang Chen and Iavor Bojinov for their helpful comments regarding the analysis in Part IV of this Article.

INTRODUCTION

Plaintiffs moving for class certification pursuant to Rule 23(b)(3) of the Federal Rules of Civil Procedure must demonstrate, among other things, that “questions of law or fact common to class members predominate over any questions affecting only individual members.”¹ To establish predominance, plaintiffs must show that common classwide proof can resolve one or more liability or damages issues, and that individual inquiries—those necessitating the introduction of evidence that varies from member to member—are not qualitatively more significant than common issues. Putative class plaintiffs have proposed methods to prove classwide liability or damages by proffering sampling evidence, contending that common questions predominate over individual questions notwithstanding the existence of differences among class members’ claims. They have argued that it is appropriate to extrapolate from adjudications of a representative sample of class members’ claims to determine classwide liability or to calculate classwide damages. However, federal courts have rarely accepted these arguments.

In 2011, in *Wal-Mart Stores, Inc. v. Dukes*, the United States Supreme Court rejected as “Trial by Formula” a proposal to extrapolate classwide liability and damages based on a small sample of claims selected for individual adjudications.² Two years later, in *Comcast Corp. v. Behrend*, after confirming that federal district courts must perform a “rigorous analysis” to test whether plaintiffs have satisfied the Rule 23(b)(3) predominance standard, the Court arguably raised the evidentiary bar when it reversed a class certification order.³ The Court held that the Third Circuit committed reversible error by refusing to consider whether a proposed statistical model was capable of proving classwide damages.⁴ *Wal-Mart* and *Comcast* were generally viewed as the death knell for sampling evidence in class actions.⁵

In 2016, however, in *Tyson Foods, Inc. v. Bouaphakeo*, the Court approved the admissibility of “representative evidence” to prove classwide liability.⁶ More specifically, it held that “[o]ne way for [class plaintiffs] to

¹ FED. R. CIV. P. 23(b)(3).

² 564 U.S. 338, 367 (2011).

³ 569 U.S. 27, 33–34 (2013).

⁴ *Id.*

⁵ *Comcast* can be read to suggest that predominance is not satisfied unless damages can be determined on a classwide basis. *See id.* at 34. However, most courts do not interpret *Comcast* so broadly, finding predominance of other common issues, even if individual damages calculations are necessary. *See, e.g.,* *Neale v. Volvo Cars of N. Am., LLC*, 794 F.3d 353, 375 (3d Cir. 2015); *Roach v. T.L. Cannon Corp.*, 778 F.3d 401, 407 (2d Cir. 2015); *In re Deepwater Horizon*, 739 F.3d 790, 815 (5th Cir. 2014); *In re Whirlpool Corp. Front-Loading Washer Prods. Liab. Litig.*, 722 F.3d 838, 859–60 (6th Cir. 2013); *Leyva v. Medline Indus. Inc.*, 716 F.3d 510, 513–14 (9th Cir. 2013). In any event, it is now settled that solely the need to calculate individual damages will not defeat certification of a Rule 23(b)(3) class. *See infra* notes 178–181 and accompanying text.

⁶ 136 S. Ct. 1036, 1046 (2016).

show . . . that the sample relied upon . . . is a permissible method of proving classwide liability is by showing that each class member could have relied on that sample to establish liability if he or she had brought an individual action.”⁷ The Court explained that the admissibility of “representative evidence” turns on its relevance and, most importantly, its reliability.⁸

Tyson may usher in a revival of sampling evidence in class action litigation.⁹ After *Tyson*, class plaintiffs are increasingly offering “representative evidence,” including sampling evidence, as a method to prove classwide liability to satisfy the predominance prerequisite of Rule 23(b)(3). But the permissibility of sampling evidence to prove classwide damages—and especially individual damages—remains uncertain and controversial. Frequently, when a putative Rule 23(b)(3) class asserts individual damages claims, federal district courts decide to certify a class to determine classwide liability issues, and perhaps even aggregate damages issues, but reserve the calculations of individual class members’ damages for separate adjudications. Even more often, if the calculations of individual damages cannot be computed formulaically, federal district courts find that individual issues predominate over common issues, and refuse to certify a Rule 23(b)(3) class.

In this Article, we argue that, in appropriate circumstances, sampling evidence is admissible to meet the predominance standard and as a method to prove individual damages when plaintiffs move for class certification pursuant to Rule 23(b)(3). We answer a threshold question: When does sampling evidence, offered to prove individual damages in a Rule 23(b)(3) class action, pass muster under governing evidentiary standards? We explain that, whether sampling evidence is admissible, and whether a federal district court will certify a Rule 23(b)(3) class based on such evidence, depend on the relevance and reliability of the proffered evidence. We explore the circumstances under which sampling evidence establishing individual damages satisfies these governing evidentiary standards.

After this brief introduction, in Part I, as historical background, we discuss the demise and revival of sampling evidence in federal court class actions. In Part II, we analyze the evidentiary law governing the admissibility of expert evidence in class certification disputes. In Part III, we examine the Rule 23(b)(3) class-action prerequisites of commonality, predominance, and superiority, with a focus on predominance. We also summarize the circumstances under which federal district courts have generally found that, in Rule 23(b)(3) class actions, common issues predominate over individual issues, notwithstanding variable individual damages. We next discuss the

⁷ *Id.* at 1046–47.

⁸ *Id.* at 1046.

⁹ See Jonah B. Gelbach, *The Triangle of Law and the Role of Evidence in Class Action Litigation*, 165 U. PA. L. REV. 1807, 1845 (2017) (“Still, *Tyson* may be a big case. It may herald a more pragmatic approach in Supreme Court assessments of class certification.”).

admissibility of expert evidence, and sampling evidence in particular, to prove individual damages in Rule 23(b)(3) class actions. In Part IV, we propose a statistical method to support the introduction of sampling evidence during class certification proceedings that satisfies *Daubert*, Federal Rule of Evidence 702 standards that govern the admissibility of expert evidence, and the predominance prerequisite of Rule 23(b)(3). We develop a method and derive a threshold to determine whether class damages claims are sufficiently homogeneous to justify the admissibility of sampling evidence to prove individual damages. We argue that accuracy is an appropriate standard for evidentiary reliability and show that, when judgment variability exceeds claim variability (terms we define), sampling evidence improves accuracy and evidentiary reliability, and is admissible. Finally, we conclude that our proposed method to prove individual damages achieves the Supreme Court's stated goals for Rule 23(b)(3) class actions, "economies of time, effort, and expense" and the promotion of "uniformity of decision as to persons similarly situated, without sacrificing procedural fairness or bringing about other undesirable results."¹⁰

I. DEMISE AND REVIVAL OF SAMPLING EVIDENCE IN FEDERAL COURT CLASS ACTIONS

After the Supreme Court's 2011 decision in *Wal-Mart Stores, Inc. v. Dukes*, several commentators declared that the use of statistical evidence to prove liability or damages in class actions was dead or, at best, moribund.¹¹ In 2016, however, in *Tyson Foods, Inc. v. Bouaphakeo*, the Court clarified that: "*Wal-Mart* does not stand for the broad proposition that a representative sample is an impermissible means of establishing classwide liability."¹² Thus, to paraphrase Mark Twain, the reports of the death of statistical evidence in class actions proved to be an exaggeration. The *Tyson* Court proclaimed: "Whether and when statistical evidence can be used to establish classwide liability will depend on the purpose for which the evidence is being

¹⁰ *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 615, 634 (1997) (quoting FED. R. CIV. P. 23 advisory committee's notes to 1966 amendment).

¹¹ See, e.g., Jay Tidmarsh, *Resurrecting Trial by Statistics*, 99 MINN. L. REV. 1459, 1459 (2015) ("[T]rial by statistics died on June 20, 2011 [date of *Wal-Mart* decision]. . . . The notion that a court could try a representative sample of monetary claims and extrapolate the average result to the remainder of the cases was finished."); Wenbo Whang, *Tyson Foods, Inc. v. Bouaphakeo: The Use of Statistical Evidence in Class Actions*, 11 DUKE J. CONST. L. & PUB. POL'Y SIDEBAR 133, 133 (2016), https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1130&context=djclpp_sidebar [<https://perma.cc/R758-NXS6>] ("[T]he Supreme Court has largely foreclosed arguments for trial by statistics in *Wal-Mart Stores v. Dukes*."). But see, e.g., *In re Urethane Antitrust Litig.*, 768 F.3d 1245, 1257 (10th Cir. 2014) (noting that "*Wal-Mart* does not prohibit [class] certification based on the use of extrapolation to calculate damages").

¹² *Tyson*, 136 S. Ct. at 1048.

introduced and on ‘the elements of the underlying cause of action.’”¹³ With that dicta, the Court resisted the urgings of the *Tyson* parties and their respective amici to adopt “broad and categorical rules governing the use of representative and statistical evidence in class actions.”¹⁴ As the Court explained:

¹³ *Id.* at 1046 (quoting *Erica P. John Fund, Inc. v. Halliburton Co.*, 563 U.S. 804, 809 (2011)). The *Tyson* Court did not elucidate what it meant by “the purpose for which the [statistical] evidence is being introduced.” The results of sampling are often admitted into evidence in various legal contexts. *See, e.g.*, 1 DAVID L. FAIGMAN ET AL., *MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY* 479, 483–86 (2015–16 ed.) (discussing uses of sampling surveys in cases involving, among others, employment discrimination, juror bias, obscenity, and trademark litigation); 2 JOSEPH L. GASTWIRTH, *STATISTICAL REASONING IN LAW AND PUBLIC POLICY* 483–528 (1988) (discussing uses of sampling in cases involving, among others, false labeling and gender-based equal protection). By way of examples only, statistical evidence, and more particularly sampling evidence, is frequently introduced to prove “likelihood of confusion” in trademark infringement cases, *see* 4 J. THOMAS MCCARTHY, *MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION* §§ 23:1–23:11.50 (4th ed. 2010); “disparate impact” in Title VII (42 U.S.C. §§ 2000e-1–2000e-17) cases, *see* GEORGE RUTHERGLEN, *EMPLOYMENT DISCRIMINATION LAW: VISIONS OF EQUALITY IN THEORY AND DOCTRINE* 79–89 (3d ed. 2010); and antitrust damages, *see, e.g.*, *Zenith Radio Corp. v. Hazeltine Research, Inc.*, 395 U.S. 100, 116 & n.11 (1969). *See generally* Joseph B. Kadane, *Probability Sampling in Litigation*, 18 CONN. INS. L.J. 297, 299–303 (2011). As one commentator suggested, “broadly interpreted,” *Tyson* “stands for the proposition that sampling can be used to overcome any serious proof obstacle that systematically deprives a large number of injured parties of compensation, impedes enforcement of the substantive law, and leaves the defendant free to retain the benefits of its unlawful conduct—provided, of course, that sampling is otherwise consistent with the applicable substantive law.” Robert Bone, *Tyson Foods and the Future of Statistical Adjudication*, 95 N.C. L. REV. 607, 636 (2017).

The *Tyson* Court also did not explain its quotation from *Erica P. John Fund*—“the elements of the underlying cause of action.” The sole dispute in *Erica P. John Fund* was whether plaintiff satisfied the predominance prerequisite of Rule 23(b)(3). The Court stated “[c]onsidering” predominance “begins, of course, with the elements of the underlying cause of action.” *Erica P. John Fund*, 563 U.S. at 809. There, plaintiff’s claims were based on alleged violations of section 10(b) of the Securities Exchange Act of 1934 (15 U.S.C. § 78j(b)) and Securities and Exchange Commission Rule 10b-5 (17 C.F.R. § 240.10b-5). *Id.* at 809–10. Whether common questions of law or fact predominated turned on, as it often does in securities fraud actions, the element of reliance. The narrow question presented for review by the Supreme Court was whether plaintiff needed to establish “loss causation as a precondition for invoking *Basic*’s [“fraud-on-the-market”] rebuttable presumption of reliance.” *Id.* at 812; *Basic Inc. v. Levinson*, 485 U.S. 224, 243 (1988). The Supreme Court held that the Fifth Circuit erred in requiring plaintiff to make this showing as a precondition of class certification. *Erica P. John Fund*, 563 U.S. at 813. In short, the Court ruled that an inability to prove “loss causation” does not prevent a plaintiff from invoking the “fraud-on-the-market” rebuttable presumption of reliance as evidence to satisfy that “element[] of the underlying cause of action.” *Id.* at 809. Taken together, *Tyson* and *Erica P. John Fund* teach that introducing statistical evidence to prove one or more elements of class claims will not only support the admissibility of that evidence, but also undergird the satisfaction of the predominance prerequisite of Rule 23(b)(3).

¹⁴ *Tyson*, 136 S. Ct. at 1040. The death of Justice Antonin Scalia on February 13, 2016, undoubtedly influences the jurisprudence of the current sharply-divided Supreme Court. Justice Scalia heard oral argument in *Tyson* on November 10, 2015, but died before the opinion was issued on March 22, 2016. He was an avowed proponent of curtailing class actions, as evidenced by his authorship of the majority opinions in *Comcast Corp. v. Behrend*, 569 U.S. 27 (2013), *Wal-*

A representative or statistical sample, like all evidence, is a means to establish or defend against liability. Its permissibility turns not on the form a proceeding takes—be it a class or individual action—but on the degree to which the evidence is reliable in proving or disproving the elements of the relevant cause of action.¹⁵

A comparison of *Wal-Mart* and *Tyson* illuminates the superficial inconsistency of the Court's two decisions. However, in *Tyson*, the Court attempted to harmonize them.

A. Wal-Mart v. Dukes

The “crux” of the *Wal-Mart* case is “commonality.”¹⁶ The *Wal-Mart* Court found that the only corporate policy established by plaintiffs’ evidence was “Wal-Mart’s ‘policy’ of *allowing discretion* by local supervisors over employment matters. . . . just the opposite of a uniform employment practice that would provide the commonality needed for a class action”¹⁷

In *Wal-Mart*, a majority of the Court rejected regression analyses performed by plaintiffs’ expert statistician and expert labor economist. It found that the regional and national data used by the experts “does not establish the existence of disparities at individual stores, let alone raise the inference that a company-wide policy of discrimination is implemented by discretionary decisions at the store and district level.”¹⁸ The *Wal-Mart* Court also identified a “more fundamental . . . respect in which [plaintiffs’] statistical proof fails.”¹⁹ It found that the “nature and effects” of Wal-Mart managers’ applying “performance-based criteria . . . will differ from store to store,” and ruled that “[m]erely showing that Wal-Mart’s policy of discretion has produced an overall sex-based disparity does not suffice.”²⁰

Plaintiffs in *Wal-Mart* proposed to select a sample set of class members and to determine liability and damages for those members by depositions

Mart Stores, Inc. v. Dukes, 564 U.S. 338 (2011), and *AT&T Mobility LLC v. Concepcion*, 563 U.S. 333 (2011).

¹⁵ *Tyson*, 136 S. Ct. at 1046 (citing FED. R. EVID. 401, 403, 702). As will be discussed, reliability is paramount when considering the admissibility of “[a] representative or statistical sample.” See *infra* note 190 and accompanying text and Part IV.

¹⁶ 564 U.S. at 349; see FED. R. CIV. P. 23(a)(2) (“[T]here are questions of law or fact common to the class . . .”).

¹⁷ 564 U.S. at 355.

¹⁸ *Id.* at 356–57 (quoting *Dukes v. Wal-Mart Stores, Inc.*, 603 F.3d 571, 637 (9th Cir. 2010) (Ikuta, J., dissenting), *rev’d*, 564 U.S. 338 (2011)).

¹⁹ *Id.* at 357.

²⁰ *Id.* The four dissenting justices in *Wal-Mart* (Ginsburg, Breyer, Sotomayor, and Kagan, JJ.) contended that plaintiffs’ expert statistician’s regression analyses showed that there were disparities within stores, not just at the regional and national level, and that the majority’s contrary view “reflect[ed] only an arcane disagreement about statistical method.” *Id.* at 372 n.5 (Ginsburg, J., concurring in part and dissenting in part).

supervised by a master.²¹ Under their proposal, classwide aggregate damages would be derived by taking the “percentage of claims determined to be valid” from the sample and applying that percentage to non-sample class members, and then multiplying the “number of (presumptively) valid claims” by the average damages in the sample.²² The *Wal-Mart* Court found that plaintiffs’ proposed statistical evidence would not tend to prove classwide claims and held that this “Trial by Formula” violated the Rules Enabling Act because it “enlarge[d]” the class members’ “substantive right[s]” and deprived Wal-Mart of its right to raise statutory defenses to individual claims.²³

²¹ 564 U.S. at 367.

²² *Id.*

²³ *Id.* (quoting 28 U.S.C. § 2072(b) (2006)). Responses to arguments against the use of statistical evidence in class actions based on the Rules Enabling Act, Article III (federal court authority), Seventh Amendment (jury trial right and Reexamination Clause), and Fourteenth Amendment (Due Process Clause) are beyond the scope of this article. U.S. CONST. art. III, amend. VII, XIV; 28 U.S.C. § 2072(b) (2012).

On April 4, 2016, two weeks after its *Tyson* decision, the Supreme Court denied a petition for a writ of certiorari to review a \$188 million class action judgment against Wal-Mart in a Pennsylvania wage and hour case. *Braun v. Wal-Mart Stores, Inc.*, 106 A.3d 656, 658 (Pa. 2014), *cert. denied*, 136 S. Ct. 1512 (2016) (mem.). In that state court trial, plaintiffs’ experts extrapolated sample evidence to calculate classwide aggregate damages. *Id.* at 661–62. Despite Wal-Mart’s contention that this “Trial by Formula” violated its constitutional rights and was proscribed by *Wal-Mart*, the Supreme Court declined to review the case without comment. *Id.* at 663–67.

After *Wal-Mart*, several federal courts latched on to the “Trial by Formula” rubric and disapproved sampling methodologies to prove classwide liability or damages. *See, e.g.*, *Jimenez v. Allstate Ins. Co.*, 765 F.3d 1161, 1167 (9th Cir. 2014) (affirming bifurcated class action order because “statistical sampling and representative testimony are acceptable ways to determine liability so long as the use of these techniques is not expanded into the realm of damages”); *Davis v. Cintas Corp.*, 717 F.3d 476, 490–91 (6th Cir. 2013) (characterizing sampling method used as “worse than the system that the Supreme Court unanimously rejected in” *Wal-Mart* and rejecting “shortfall-based” model” to order hiring of randomly selected class members and to calculate aggregate damages that would be distributed to eligible class members pro rata); *In re Elec. Books Antitrust Litig.*, No. 11 MD 2293-DLC, 2014 WL 1282293, at *22–23 (S.D.N.Y. Mar. 28, 2014) (describing *Wal-Mart* “Trial by Formula” as “a plan to try a sample set of class members’ claims . . . and then multiply the average backpay award to determine the class-wide recovery without further individualized proceedings”); *Slipchenko v. Brunel Energy, Inc.*, No. H-11-1465-LHR, 2013 WL 4677918, at *10 n.8 (S.D. Tex. Aug. 30, 2013) (distinguishing *Wal-Mart* “Trial by Formula” because plaintiffs did “not propose[] a sampling-based method to determine penalties”); *Jacob v. Duane Reade, Inc.*, 293 F.R.D. 578, 588–93 (S.D.N.Y. 2013) (bifurcating liability and damages issues and partially decertifying damages issues of Rule 23(b)(3) class due to sampling concerns); *Acosta v. Tyson Foods, Inc.*, No. 8:08 CV 86-JFB, 2013 WL 7849473, at *18 (D. Neb. May 31, 2013) (“Although a ‘trial by formula,’ wherein damages are determined for a sample set of class members and then applied by extrapolation to the rest of the class ‘without further individualized proceedings,’ has been disapproved, that disapproval cannot fairly be interpreted to apply to the use of representative testimony in a class action for unpaid wages.” (quoting *Wal-Mart*, 564 U.S. at 367)), *order clarified* 2013 WL 3716445 (July 12, 2013); *George v. Nat’l Water Main Cleaning Co.*, 286 F.R.D. 168, 181–82 (D. Mass. 2012) (contrasting “Trial by Formula” because plaintiffs alleged defendant’s “wage policies facially violated state law” and remedy only “involves reconstructing the correct wage algorithm,” and noting “defendants are entitled to any additional proceedings required to ensure that their due process rights are protected” (citations omitted)); *Stone v. Advance Am.*, 278 F.R.D. 562, 566 n.1 (S.D. Cal. 2011)

At bottom, the putative *Wal-Mart* class failed to meet the commonality prerequisite of Rule 23(a)(2) because plaintiffs did not prove that each class member was subjected to a common policy of discrimination.²⁴

B. *Tyson v. Bouaphakeo*

In *Tyson*, plaintiffs relied on “representative evidence,” including class members’ testimony, videotaped observations of certain employees’ activities, and “most important, a study performed by an industrial relations expert.”²⁵ Based on the videotaped observations, plaintiffs’ industrial relations expert calculated average times for the observed employees to perform the disputed activities, “donning” and “doffing” safety gear.²⁶ Another plaintiffs’ expert extrapolated those averages to all class members and, by using *Tyson* company time records, estimated the number of class members who had viable claims (that is, those who worked more than forty hours in one or more

(stating that *Wal-Mart* “largely eliminates a ‘trial by formula’ approach to use statistics to extrapolate average damages for an entire class”); *United States v. City of New York*, 276 F.R.D. 22, 38–39 n.7 (E.D.N.Y. 2011) (rejecting methodology for distributing aggregate damages because it “resembles the method the Supreme Court termed ‘Trial by Formula’”).

At least one Circuit Court has arguably narrowed the scope of *Wal-Mart*. See *Jimenez*, 765 F.3d at 1167. In *Jimenez*, the Ninth Circuit opined that the *Wal-Mart* Court “relied on two key factors,” namely that the holding was made “under Rule 23(b)(2), which contains fewer procedural safeguards than Rule 23(b)(3) [e.g., opt-out],” and Title VII “which explicitly includes affirmative defenses relating to motive and alternative explanations.” *Id.* The Ninth Circuit suggested that these “two key factors” limit the *Wal-Mart* holding. See *id.*

²⁴ 564 U.S. at 359. The conventional wisdom is that *Wal-Mart* makes it more difficult to satisfy the commonality prerequisite of Rule 23(a)(2). See, e.g., *Brown v. Nucor Corp.*, 785 F.3d 895, 903 (4th Cir. 2015) (noting “*Wal-Mart*’s heightened requirement of commonality”); *M.D. ex rel. Stukenberg v. Perry*, 675 F.3d 832, 839 (5th Cir. 2012) (“[T]he *Wal-Mart* decision has heightened the standards for establishing commonality under Rule 23(a)(2)”); Christine P. Bartholomew, *Redefining Prey and Predator in Class Actions*, 80 BROOK. L. REV. 743, 772 (2015) (stating that *Wal-Mart* “heightened the [commonality] requirement”); Elizabeth Chamblee Burch, *Constructing Issue Classes*, 101 VA. L. REV. 1855, 1863–64 (2015) (“*Wal-Mart Stores, Inc. v. Dukes* strengthened the commonality standard under Rule 23(a)”). Class certification orders filed after *Tyson* appear to confirm the conventional wisdom. See, e.g., *Dilts v. Penske Logistics, LLC*, No. 08-CV-0318-CAB-BLM, 2016 WL 4072485, at *3–6 & n.1 (S.D. Cal. July 20, 2016) (decertifying certain subclasses and rejecting plaintiff’s proposal to call forty random absent class members to testify at trial and to extrapolate from that testimony); *Brown v. United States*, 126 Fed. Cl. 571, 580–81 (2016).

²⁵ 136 S. Ct. at 1043. Plaintiffs’ claims were based on alleged violations of the Fair Labor Standards Act of 1938, 52 Stat. 1060, as amended, 29 U.S.C. §§ 201–219 (“FLSA”). “Representative evidence” and “representative testimony” are terms of art in FLSA actions that refer to evidence regarding selected employees that support findings for all “similarly situated” employees. See, e.g., *Monroe v. FTS USA, LLC*, 815 F.3d 1000, 1018, 1020–22 (6th Cir. 2016), *cert. granted, vacated*, 137 S. Ct. 590 (2016) (judgment vacated, and case remanded “in light of” *Tyson*). Cf. *Espenscheid v. DirectStat USA, LLC*, 705 F.3d 770, 775 (7th Cir. 2013) (finding no “representative evidence”).

²⁶ *Tyson*, 136 S. Ct. at 1043. Pre-*Tyson*, class plaintiffs relied on similar studies in *Garcia v. Tyson Foods, Inc.*, 770 F.3d 1300, 1306–07 (10th Cir. 2014), and *Perez v. Mountaire Farms, Inc.*, 650 F.3d 350, 370–72 (4th Cir. 2011).

weeks) and the aggregate amount of class damages.²⁷ That expert's "calculations supported an aggregate [damages] award of approximately \$6.7 million."²⁸ The jury awarded about \$2.9 million in damages.²⁹

²⁷ *Tyson*, 136 S. Ct. at 1043–44.

²⁸ *Id.* at 1044. The *Tyson* Court noted that Tyson "did not raise a challenge to [plaintiffs'] experts' methodology under *Daubert*; and, as a result, there is no basis in the record to conclude it was legal error to admit that evidence." *Id.* at 1049. The Court emphasized that, once admitted, the "persuasiveness" of evidence "is, in general, a matter for the jury" and "[r]esolving" whether plaintiffs' industrial relations expert's average times "is probative as to the time actually worked by each employee" "is the near-exclusive province of the jury." *Id.*

It is curious that Tyson neither challenged the validity of plaintiffs' experts' studies under *Daubert* nor introduced rebuttal expert testimony, *id.* at 1044, 1049, despite the fact that it listed an expert witness that it expected to call at trial, Def.'s Mem. Supporting Exclusion of Any Reference to Dr. Paul Adams, or His Ops., at Trial at 2–3, *Bouaphakeo v. Tyson Foods, Inc.*, 564 F. Supp. 2d 870 (N.D. Iowa 2008). Tyson's "primary defense was to show that [plaintiffs'] industrial relations expert's] study was unrepresentative or inaccurate." *Tyson*, 136 S. Ct. at 1047. In opposing class certification and to the jury, Tyson argued, unsuccessfully, that plaintiffs' expert's calculated average times were overstated and that the varying amounts of time that the observed employees performed the disputed activities made the claims "too speculative for classwide recovery." *Id.* at 1044, 1047; see Andrew J. Trask, *Litigation Matters: The Curious Case of Tyson Foods v. Bouaphakeo*, 2016 CATO S. CT. REV. 279. The fact that Tyson did not object to plaintiffs' expert evidence precluded Tyson from arguing that it was not admissible, but it did not prevent Tyson from arguing that the "evidence failed 'to show that the case is susceptible to awarding damages on a class-wide basis.'" *Comcast*, 569 U.S. at 32 n.4 (quoting question presented for review in *Comcast Corp. v. Behrend*, 567 U.S. 933 (2012)).

²⁹ *Tyson*, 136 S. Ct. at 1044. At Tyson's request, the jury was instructed that non-testifying class members could recover only if the evidence established that they "suffered the same harm as a result of the same unlawful decision or policy" as the testifying class members. *Id.* (citing J.A. at 101, 471–72, *Tyson Foods, Inc. v. Bouaphakeo*, No. 14-1146 (Aug. 7, 2015)). The trial court also adopted Tyson's proposed jury verdict form calling for an aggregate damages award. Brief for Respondents at 20, *Tyson Foods, Inc. v. Bouaphakeo*, No. 14-1146 (Sept. 22, 2015).

Tyson originally argued that class certification was proper only if plaintiffs established that all putative class members had valid claims, but it abandoned that argument in favor of a narrower challenge to the distribution of the aggregate damages awarded by the jury. *Tyson*, 136 S. Ct. at 1049. The *Tyson* Court did not address damages issues. However, in his concurring opinion, Chief Justice John Roberts stated that, based on the jury's damages award, the "jury obviously did not credit [plaintiffs'] industrial relations expert's] averages" and he stated further that he is "not convinced that the District Court will be able to devise a means of distributing the aggregate award only to injured class members." *Tyson*, 136 S. Ct. at 1051 (Roberts, C.J., concurring); see *id.* at 1044 (demonstrating that it was undisputed that over 200 class members did not have valid claims). If not, according to Chief Justice Roberts, "it remains to be seen whether the jury verdict can stand" because "Article III does not give federal courts the power to order relief to any uninjured plaintiff, class action or not." *Id.* at 1053.

After remand, in an October 6, 2016, Order regarding the distribution of classwide aggregate damages, the trial court noted that the *Tyson* "parties agreed to . . . 'filters' in order to ensure uninjured individuals do not receive any of the jury's aggregate award." *Bouaphakeo v. Tyson Foods, Inc.*, No. 5:07-cv-04009-JAJ, 2016 WL 5868081, at *4 (N.D. Iowa Oct. 6, 2016). First, workers who did not "reach 40 hours in a week without the aid of" plaintiffs' industrial relations expert's "donning" and "doffing" times estimates were "excluded from the award." *Id.* Second, only workers who are "owed at least \$50" after adding those estimates will share in the award. *Id.* According to the trial court, this "'buffer zone' further ensures that uninjured individuals will not receive damages by withholding damages from lesser-injured parties." *Id.* As a result, 199 workers "are not uninjured, rather, they are exempt from the award by agreement of the parties." *Id.*

The *Tyson* Court ruled that one way to show that a sample is a “permissible method of proving classwide liability is by showing that each class member could have relied on that sample to establish liability . . . [in] an individual action.”³⁰ The Court concluded that plaintiffs’ expert’s study was admissible and “could have been sufficient to sustain a jury finding . . . if it were introduced in each employee’s individual action.”³¹ Moreover, as the Court pointed out, “[i]n many cases, a representative sample is ‘the only practicable means to collect and present relevant data’ establishing a defendant’s liability.”³²

“Finally, by choosing to award any damages in light of the jury instructions regarding compensation of uninjured parties, the jury reinforced the ‘injured’ status of the class members.” *Id.*

On July 11, 2017, the parties filed a Joint Notice of Settlement of All Contested Issues Between the Parties and Joint Request to Vacate Current Court Deadlines for Plaintiffs to File Supplements to Their Petition for Award of Attorneys Fees and Costs, by which the parties notified the trial court that, among other things, they had “establish[ed] a process involving a third-party settlement administrator, . . . , to disburse all payments to the Class Members from the aggregate jury award in a manner consistent with this Court’s Order of October 6, 2016.” *Bouaphakeo v. Tyson Foods, Inc.*, No. 5:07-cv-04009-JAJ (N.D. Iowa July 11, 2017), ECF No. 432.

³⁰ 136 S. Ct. at 1046. The Court unanimously shared this view; *see also id.* at 1050 (Roberts, C.J., concurring); *id.* at 1053 (Thomas, J., dissenting). *Cf. Davenport v. Charter Commc’ns, LLC*, Case No. 4:12 CV 00007 AGF, 2017 WL 878029, at *8–9 (E.D. Mo. Mar. 6, 2017) (finding a “lack of evidence of a common policy” and therefore “little role for representative evidence”); *Angeles v. US Airways, Inc.*, No. C 12 -05860 CRB, 2017 WL 587658, at *4 (N.D. Cal. Feb. 13, 2017) (“Representative evidence would not fly for individual [class members] and it will not fly for a class of them.”).

³¹ *Tyson*, 136 S. Ct. at 1048.

³² *Id.* at 1046 (citing MANUAL FOR COMPLEX LITIGATION (FOURTH) § 11.493, at 102 (2004)). The 2015 edition of the *Manual for Complex Litigation* expands on this point: “Statistical methods can often estimate, to specified levels of accuracy, the characteristics of a ‘population’ . . . by observing those characteristics in a relatively small segment, or sample, of the population.” MANUAL FOR COMPLEX LITIGATION (FOURTH) § 11.493, at 102 (2015). In the *Reference Manual on Scientific Evidence* published by the Federal Judicial Center, the *Reference Guide on Statistics* begins: “Statistical assessments are prominent in many kinds of legal cases, including antitrust, employment discrimination, toxic torts, and voting rights cases.” David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 211, 213 (Federal Judicial Center, 3d ed. 2011) [hereinafter *Reference Guide on Statistics*].

As the Brief of Amici Curiae Complex Litigation Law Professors in *Tyson* highlighted: “Statistical techniques and similar approaches that rely upon a sample are an appropriate way to determine common issues of liability in a variety of settings, including antitrust, securities fraud, and employment discrimination litigation. In all of these settings statistical techniques have been especially useful at the class certification stage to determine whether the predominance requirement of Rule 23(b)(3) has been satisfied.” Brief for Complex Litigation Law Professors as Amici Curiae Supporting Respondents at 4–5, *Tyson Foods, Inc. v. Bouaphakeo*, No. 14-1146 (Sept. 29, 2015) (citing, *inter alia*, *Comcast*, 569 U.S. at 31–32) (antitrust); *Halliburton Co. v. Erica P. John Fund, Inc.*, 134 S. Ct. 2398, 2407 (2014) (securities fraud); *Messner v. Northshore Univ. Health Sys.*, 669 F.3d 802, 808 (7th Cir. 2012) (antitrust); *In re Hydrogen Peroxide Antitrust Litig.*, 552 F.3d 305, 311 (3d Cir. 2008) (antitrust); *Munoz v. Orr*, 200 F.3d 291, 300 (5th Cir. 2000) (employment discrimination); *Barnes v. GenCorp. Inc.*, 896 F.2d 1457, 1466 (6th Cir. 1990) (employment discrimination)); *see* MANUAL FOR COMPLEX LITIGATION (FOURTH)

In *Tyson*, the “central dispute” was whether the “average time” observed in the sample could be extrapolated to each member of the population was a “permissible” “inference.”³³ For this reason, the Court opined that its *Tyson* holding “is in accord with” *Wal-Mart* because the “underlying question” in both cases “was whether the sample at issue could have been used to establish liability in an individual action.”³⁴

§ 23.1, at 469–72 (2004); 6 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 18:45, at 155–56 (5th ed. 2012).

³³ 136 S. Ct. at 1046. The Brief of Economists and Other Social Scientists as Amici Curiae in Support of Respondents in *Tyson* offered this overview:

Inferential statistics, . . . comprises methods that allow us to use what is known to make estimates and predictions about the unknown. All inferential techniques in statistics are, in one way or another, based on well-accepted concepts of sampling and extrapolation. Sampling refers to the idea that observed information is an incomplete, grainy snapshot—a sample—taken from a larger universe of potentially observable information, called the population. Extrapolation means that this sample, although incomplete, can still be reliably informative about the population from which it is obtained. . . .

Good statistical practice calls for data collection procedures that minimize the risk of unrepresentative sampling, analytical tools appropriate to deal with a given sample, and trained interpretation that recognizes the potential limitations of both data and techniques.

Brief of Economists and Other Social Scientists as Amici Curiae in Support of Respondents at 7–8, *Tyson Foods, Inc. v. Bouaphakeo*, No. 14-1146 (Sept. 29, 2015).

Among other grounds for criticizing plaintiffs’ experts’ statistical evidence in *Tyson*, Justice Clarence Thomas in dissent objected to the sample size, “about 53,” that was used “to extrapolate averages for the 3,344-person class.” *Tyson*, 136 S. Ct. at 1055 (Thomas, J., dissenting). While discussing *Wal-Mart*, Justice Thomas alluded to the “1:8 ratio of anecdotes [samples] to class members that our prior cases accepted.” *Id.* at 1060 (citing *Wal-Mart*, 564 U.S. at 358); see also *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 338 (1977).

³⁴ *Tyson*, 136 S. Ct. at 1048. Some courts and commentators believe that the *Tyson* Court’s decision to accept “representative evidence” can be explained by, and therefore should be limited to, its substantive law context. As previously noted, in *Tyson*, plaintiffs’ claims were based on alleged violations of the FLSA. See *supra* note 25. The FLSA requires employers, such as *Tyson*, to “make, keep, and preserve” employee wage and hour records. 29 U.S.C. § 211(c) (2012). More than seventy years ago, the Supreme Court, in *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680 (1946), held that when an employer violates its FLSA statutory duty to keep proper records, and employees thereby have no way to establish the time spent doing uncompensated work, “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.” *Id.* at 687. “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.” *Id.* at 687–88. Following and applying the evidentiary burden-shifting procedure approved in *Mt. Clemens*, the *Tyson* Court found that plaintiffs “sought to introduce a representative sample to fill an evidentiary gap created by [Tyson’s] failure to keep adequate records.” *Tyson*, 136 S. Ct. at 1047. “Courts have frequently granted back wages under the FLSA to non-testifying employees based upon the representative testimony of a small percentage of the employees.” *Donavan v. Bel-Loc Diner, Inc.*, 780 F.2d 1113, 1116 (4th Cir. 1985), *abrogated on other grounds by* *McLaughlin v. Richard Shoe Co.*, 486 U.S. 128 (1988); see *Morgan v. Family Dollar Stores, Inc.*, 551 F.3d 1233, 1278–79 (11th Cir. 2008) (collecting cases); cf. *Day v. Celadon*

Plaintiffs in both *Wal-Mart* and *Tyson* proposed to use sampling evidence to prove classwide liability and damages, but both Supreme Court decisions were limited to classwide liability issues. Because the aggregate damages awarded at the *Tyson* trial had not been disbursed, and the record at the Supreme Court did not indicate how they would be disbursed, the *Tyson* Court deemed any consideration of damages issues premature.³⁵ Thus, the Supreme Court did not in *Tyson*, and has not to date, provided any specific guidance regarding the permissible use of sampling evidence to prove aggregate damages or individual damages in class actions.³⁶

C. Harmonizing Wal-Mart and Tyson

At first blush, it is difficult to reconcile the *Wal-Mart* and *Tyson* decisions. The salient facts are similar. In both cases, the putative class members' damages claims were too small to make separate adjudications feasible ("negative value cases"), and therefore the only realistic way for them to obtain relief was through a class action (or other aggregation procedure). In both cases, plaintiffs offered a sample average to determine aggregate damages.

The *Tyson* Court distinguished *Wal-Mart* by contrasting the heterogeneity of the two proposed classes: "While the experiences of the employees in *Wal-Mart* bore little relationship to one another, in this [*Tyson*] case each employee worked in the same facility, did similar work, and was paid under the same policy."³⁷ The *Tyson* Court characterized the "underlying question" in both cases as "whether the sample at issue could have been used

Trucking Servs., Inc., No. 15-1711, 2016 WL 3606682, at *8 (8th Cir. July 5, 2016) (applying *Mt. Clemens* evidentiary burden-shifting procedure in a Worker Adjustment and Retraining Notification ("WARN") Act case). Compare *Atis v. Freedom Mortg. Corp.*, No. 15-3424-RBK-JS, 2016 WL 7440465, at *6 (D.N.J. Dec. 27, 2016) (finding *Tyson* inapposite because the employer "had no statutory duty to track the hours worked" by exempt employees), with *Ridgeway v. Wal-Mart Stores, Inc.*, No. 08-cv-05221-SI, 2017 WL 363214, at *3 (N.D. Cal. Jan. 25, 2017) (stating a jury accepting "representative sampling [figures] used to reach an average amount of time spent on a certain activity was approved by the Supreme Court in *Tyson*").

Justice Thomas reads *Mt. Clemens* even more narrowly, concluding that "[a]ll *Mt. Clemens* decided was that the lack of precise data about the amount of time each employee worked was not fatal to their case," and that its holding is limited "to instances where the employer's FLSA violation was 'certain.'" 136 S. Ct. at 1057-59 (Thomas, J., dissenting) (citing *Mt. Clemens*, 328 U.S. at 688).

³⁵ 136 S. Ct. at 1049-50; see *supra* note 29 and accompanying text.

³⁶ See Bone, *supra* note 13, at 632 (opining that *Tyson* "opens the door to broader use of sampling but provides little guidance to lower court judges, who must decide when and how to use it"); Note, *Civil Procedure—Representative Evidence—Tyson Foods Inc. v. Bouaphakeo*, 130 HARV. L. REV. 407, 411 (2016) ("Though the Court reached a sound conclusion on the admissibility question, it failed to guide trial judges on how to handle representative studies going forward."); *id.* at 416 ("Trial judges would have benefited from some discussion of managing statistical evidence when it is admissible.").

³⁷ *Tyson*, 136 S. Ct. at 1048; see *Wal-Mart*, 564 U.S. at 359.

to establish liability in an individual action,” and, because the *Wal-Mart* putative class members were not “similarly situated,” the Court stated that there “would be little or no role for representative evidence” in any individual adjudication in that case.³⁸ The *Tyson* Court concluded: “Permitting the use of that sample in a class action, therefore, would have violated the Rules Enabling Act by giving plaintiffs and defendants different rights in a class proceeding than they could have asserted in an individual action.”³⁹ In short, plaintiffs’ sampling evidence was rejected in *Wal-Mart* because “the sample chosen was [not] representative of [the] population.”⁴⁰

As the *Wal-Mart* Court itself stated, the “crux” of the *Wal-Mart* case is “commonality,” that is, whether each class member suffered from a common corporate policy of discrimination.⁴¹ It is therefore a stretch to intimate that the “underlying question” in that case “was whether the sample at issue could have been used to establish liability in an individual action.”⁴²

However, *Wal-Mart* and *Tyson* can be harmonized. The principal learning of *Wal-Mart*, as interpreted by the *Tyson* Court, is that the admissibility of sampling evidence, “like all evidence,” “turns . . . on the degree to which the evidence is reliable in proving or disproving the elements of the relevant cause of action.”⁴³ Thus, there is no reason why reliable sampling evidence may not be introduced to prove both classwide liability and classwide damages, including individual damages.

In this Article, we focus on proving the element of classwide damages, and more particularly class members’ individual damages, and satisfying the predominance prerequisite of Rule 23(b)(3). As we discuss, to be admissible, sampling evidence must be relevant and reliable and satisfy *Daubert* and Federal Rule of Evidence 702 standards. One avenue to admissibility is to demonstrate that an individual class member could rely on the sampling evidence to prove liability or damages in an individual action, a proposition endorsed by the entire *Tyson* Court.⁴⁴ Before exploring each of these topics, we offer a brief history of attempts to use sampling evidence in class actions

³⁸ *Tyson*, 136 S. Ct. at 1048.

³⁹ *Id.*

⁴⁰ MANUAL FOR COMPLEX LITIGATION (FOURTH) § 11.493, at 103 (2004); see *Wal-Mart*, 564 U.S. at 357. The sampling evidence in *Wal-Mart* was also rejected because it purportedly would deprive *Wal-Mart* of its right “to litigate its [Title VII] statutory defenses to individual claims.” *Wal-Mart*, 564 U.S. at 367. Rejecting sampling evidence because it is not representative is logical. But, the latter proposition is a *non sequitur*. The *Tyson* Court did not explain how class heterogeneity detrimentally affects a class defendant’s right to interpose statutory defenses to individual claims.

⁴¹ 564 U.S. at 349.

⁴² *Tyson*, 136 S. Ct. at 1048.

⁴³ *Id.* at 1046 (citing FED. R. EVID. 401, 403, 702); see *Monroe v. FTS USA, LLC*, 860 F.3d 389, 399 (6th Cir. 2017) (“*Tyson* did not create a rule limiting representative evidence beyond the well-established standards of admissibility.”).

⁴⁴ See *supra* note 30 and accompanying text.

before *Wal-Mart* and *Tyson*, which may help to explain the prior judicial reluctance to accept it.⁴⁵

D. Brief Bleak History of Sampling Evidence in Federal Court Class Actions

Even before *Wal-Mart* and the Supreme Court's condemnation of "Trial by Formula," few federal courts embraced sampling evidence to determine classwide issues. Judicial reluctance to admit such evidence is explained, at least in part, not by any fault with this statistical methodology, but with its improper application.⁴⁶ A brief history is illuminating. The leading federal court cases before *Wal-Mart* and *Tyson* are discussed immediately below.

1. Cimino v. Raymark Industries

The tortuous history of mass tort asbestos litigation in the Eastern District of Texas and the Fifth Circuit is emblematic of the difficulties that trial courts face attempting to efficiently and fairly manage class action claims by using sampling. The "odyssey" began in 1986 when District Judge Robert M. Parker certified a Rule 23(b)(3) class for the trial of certain common issues, including the viability of the "state of the art" defense interposed by defendant asbestos manufacturers.⁴⁷ After an interlocutory appeal, the Fifth Circuit affirmed District Judge Parker's order to certify the Rule 23(b)(3) class.⁴⁸

District Judge Parker then issued pre-trial orders, consolidating over 3,000 personal injury cases for a single jury trial on the "state of the art" defense and punitive damages issues, pursuant to Fed. R. Civ. P. 42(a), and certifying a class "for the remaining issues of exposure and actual damages," pursuant to Fed. R. Civ. P. 23(b)(3).⁴⁹ The trial was planned to proceed in three phases. Phase I would follow the procedure approved in *Jenkins* to decide the "state of the art" defense and punitive damages issues.⁵⁰ Phase II, before the same jury, would include a full trial of liability and damages for the eleven class representatives, with additional evidence from thirty illustrative

⁴⁵ See *infra* notes 46–81 and accompanying text.

⁴⁶ There is ample guidance about the proper use of statistics. See, e.g., PAUL C. GIANNELLI ET AL., SCIENTIFIC EVIDENCE §§ 4–7 (5th ed. 2012); DAVID H. KAYE ET AL., THE NEW WIGMORE: EXPERT EVIDENCE §§ 12.1–12.10 (2014); JOSEPH M. McLAUGHLIN, McLAUGHLIN ON CLASS ACTIONS §§ 8:6–8:13 (14th ed. 2017); 4 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS §§ 11:1–11:21 (5th ed. 2014); *Reference Guide on Statistics*, *supra* note 32, at 211–302.

⁴⁷ *Jenkins v. Raymark Indus. Inc.*, 782 F.2d 468, 470–71 (5th Cir. 1986) (discussing viability of "state of the art" defense); *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 652 (E.D. Tex. 1990), *rev'd in part, vacated in part by* 151 F.3d 297 (5th Cir. 1998) (describing asbestos litigation as an "odyssey").

⁴⁸ *Jenkins*, 782 F.2d at 475.

⁴⁹ *In re Fibreboard Corp.*, 893 F.2d 706, 708 (5th Cir. 1990).

⁵⁰ *Id.*

class members, fifteen selected by plaintiffs and fifteen selected by defendants.⁵¹ In this Phase, the jury would hear expert evidence and determine aggregate damages for each subclass.⁵² Finally, in Phase III, “any awarded damages [would] be distributed utilizing various techniques.”⁵³

While “encourag[ing] the district court to continue its imaginative and innovative efforts to confront these cases,”⁵⁴ the Fifth Circuit vacated the order for the Phase II trial, and the associated Rule 23(b)(3) class certification order, because, among other things, “[t]here are too many disparities [e.g., different diseases, product identification, manners and degrees of exposure, lifestyle differences] . . . for their common concerns to predominate.”⁵⁵

On remand, the case proceeded as a Rule 23(b)(3) class action jury trial: Phase I on the issues of liability and punitive damages and “complete cases of ten class representatives.”⁵⁶ “After an eight week trial, the jury found for the [p]laintiffs and awarded nine of the class representatives approximately \$3.5 million in actual damages.”⁵⁷ The jury also found defendant asbestos manufacturers grossly negligent and assessed punitive damages.⁵⁸ The need for Phase II was obviated when the parties stipulated to asbestos “exposure” issues.⁵⁹

To resolve damages issues, in Phase III, almost 2,300 class members were divided into five disease categories (*i.e.*, subclasses) based on plaintiffs’ injury claims.⁶⁰ District Judge Parker selected a random sample of class members from each category for trial before a jury.⁶¹ He ruled that “[e]ach plaintiff whose damage case was submitted to the jury is to be awarded his individual verdict and the average verdict for each disease category will constitute the damage award for each non-sample class member.”⁶² The claims of the randomly selected 160 class members were tried over 133 days.⁶³ District Judge Parker ordered that each of the more than 2,100 remaining class members (“extrapolation cases”) would be entitled to

⁵¹ *Id.* at 709.

⁵² *Id.*

⁵³ *Id.* at 707.

⁵⁴ *Id.* at 712.

⁵⁵ *Id.*

⁵⁶ *Cimino v. Raymark Indus., Inc.*, 739 F. Supp. 328, 329 (E.D. Tex. 1990), *aff’d in part, vacated in part*, 151 F.3d 297 (5th Cir. 1998).

⁵⁷ *Id.* at 330.

⁵⁸ *Id.*

⁵⁹ *Cimino*, 751 F. Supp. at 654.

⁶⁰ *Id.* at 653; *see* FED. R. CIV. P. 23(c)(5) (“When appropriate, a class may be divided into subclasses that are each treated as a class under this rule.”).

⁶¹ *Cimino*, 751 F. Supp. at 653.

⁶² *Id.* Plaintiffs agreed to this procedure, waiving their rights to individual damages determinations. *See id.*; *Cimino*, 739 F. Supp. at 329.

⁶³ *Cimino*, 751 F. Supp. at 653. According to District Judge Parker, Phase III “utilizes the science of statistics, or more specifically, inferential statistics.” *Id.* at 659.

individual damages equal to the average of the verdicts in the sample cases in the same disease category.⁶⁴

Almost eight years after the trials, the Fifth Circuit reversed the judgments entered in the Phase III sample cases as well as the “extrapolation cases,” holding that, under the governing law of Texas, “causation must be determined as to individuals, not groups,” and that defendants had Seventh Amendment rights to have a jury make that determination.⁶⁵

2. *Hilao v. Estate of Marcos*

During the protracted interval between the trial court judgments in *Cimino* and their reversal by the Fifth Circuit, the District of Hawaii used inferential statistics to calculate individual damages for over 9,500 Rule 23(b)(3) class members in a class action against the Estate of Ferdinand Marcos, the former dictator of the Philippines, based on alleged violations of

⁶⁴ *Id.* at 664–65; *Cimino v. Raymark Indus., Inc.*, 151 F.3d 297, 300 (5th Cir. 1998).

⁶⁵ *Cimino*, 151 F.3d at 319 (internal quotation marks omitted). Significantly, the Fifth Circuit also observed that none of plaintiffs’ experts “purported to say that the damages suffered by the phase III plaintiffs in a given disease category (whether as disclosed by the phase III evidence or as found by the jury) were to any extent representative of the damages suffered by the extrapolation plaintiffs in the same disease category.” *Id.* at 320–21 nn.23–25, 48.

The three-judge Fifth Circuit panel in *Cimino* (Garza, Garwood and Davis, JJ.) distinguished the opinion and decision of another panel (Jones, DeMoss and Parker, JJ.) in *In re Chevron U.S.A., Inc.*, 109 F.3d 1016 (5th Cir. 1997), which had been issued during the almost eight-year hiatus between the federal district court sample trials and the issuance of the appellate decision in *Cimino*. Interestingly, *In re Chevron* was authored by then Circuit Judge Robert M. Parker, who had ordered and presided over the *Cimino* trials as a federal district court judge before his elevation to the Circuit Court.

In re Chevron involved claims by approximately 3,000 neighboring property owners for personal injuries and property damage allegedly caused by contamination from Chevron’s abandoned crude oil storage waste pit. 109 F.3d at 1017. The trial court directed that thirty individuals be chosen, fifteen by plaintiffs and fifteen by defendants, and that there be “a unitary trial on the issues of ‘general liability or causation’ on behalf of the remaining plaintiffs, as well as the individual causation and damage issues of the [thirty] selected plaintiffs.” *Id.* If the “unitary trial” established that Chevron was responsible for the “pollutants that, allegedly, give rise to all of the plaintiffs’ claims,” then individual causation and damages issues for the unchosen claimants would be determined subsequently at separate trials. *Id.* at 1019.

The *In re Chevron* Court (per Parker, J.) found that the thirty selected claimants were neither shown, nor chosen, to be representative of the other claimants, and opined that “[a] bellwether trial designed to achieve its value ascertainment function for settlement purposes or to answer troubling causation or liability issues common to the universe of claimants has as a core element representativeness.” *Id.* For a discussion of “bellwether” trials, see *infra* note 197 and accompanying text. The Fifth Circuit granted Chevron’s petition for mandamus prohibiting “utilization of the results obtained from the trial of the thirty (30) selected cases for any purpose affecting issues or claims of, or defenses to, the remaining untried cases.” *Id.* at 1021.

Later in its appellate decision in *Cimino*, the Fifth Circuit described the language in Circuit Judge Parker’s opinion “generally looking with favor on the use of bellwether verdicts when shown to be statistically representative,” see *In re Chevron*, 109 F.3d at 1019–20, as “plainly *dicta* . . . insofar as it might suggest that representative bellwether verdicts could properly be used to determine *individual* causation and damages for other plaintiffs.” *Cimino*, 151 F.3d at 318.

the Alien Tort Statute and the Tort Victim Protection Act.⁶⁶ The federal district court segregated the trial into three phases: liability, exemplary damages, and compensatory damages.⁶⁷ After finding liability and awarding exemplary damages, in the compensatory damages phase of the trial, the trial court “allowed the jury to consider” the “damages sustained by a random sample of plaintiffs as representative of damages suffered by the entire class.”⁶⁸ The Special Master appointed by the trial court reviewed the depositions of 137 randomly selected claimants.⁶⁹ He reported to the jury “recommending the damages suffered by the 137 claimants, to give the jury a statistically valid representation of damages suffered by the entire class.”⁷⁰ The trial court cited and followed District Judge Parker’s trial plan in *Cimino* as “precedent,” dividing the class members into three subclasses based upon each plaintiff’s claims, and selecting a random sample from the class population.⁷¹ Each plaintiff who was a member of the random sample, and who had a valid claim, was awarded individual damages.⁷² The average verdict for each subclass was awarded to the non-sample class members.⁷³

On appeal, the Estate’s challenge to the trial procedure adopted by the trial court was “very narrow”: “[i]t challenges specifically only ‘the method by which [the district court] allowed the validity of the class claims to be determined’: the master’s use of a representative sample to determine what percentage of the total claims were invalid.”⁷⁴ To the Ninth Circuit, the grounds for the Estate’s “challenges” were “unclear” and “poorly presented,” but, still, the “Estate’s due-process claim does raise serious questions.”⁷⁵ Applying the due process balancing test set forth in *Connecticut v. Doe*⁷⁶ and *Matthews v. Eldridge*,⁷⁷ the Ninth Circuit weighed the Estate’s “only interest,” the “total amount of damages for which it will be liable,” with the

⁶⁶ *In re Estate of Ferdinand E. Marcos Human Rights Litig.*, 910 F. Supp. 1460 (D. Haw. 1995); see Alien Tort Statute, 28 U.S.C. § 1350 (2012).

⁶⁷ *In re Marcos*, 910 F. Supp. at 1462.

⁶⁸ *Id.* at 1462, 1464.

⁶⁹ *Id.*; see FED. R. EVID. 706 (Court-Appointed Expert Witnesses).

⁷⁰ *In re Marcos*, 910 F. Supp. at 1464. It is noteworthy that defendant Estate “did not appear at any of the depositions” or “depose any of the . . . class members to test the procedure employed by the Court, or to acquire evidence to refute the fairness to the defendant of this random selection process using inferential statistical methodology.” *Id.* at 1465.

⁷¹ *Id.* at 1467.

⁷² *Id.*

⁷³ *Id.*

⁷⁴ *Hilao v. Estate of Marcos*, 103 F.3d 767, 784 (9th Cir. 1996) (alteration in the original). The Ninth Circuit stated that District Judge Parker’s opinion in *Cimino* “apparently helped persuade the district court [of Hawaii] to use this method.” *Id.*; see *Cimino*, 751 F. Supp. at 659–67.

⁷⁵ *Hilao*, 103 F.3d at 784–85. The Ninth Circuit distinguished the Fifth Circuit’s decision in *In re Fibreboard Corp.* on the ground that it was based on Texas substantive law. *Id.*; see *In re Fibreboard Corp.*, 893 F.2d 706, 711–12 (5th Cir. 1990).

⁷⁶ 501 U.S. 1, 10 (1991).

⁷⁷ 424 U.S. 319, 335 (1976).

class members' "enormous" interest in the sampling methodology, "since adversarial resolution of each class member's claim would pose insurmountable practical hurdles," and with the judiciary's "ancillary" interest" in avoiding "individual adversarial determinations of claim validity [that] would clog the docket of the district court for years."⁷⁸ It held that the trial court's "unorthodox" statistical "methodology in determining valid claims" passed due process muster.⁷⁹

Significantly, the Ninth Circuit's *Hilao* decision pre-dated the Fifth Circuit's decision in *Cimino*, which rejected District Judge Parker's trial plan that had "persuad[ed]" the District of Hawaii to use a sampling methodology.⁸⁰ Moreover, it is doubtful that the vitality of the *Hilao* sampling methodology survives *Wal-Mart*.⁸¹

Suffice it to say, the sampling methodologies proposed in *Cimino* and *Hilao* did not present sampling evidence in its best light. *Tyson* may represent

⁷⁸ *Hilao*, 103 F.3d at 786–87.

⁷⁹ *Id.*

⁸⁰ See *Cimino* 151 F.3d at 297; *Hilao* 103 F.3d at 767. Another case that raised statistical evidence issues is *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris, Inc.*, 178 F. Supp. 2d 198 (E.D.N.Y. 2001), *rev'd. in part*, 344 F.3d 211 (2d Cir. 2003). In that case, plaintiffs' experts used a random sample of subscriber depositions "to extrapolate statistically meaningful inferences about the plaintiff subscriber population as a whole" and "statistical evidence about the effect of the defendants' misleading statements on smoking behavior." *Id.* at 226–28.

On appeal, the Second Circuit found "without merit" defendants' arguments that plaintiffs' reliance on "aggregate (*i.e.*, statistical) proof of causation and damages" violated their constitutional rights to a jury trial and to due process, but certified to the New York Court of Appeals the question of whether, under the facts of the case, New York's consumer protection statute, *N.Y. Gen. Bus. L.* § 349, requires individual proof of harm. *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris USA, Inc.*, 344 F.3d 211, 225–28 (2d Cir. 2003). The New York Court of Appeals did not reach that question because it answered in the affirmative another certified question: whether the claims of a third-party payer of health care costs, such as *Blue Cross & Blue Shield*, were too remote to be recovered derivatively under the statute, thereby mooting the certified individual proof of harm question. *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris USA, Inc.*, 818 N.E. 2d 1140, 1146 (N.Y. 2004). As a result of the New York Court of Appeals response to that certified question, the Second Circuit reversed the jury trial verdict on a ground that was not related to the expert statistical evidence introduced at trial. *Empire Healthchoice, Inc. v. Philip Morris USA, Inc.*, 393 F.3d 312 (2d Cir. 2004).

Subsequently, in a putative Rule 23(b)(3) class action based on claims under the Racketeer Influenced and Corrupt Organizations Act ("RICO"), 18 U.S.C. §§ 1961–1968 (2012), the Second Circuit ruled that common issues did not predominate over individual issues of reliance, injury, and damages. *McLaughlin v. Am. Tobacco Co.*, 522 F.3d 215 (2d Cir. 2008). More to the point, the court rejected, as violative of the Rules Enabling Act and the Due Process Clause, plaintiffs' "fluid recovery" proposal whereby "defendant's aggregate liability is determined in a single, class-wide adjudication" and "individual class members are afforded an opportunity to collect their individual shares, 'usually through a simplified proof of claim procedure.'" *Id.* at 231–33. The court held that "fluid recovery" could not be used to "mask the prevalence of individual issues." *Id.* at 232.

⁸¹ See *supra* note 23. The Supreme Court's discussion of "Trial by Formula" in *Wal-Mart* does not refer to *Hilao*, but the majority noted that the Ninth Circuit approved the trial plan in *Wal-Mart* in reliance on *Hilao*. See *Wal-Mart*, 564 U.S. at 348; *Dukes v. Wal-Mart Stores, Inc.*, 603 F.3d 571, 627–28 (9th Cir. 2010), *rev'd*, 564 U.S. 338 (2011).

the advent of a more open-minded view by the Supreme Court toward the proper use of sampling evidence in class actions.

We now turn to the standards governing the admissibility of sampling evidence in class actions.

II. SAMPLING EVIDENCE AND CLASS CERTIFICATION DISPUTES

A. Admissibility of Expert Evidence

Daubert v. Merrell Dow Pharmaceuticals established new standards for the admissibility of expert evidence in federal courts.⁸² Further, because a decision to grant or deny a motion for class certification is often outcome-determinative of an entire case, and a class certification ruling is frequently based on expert evidence, expert evidence is now *de rigueur* in class certification disputes and *Daubert* and its progeny are often on center stage.⁸³

⁸² 509 U.S. 579, 585–89 (1993) (holding the “general acceptance” test of *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923), for the admissibility of scientific evidence that had been applied by many federal and state courts had been superseded by the enactment of the Federal Rules of Evidence in 1975, specifically Rule 702); see ERIC R. HARLAN ET AL., AM. BAR ASS’N SECTION OF LITIG. TRIAL EVIDENCE COMM., *DAUBERT V. FRYE: A FIFTY-STATE SURVEY OF THE ADMISSIBILITY OF EXPERT TESTIMONY* (2016).

⁸³ See generally Robert H. Klonoff, *The Decline of Class Actions*, 90 WASH. U. L. REV. 729 (2013) (arguing that, in recent years, the ability to bring class action lawsuits has become problematic).

For a class action defendant, “[c]ertification of a large class may so increase the defendant’s potential damages liability and litigation costs that he may find it economically prudent to settle and to abandon a meritorious defense.” *Coopers & Lybrand v. Livesay*, 437 U.S. 463, 476 (1978); see Richard A. Nagareda, *Class Certification in the Age of Aggregate Proof*, 84 N.Y.U. L. REV. 97, 99 (2009) (“With vanishingly rare exception, class certification sets the litigation on a path toward resolution by way of settlement”); see also, e.g., *Castano v. Am. Tobacco Co.*, 84 F.3d 734, 746 (5th Cir. 1996) (“The risk of facing an all-or-nothing verdict presents too high a risk, even when the probability of an adverse judgment is low.”); *In re Am. Med. Sys., Inc.*, 75 F.3d 1069, 1086–87 & n.19 (6th Cir. 1996); *In re Rhone-Poulenc Rorer Inc.*, 51 F.3d 1293, 1298 (7th Cir. 1995) (noting that defendants facing large classwide damages awards “will be under intense pressure to settle”).

For a class action plaintiff, “a refusal to certify a class . . . may induce a plaintiff to abandon his individual claim.” *Coopers & Lybrand*, 437 U.S. at 470 (discussing interlocutory appeals and “death knell” doctrine); MANUAL FOR COMPLEX LITIGATION (FOURTH) § 11.213, at 40 (2004) (“Denial of class certification may effectively end the litigation.”).

Note that *Coopers & Lybrand* was decided before the 1998 adoption of Rule 23(f) authorizing discretionary interlocutory appellate review of class certification rulings. Following an instruction by the Advisory Committee on Civil Rules (“[t]he courts of appeal will develop standards for granting review” under Rule 23(f)), Circuit Courts have established general standards to decide whether a class certification ruling warrants interlocutory review. FED. R. CIV. P. 23 advisory committee’s notes to 1998 amendment; see, e.g., *Vallario v. Vandehey*, 554 F.3d 1259, 1263–64 (10th Cir. 2009); *In re Lorazepam & Clorazepate Antitrust Litig.*, 289 F.3d 98, 102–06 (D.C. Cir. 2002); *Newton v. Merrill Lynch, Pierce, Fenner & Smith, Inc.*, 259 F.3d 154, 165 (3d Cir. 2001); *Blair v. Equifax Check Servs.*, 181 F.3d 832, 834–35 (7th Cir. 1999).

Case studies have suggested that “aggregation of related cases at a very young stage may benefit defendants, while aggregation of related cases in an ‘adolescent’ stage (i.e., after plaintiffs have achieved early breakthrough victories) may favor plaintiffs. Moreover, experimental studies

After *Daubert*, a federal district court acts as the “gatekeep[er]” to ensure that, to be admissible, proffered expert evidence is “not only relevant, but reliable.”⁸⁴

Evidence is relevant if “it has any tendency to make a fact more or less probable than it would be without the evidence” and “the fact is of consequence in determining the action.”⁸⁵ “The Rules’ basic standard of relevance thus is a liberal one.”⁸⁶ Relevant evidence is generally admissible.⁸⁷ Pursuant to Rule 702(a) of the Federal Rules of Evidence, expert evidence is admissible if it “will help the trier of fact to understand the evidence or to determine a fact in issue.”⁸⁸ Therefore, expert evidence must be based on the facts of the case: “Rule 702’s ‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.”⁸⁹ In general, the Federal Rules of Evidence embody a “liberal thrust” and a “general approach of relaxing the traditional barriers to ‘opinion’ testimony.”⁹⁰

To be reliable, the subject of an expert’s testimony must be “scientific knowledge.”⁹¹ “[S]cientific” implies a grounding in the methods and procedures of science,” and “knowledge” . . . applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds.”⁹² To qualify as “scientific knowledge,” therefore, an inference based on statistics must be “derived by the scientific method,” and related

have shown that aggregation helps plaintiffs with weak cases and hurts plaintiffs with strong cases in terms of damages, and have generally (though not entirely) pro-plaintiff effects in terms of liability.” JAY TIDMARSH & ROGER TRANGSRUD, *COMPLEX LITIGATION: PROBLEMS IN ADVANCED CIVIL PROCEDURE* 11 (2002) (citations omitted); see BARBARA J. ROTHSTEIN & THOMAS E. WILLGING, *FED. JUDICIAL CTR., MANAGING CLASS ACTION LITIGATION: A POCKET GUIDE FOR JUDGES* 6 (2005) (finding that 90% of class actions settle); Thomas E. Willging & Shannon R. Wheatman, *Attorney Choice of Forum in Class Action Litigation: What Difference Does It Make?*, 81 NOTRE DAME L. REV. 591, 606–07 (2006) (noting that, in an empirical study of state and federal class actions, “all certified class actions settled on a class-wide basis”).

⁸⁴ *Daubert*, 509 U.S. at 589, 597; see FED. R. EVID. 401, 702. The trial “court must decide any preliminary questions about whether . . . evidence is admissible.” FED. R. EVID. 104(a).

⁸⁵ FED. R. EVID. 401.

⁸⁶ *Daubert*, 509 U.S. at 587.

⁸⁷ See FED. R. EVID. 402.

⁸⁸ FED. R. EVID. 702(a).

⁸⁹ *Daubert*, 509 U.S. at 591–92. In explaining the required relevancy of expert testimony, the *Daubert* Court adopted the characterization of Circuit Judge Edward R. Becker in *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985), and *United States v. Downing*, 609 F. Supp. 784, 791–92 (E.D. Pa.), *aff’d*, 780 F.2d 1017 (3d Cir. 1985), “as one of ‘fit.’” 509 U.S. at 591. The relevancy inquiry is: Does the expert evidence “fit” the case?

⁹⁰ *Daubert*, 509 U.S. at 588 (internal quotation marks omitted) (quoting *Beech Aircraft Corp. v. Rainey*, 488 U.S. 153, 169 (1988)).

⁹¹ *Id.* at 589–90.

⁹² *Id.* at 590 (internal quotation marks omitted) (quoting WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 1252 (1986)).

expert statistical evidence “must be supported by appropriate validation.”⁹³ This “establishes a standard of evidentiary reliability.”⁹⁴

The *Daubert* Court noted “that scientists typically distinguish between ‘validity’ (does the principle support what it purports to show?) and ‘reliability’ (does application of the principle produce consistent results?).”⁹⁵ The Court emphasized that it was referring to “evidentiary reliability—that is, trustworthiness,” and instructed, “[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*.”⁹⁶

Thus, to discharge its responsibility as “gatekeeper,” a federal district court must determine “whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or

⁹³ *Id.*

⁹⁴ *Id.*

⁹⁵ *Id.* at 590 n.9 (citing Bert Black, *A Unified Theory of Scientific Evidence*, 56 FORDHAM L. REV. 595, 599 (1988)). According to the author of the cited article, scientific validity depends on the acceptance of the “scientific practice and the soundness and cogency of the entire pattern of reasoning leading to the expert’s conclusion.” Black, *supra* at 600. He continues, “[w]idespread consensus and acceptance, therefore, is the central test that scientists use to decide the validity of theories and reasoning in any given context, which is a test that the law can adopt and use successfully.” *Id.* at 601 (footnotes omitted). The citation to this article is curious in light of the *Daubert* Court’s rejection of the *Frye* test.

⁹⁶ *Daubert*, 509 U.S. at 590 n.9. In *Kumho Tire Co. v. Carmichael*, the Supreme Court unanimously held that a federal district court’s “gatekeeper” role encompasses all expert evidence, not only “scientific” evidence, pointing out that Rule 702 “makes no relevant distinction between ‘scientific’ knowledge and ‘technical’ or ‘other specialized’ knowledge,” and “applies its reliability standard to all . . . matters within its scope.” 526 U.S. 137, 147 (1999). As the Supreme Court recognized in a decision after *Daubert*, under Rule 702, all expert evidence must comply with “exacting standards of reliability.” *Weisgram v. Marley Co.*, 528 U.S. 440, 455 (2000).

Compare the *Daubert* Court’s explanation with the discussion regarding reliability and validity of statistics in the *Reference Guide on Statistics*:

Reliability and validity are two aspects of accuracy in measurement. In statistics, reliability refers to reproducibility of results. A reliable measuring instrument returns consistent measurements. . . .

Reliability can be ascertained by measuring the same quantity several times; the measurements must be made independently to avoid bias. Given independence, the correlation coefficient . . . between repeated measurements can be used as a measure of reliability. . . .

Two different aspects of reliability should be considered. First, the “within-observer variability” of judgments should be small—the same evaluator should rate essentially identical cases in similar ways. Second, the “between-observer variability” should be small—different evaluators should rate the same cases in essentially the same way. . . .

Reliability is necessary but not sufficient to ensure accuracy. In addition to reliability, validity is needed. A valid measuring instrument measures what it is supposed to. . . .

When there is an established way of measuring a variable, a new measurement process can be validated by comparison with the established one. . . . A common measure of validity is the correlation coefficient between the predictor and the criterion. . . .

determine a fact in issue.”⁹⁷ As the *Daubert* Court explained, “[t]his entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.”⁹⁸ “Pertinent evidence based on scientifically valid principles will satisfy those demands.”⁹⁹

The *Daubert* Court offered federal courts five non-exclusive factors to guide the “flexible” “inquiry envisioned by Rule 702” regarding whether “a theory or technique is scientific knowledge.”¹⁰⁰ To assess whether a “theory or technique” exhibits scientific validity, federal courts should consider: “whether it can be (and has been) tested” (“[o]rdinarily, a key question”); “whether [it] has been subjected to peer review and publication”; “known or potential rate of error”; “existence and maintenance of standards controlling [its] operation”; and “general acceptance.”¹⁰¹

In response to *Daubert* and the many cases applying it, Rule 702 was amended in 2000.¹⁰² As pointed out by the Advisory Committee on Civil Rules, that amendment “provides some general standards that the trial court *must* use to assess the reliability and helpfulness of proffered expert testimony.”¹⁰³ Those mandatory standards are: “the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue”; “the testimony is based on sufficient facts or data”; “the testimony is the product of reliable principles and methods”; and “the expert has reliably applied the principles and methods to the facts of the case.”¹⁰⁴

⁹⁷ *Daubert*, 509 U.S. at 592.

⁹⁸ *Id.* at 592–93.

⁹⁹ *Id.* at 597.

¹⁰⁰ *Id.* at 593–94.

¹⁰¹ *Id.* at 593–94. In *General Elec. Co. v. Joiner*, a unanimous Supreme Court held that an appellate court should apply an abuse of discretion standard when reviewing a federal district court’s *Daubert* evidentiary ruling to admit or exclude expert testimony, whether or not the ruling is “outcome determinative.” 522 U.S. 136, 141–43, 146 (1997).

A party will likely have only one chance to pass by the federal district court “gatekeeper.” In *Weisgram*, plaintiff won a jury verdict. The Eighth Circuit, however, held that plaintiff’s expert’s testimony was speculative and should have been excluded, and granted judgment as a matter of law for defendants. *Weisgram*, 528 U.S. at 455. Plaintiff argued that he should have a right to a new trial and to introduce additional expert evidence, but the Supreme Court rejected the argument: “Since *Daubert* . . . parties relying on expert evidence have had notice of exacting standards of reliability such evidence must meet. . . . It is implausible to suggest, post-*Daubert*, that parties will initially present less than their best expert evidence in the expectation of a second chance should their first try fail.” *Id.*

¹⁰² FED. R. EVID. 702. Rule 702 was amended again in 2011: “These changes are intended to be stylistic only. There is no intent to change any result in any ruling on evidence admissibility.” FED. R. EVID. 702 advisory committee’s notes to 2011 amendment.

¹⁰³ FED. R. EVID. 702 advisory committee’s notes to 2000 amendment (emphasis added).

¹⁰⁴ FED. R. EVID. 702.

The Advisory Committee also enumerated factors to add to the *Daubert* Court's "non-exclusive checklist" that are "relevant in determining whether expert testimony is sufficiently reliable to be considered by the trier of fact," including:

- (1) Whether experts are proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying.
- (2) Whether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion.
- (3) Whether the expert has adequately accounted for obvious alternative explanations.
- (4) Whether the expert is being as careful as he would be in his regular professional work outside his paid litigation consulting.
- (5) Whether the field of experience claimed by the expert is known to reach reliable results for the type of opinion the expert would give.¹⁰⁵

The 2000 amendment to Rule 702 expressly provides that a federal district court "must scrutinize not only the principles and methods used by the expert, but also whether those principles and methods have been properly applied to the facts of the case."¹⁰⁶

B. Admissibility of Sampling Evidence

As the 2015 edition of the Manual for Complex Litigation (Fourth) succinctly summarizes, when "sampling to generate data about a population so the data will be verified or declared true," "the reliability and validity of estimates about the population derived from sampling are critical."¹⁰⁷ Relevant factors to determine whether "sampling methods . . . conform to generally recognized statistical standards" include: "the population was properly chosen and defined"; "the sample chosen was representative of that

¹⁰⁵ FED. R. EVID. 702 advisory committee's notes to 2000 amendment (internal quotation marks and citations omitted).

¹⁰⁶ *Id.* As the *Daubert* Court made clear, "[t]he focus, of course, must be solely on principles and methodology, not on the conclusions that they generate." 509 U.S. at 595. That said, "conclusions and methodology are not entirely distinct from one another . . . [N]othing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered." *Joiner*, 522 U.S. at 146; see *ZF Meritor, LLC v. Eaton Corp.* 696 F.3d 254, 291 (3d Cir. 2012) (stating that reliability "applies to all aspects of an expert's testimony: the methodology, the facts underlying the expert's opinion, [and] the link between the facts and the conclusion" (quoting *Heller v. Shaw Indus., Inc.*, 167 F.3d 146, 155 (3d Cir. 1999))).

¹⁰⁷ MANUAL FOR COMPLEX LITIGATION (FOURTH) § 11.493, at 102 (2015).

population”; “the data gathered were accurately reported”; and “the data were analyzed in accordance with accepted statistical principles.”¹⁰⁸ Adhering to these standards enables the proponent of sampling evidence to convincingly argue that the proffered evidence is “scientific knowledge” and “will assist the trier of fact to understand or determine a fact in issue,” and is, consequently, admissible.¹⁰⁹ And, as stated in the *Reference Guide on Statistics*:

Statistical studies suitably designed to address a material issue generally will be admissible under the Federal Rules of Evidence. . . . Because most statistical methods relied on in court are described in textbooks or journal articles and are capable of producing useful results when properly applied, these methods generally satisfy important aspects of the “scientific knowledge” requirement in [*Daubert*]. . . . Of course, a particular study may use a method that is entirely appropriate but that is so poorly executed that it should be inadmissible under Federal Rules of Evidence 403 and 702.¹¹⁰ Or, the method may be inappropriate for the problem at hand and thus lack the “fit” spoken of in *Daubert*. Or the study might rest on data of the type not reasonably relied on by statisticians or substantive experts and hence run afoul of Federal Rules of Evidence 703.¹¹¹ Often, however, the battle over statistical evidence concerns weight or sufficiency rather than admissibility.¹¹²

¹⁰⁸ *Id.* at 103.

¹⁰⁹ *Daubert*, 509 U.S. at 592. The burden is on the proponent of the expert evidence to show that it meets all *Daubert* requirements. *Id.*

¹¹⁰ FED. R. EVID. 403 (internal footnote supplied). The Rule, titled “Excluding Relevant Evidence for Prejudice, Confusion, Waste of Time, or Other Reasons,” states: “The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.” *Id.*

¹¹¹ FED. R. EVID. 703 (internal footnote supplied). The Rule, titled “Bases of an Expert,” states:

An expert may base an opinion on facts or data in the case that the expert has been made aware of or personally observed. If experts in the particular field would reasonably rely on those kinds of facts or data in forming an opinion on the subject, they need not be admissible for the opinion to be admitted. But if the facts or data would otherwise be inadmissible, the proponent of the opinion may disclose them to the jury only if their probative value in helping the jury evaluate the opinion substantially outweighs their prejudicial effect.

Id.

¹¹² *Reference Guide on Statistics*, *supra* note 32, at 214 (original footnotes omitted).

C. Generally Accepted Statistical Standards

An expert statistician can randomly select a sample from a population of class members' damages claims for individual adjudications and estimate the average value of the claims. "The precision of an estimate is usually reported in terms of the standard error and a confidence interval."¹¹³

An estimate based on a sample of class members' damages claims may be inaccurate because of random error. The standard error measures the "likely magnitude of this random error."¹¹⁴ The standard deviation measures heterogeneity.¹¹⁵ "The less heterogeneity in the values, the smaller the standard error."¹¹⁶

"Usually, a confidence interval for the population average is centered at the sample average; the desired confidence level is obtained by adding and subtracting a suitable multiple of the standard error."¹¹⁷ Many statisticians use ± 1.96 standard errors to arrive at a 95% confidence interval.¹¹⁸ Using this heuristic, a statistician who opines that the population average is "within 2 standard errors" of the sample average will be correct about 95% of the time.¹¹⁹

In a sampling study, it is thus possible that an observation (*e.g.*, estimated average value) is due to sampling error.¹²⁰ An alternative (or complement) to using confidence intervals to express the uncertainty of an estimate, or to make decisions notwithstanding the uncertainty of an estimate, is to use *p*-values. "The *p*-value is the probability of getting data as extreme as, or more extreme than, the actual data—given that the null hypothesis [or default hypothesis, such as 'no difference' between two populations] is true."¹²¹ If the *p*-value of the observed result is less than a defined significance level, it may be probabilistically concluded that the null hypothesis should be rejected in favor of an alternative hypothesis.¹²² Typically, a significance level of five percent is chosen. Below this threshold,

¹¹³ *Id.* at 241 (footnotes omitted).

¹¹⁴ *Id.* at 243.

¹¹⁵ *Id.* Statisticians distinguish between "the standard error of the sample average, which measures the likely size of the random error in the sample average," and "the standard deviation of the sample, which measures the spread in the sample data." *Id.* at 243 n.87.

¹¹⁶ *Id.* at 243.

¹¹⁷ *Id.* at 244.

¹¹⁸ *Id.* at 244 n.89.

¹¹⁹ *Id.* at 244.

¹²⁰ See EARL R. BABBIE, *THE PRACTICE OF SOCIAL RESEARCH* 185–226 (13th ed. 2013) (discussing survey sampling); VINCENT FLAHERTY, *COMPASSIONATE STATISTICS: APPLIED QUANTITATIVE ANALYSIS OF SOCIAL SCIENCES* 127–38 (2008) (discussing probability and statistical significance).

¹²¹ *Reference Guide on Statistics*, *supra* note 32, at 250.

¹²² See R. MARK SIRKIN, *STATISTICS FOR THE SOCIAL SCIENCES* 271–316 (3d ed. 2005).

the observed result is called “statistically significant,” that is, not due to sampling error.¹²³

Using these generally accepted statistical concepts (standard errors, confidence intervals, and *p*-values) in measuring the accuracy of an estimated average value of individual damages claims in a class action would easily satisfy the “scientific knowledge” and “fit” requirements of *Daubert* and Rules 702 and 703. However, these concepts do not recommend a methodology to follow to satisfy the Rule 23(b)(3) predominance prerequisite where, as is often the case, individual class member’s damages claims vary.

Recalling that the Supreme Court distinguished the classes (and decisions) in *Wal-Mart* and *Tyson* based on the degree of heterogeneity within the employee populations, it is clear that the Court was concerned about the reliability of the sampling methodologies employed by plaintiffs’ experts in both cases. In *Wal-Mart*, the Court found that the employees were not “similarly situated” and, therefore, “none of them could have prevailed in an individual suit by relying on” evidence relating to other employees.¹²⁴ In *Wal-Mart*, therefore, “there would be little or no role for representative evidence.”¹²⁵ “In contrast,” the “circumstances” in *Tyson* (“each employee worked in the same facility, did similar work, and was paid under the same policy”) “confirms” that “experiences of a subset of employees can be probative as to the experiences of all of them.”¹²⁶ Consequently, in *Tyson*, the

¹²³ See JAY L. DEVORE, PROBABILITY AND STATISTICS FOR ENGINEERING AND THE SCIENCES 300–44 (8th ed. 2011) (discussing tests for hypotheses based on sampling). According to a leading law and economics jurist, recently retired Judge Richard A. Posner of the Seventh Circuit, “statistically significant at the five percent level” means “that the probability that the investigation would have yielded this result even if the hypothesis that it was trying to test was false is no greater than five percent.” Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1510–11 (1999) (footnote omitted). He points out that social scientists consider results at the two percent or one percent level “more robust” and “highly significant.” *Id.* at 1511. Judge Posner continues:

What is true is that the higher the significance level of a statistical study, the more reliable the study is as evidence; and the lower that level, the less reliable it is. A low significance level may reflect an unsound method of statistical estimation, an incorrect specification of the hypothesis being tested, or the omission of relevant variables that if included would have caused the hypothesis to be rejected. . . . But if the study has been conducted responsibly and has withstood a hammering from the opponent’s expert, failure to reach the conventional five percent significance level would not be a good reason for excluding the evidence just because a social scientist who violated the conventions of his discipline by reporting results that do not attain the conventional significance level might be considered untrustworthy.

Id. But, Judge Posner concludes that, given the “cost of weak statistical evidence,” and “[g]iven the difficulty that judges and jurors have in understanding and weighing statistical evidence, there is an argument . . . for excluding statistical evidence that the relevant profession . . . considers weak.” *Id.* at 1512.

¹²⁴ *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1048 (2016).

¹²⁵ *Id.*

¹²⁶ *Id.*

Court held that it was proper to draw a “just and reasonable inference” from the statistical evidence to prove the “amount and extent of” work performed by each class member, and, as Chief Justice Roberts stated in his concurring opinion, the “just and reasonable inference” standard of proof “would apply in any case.”¹²⁷

In a recent article, one author summarized the evidentiary issue in this manner:

The threshold question for admitting such evidence is therefore not about whether the evidence is statistical or representative more broadly, or even about whether direct evidence, if it existed, would exhibit any differences across plaintiffs. Rather, the threshold question that evidence law poses for any such evidence is whether it is relevant [as Rule 401 defines relevance]. . . . When the evidence in question involves expert testimony, then Rule 702 and *Daubert* also come into play.¹²⁸

In short, the more heterogeneous the individual damages claims of the class members are from one another, the less confident one can be that the sample average can be accurately applied to calculate individual damages.

In Part IV of this Article, we propose a method to satisfy Rule 23(b)(3) predominance, notwithstanding variable individual damages claims. Specifically, we propose a statistically-based threshold for admissibility of sampling evidence by showing that, when judgment variability exceeds claim variability (two terms we define), applying the mean of damages awards for sampled class members’ claims, rather than individual damages awards, improves accuracy. We demonstrate that, in many circumstances, sampling evidence is more accurate and more reliable than individual damages awards and, therefore, admissible.

D. Daubert and Class Certification Proceedings

Daubert has played a critical, often decisive, role in class certification proceedings. As prerequisites to class certification, plaintiffs must show, among other things, that “there are questions of law or fact *common* to the class” and, if plaintiffs seek classwide damages, that “the questions of law or fact common to class members *predominate* over any questions affecting only individual members, and that a class action is *superior* to other available

¹²⁷ *Id.* at 1047 (internal quotation marks and citations omitted); *id.* at 1051 (Roberts, C.J., concurring) (internal quotation marks and citations omitted).

¹²⁸ Gelbach, *supra* note 9, at 1830–31. Professor Gelbach opines that the “critical question for any representative evidence” is whether “those who testify or form the sample used to create statistical evidence are similar enough in relevant ways to reliably measure” the disputed value for an individual class member. *Id.* at 1831.

methods for fairly and efficiently adjudicating the controversy.”¹²⁹ With increasing frequency, expert evidence is proffered during class certification proceedings to attempt to satisfy or defeat the commonality, predominance, and superiority prerequisites. With equally increasing frequency, *Daubert* challenges are made to the proffered expert evidence.

In *Comcast*, although the Supreme Court granted a petition for a writ of certiorari to review, among other questions presented for review, the question of whether *Daubert* standards for the admissibility of expert evidence apply in class certification proceedings, the Court did not reach that question because, at trial, Comcast failed to “timely object to or move to strike” plaintiffs’ expert’s damages model.¹³⁰ However, a majority of Circuit Courts that have considered the question, both before and after *Comcast*, have held that a full *Daubert* analysis of expert evidence is required at the class certification stage.¹³¹

In dicta, the *Wal-Mart* Court signaled its view: “[T]he District Court concluded that *Daubert* did not apply to expert testimony at the certification stage of class-action proceedings. . . . *We doubt that is so . . .*”¹³² Therefore, the clear weight of authority is that expert evidence will be scrutinized under

¹²⁹ FED. R. CIV. P. 23(a)(2), 23(b)(3) (emphasis added); see, e.g., *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 362 (2011) (“[W]e think it clear that individualized monetary claims belong in Rule 23(b)(3).”).

¹³⁰ *Comcast Corp. v. Behrend*, 569 U.S. 27, 39–40 (2013) (Ginsburg and Breyer, JJ., dissenting); see FED. R. EVID. 103(a)(1)(A) (“A party may claim error in a ruling to admit . . . evidence only if the error affects a substantial right of the party and . . . [the party] on the record . . . timely objects or moves to strike . . .”).

¹³¹ See, e.g., *In re Blood Reagents Antitrust Litig.*, 783 F.3d 183, 188 (3d Cir. 2015); *Messner v. Northshore Univ. Health Sys.*, 669 F.3d 802, 812 (7th Cir. 2012); *Ellis v. Costco Wholesale Corp.*, 657 F.3d 970, 982 (9th Cir. 2011); *Sher v. Raytheon Co.*, 419 F. App’x 887, 890–91 (11th Cir. 2011); *In re Scrap Metal Antitrust Litig.*, 527 F.3d 517, 528–32 (6th Cir. 2008); *Unger v. Amedisys Inc.*, 401 F.3d 316, 323 (5th Cir. 2005). The Eighth Circuit represents the minority view. See, e.g., *Smith v. Conoco Phillips Pipe Line Co.*, 801 F.3d 921, 925 n.2 (8th Cir. 2015); *In re Zurn Pex Plumbing Prods. Liab. Litig.*, 644 F.3d 604, 614 (8th Cir. 2011) (ruling that federal district court “did not err by conducting a focused *Daubert* analysis” at class certification stage). See generally L. Elizabeth Chamblee, *Between “Merit Inquiry” and “Rigorous Analysis”: Using Daubert to Navigate the Gray Areas of Federal Class Action Certification*, 31 FLA. ST. U. L. REV. 1041 (2004); George G. Gordon & Irene Ayzenberg-Lyman, *The Role of Daubert in Scrutinizing Expert Testimony in Class Certification*, 82 GEO. WASH. L. REV. ARGUENDO 132 (2014), http://www.gwlr.org/wp-content/uploads/2014/10/Gordon_Final_Redacted.pdf [https://perma.cc/KDC8-T2PH]; Richard L. Marcus, *Reviving Judicial Gatekeeping of Aggregation: Scrutinizing the Merits on Class Certification*, 79 GEO. WASH. L. REV. 324 (2011); Meredith M. Price, *The Proper Application of Daubert to Expert Testimony in Class Certification*, 16 LEWIS & CLARK L. REV. 1349 (2012); Heather P. Scribner, *Rigorous Analysis of the Class Certification Expert: The Roles of Daubert and the Defendant’s Proof*, 28 REV. LITIG. 71 (2008).

¹³² *Wal-Mart*, 564 U.S. at 354 (emphasis added). This view is consistent with the Court’s instruction in *Daubert* that a federal district court, “[f]aced with a proffer of expert scientific testimony . . . must determine [the admissibility of the evidence] at the outset.” *Daubert*, 509 U.S. at 592–93 & n.10.

Daubert standards during class certification proceedings. As one author argued:

If defendant contends that the evidence relied upon by plaintiffs for class certification would be inadmissible at trial, the court must decide the admissibility question at the class certification stage. For example, if testimony relied on by plaintiffs in support of certification would be admissible under one view of the substantive law but inadmissible under another, the court must decide the legal issue to ensure that, in fact, plaintiffs' evidence supporting class certification would in fact be admissible at trial.¹³³

He cogently explained, "[i]f an expert will not be permitted to testify at all at trial, plaintiff should not be able to rely on that inadmissible testimony in support of class certification."¹³⁴

In sum, a federal district court must be convinced that, "if credited by the fact finder," plaintiff's evidence supporting class certification would tend to prove one or more liability or damages claims or defenses on a classwide basis.¹³⁵ Therefore, the evidence must be admissible and, for expert evidence, it must be relevant and reliable and otherwise satisfy *Daubert* standards. "Once the court determines that plaintiff[s] evidence is . . . admissible at trial, however, it should not weigh plaintiff[s] evidence against defendant's evidence" to attempt to resolve which is more persuasive.¹³⁶

The factors identified by the Supreme Court and the Advisory Committee guide the inquiry into the relevancy and reliability of expert testimony and create a framework for determining the admissibility of sampling evidence in class actions, including during class certification proceedings.¹³⁷

III. SAMPLING EVIDENCE AND CLASS ACTION PREREQUISITES: COMMONALITY, PREDOMINANCE, AND SUPERIORITY

The answer to the question whether sampling evidence is admissible to prove classwide liability or classwide damages is rooted in the prerequisites for class certification. "A party seeking class certification must affirmatively demonstrate . . . compliance with [Federal Rule of Civil Procedure 23]."¹³⁸ Rule 23(a)'s four threshold requirements—numerosity (or impracticability of joinder), commonality, typicality, and adequacy of representation—

¹³³ Klonoff, *supra* note 83, at 758.

¹³⁴ *Id.* at 760.

¹³⁵ *Id.*

¹³⁶ *Id.* at 760–61; *see infra* notes 163–167 and accompanying text.

¹³⁷ *See supra* notes 100–106 and accompanying text.

¹³⁸ Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338, 350 (2011).

“effectively ‘limit the class claims to those fairly encompassed by the named plaintiff’s claims.’”¹³⁹ Certification is proper only “if the trial court is satisfied, after a rigorous analysis, that the prerequisites of Rule 23(a) have been satisfied.”¹⁴⁰

A. Commonality

Commonality requires plaintiffs to show that “there are questions of law or fact common to the class.”¹⁴¹ In brief, “[c]ommonality requires the plaintiff to demonstrate that the class members ‘have suffered the same injury.’”¹⁴² Consequently, the class claims must depend on a “common contention,” one that “is capable of classwide resolution—which means that determination of its truth or falsity will resolve an issue that is central to the validity of each one of the claims in one stroke.”¹⁴³ “Identifying common questions typically

¹³⁹ Gen. Tel. Co. of Sw. v. Falcon, 457 U.S. 147, 156 (1982) (quoting Gen. Tel. Co. of Nw. v. EEOC, 446 U.S. 318, 330 (1980)). Disputes regarding the admissibility of statistical evidence in class actions most often arise in the contexts of the commonality prerequisite of Rule 23(a)(2) and the predominance and superiority prerequisites of Rule 23(b)(3). For present purposes, numerosity (or impracticability of joinder), typicality, and adequacy of representation are therefore of little moment. For a thorough discussion of those prerequisites, see 1 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS §§ 3:11–3:17, at 185–227 (5th ed. 2016) (numerosity (or impracticability of joinder)); *id.* §§ 3:28–3:49, at 262–320 (typicality); *id.* §§ 3:50–3:88, at 320–443 (adequacy of representation); as well as the implicit requirements of definiteness, *id.* §§ 3:2–3:7, at 155–76, and ascertainability of class membership, *id.* §§ 3:8–3:10, at 176–85.

¹⁴⁰ *Falcon*, 457 U.S. at 161; *see Wal-Mart*, 564 U.S. at 351 (“Frequently that ‘rigorous analysis’ will entail some overlap with the merits of the plaintiff’s underlying claim.”); *Coopers & Lybrand v. Livesay*, 437 U.S. 463, 469 (1978) (“[T]he class determination generally involves considerations that are ‘enmeshed in the factual and legal issues comprising the plaintiff’s cause of action.’” (quoting *Mercantile Nat. Bank v. Langdeau*, 371 U.S. 555, 558 (1963) (footnote omitted))).

Initially, in *Eisen v. Carlisle & Jacquelin*, 417 U.S. 156 (1974), the Supreme Court pronounced “[w]e find nothing in either the language or history of Rule 23 that gives a court any authority to conduct a preliminary inquiry into the merits of a suit in order to determine whether it may be maintained as a class action.” 417 U.S. at 177. However, four years later, in *Coopers & Lybrand*, the Court retreated from this bright-line proposition by quoting approvingly from a leading treatise: “Evaluation of many of the questions entering into determination of class action questions [e.g., typicality, adequacy of representation, commonality] is intimately involved with the merits of the claims The more complex determinations required in Rule 23(b)(3) class actions entail even greater entanglement with the merits” 437 U.S. at 469 n.12 (quoting 15 CHARLES ALAN WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 3911, at 485 n.45 (1976)). Then, in *Falcon*, the Court surmised that “sometimes it may be necessary for the court to probe behind the pleadings before coming to rest on the certification question.” 457 U.S. at 160. The Court subsequently made it clear that Rule 23 does not authorize “free-ranging merits inquiries at the certification stage. Merits questions may be considered to the extent—but only to the extent—that they are relevant to determining whether Rule 23 prerequisites for class certification are satisfied.” *Amgen, Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 466 (2013).

¹⁴¹ FED. R. CIV. P. 23(a)(2).

¹⁴² *Wal-Mart*, 564 U.S. at 349–50 (quoting *Falcon*, 457 U.S. at 157).

¹⁴³ *Id.* at 350.

requires examining the parties' claims and defenses, identifying the type of proof the parties expect to present, and deciding the extent to which there is a need for individual, as opposed to common, proof."¹⁴⁴ For purposes of Rule 23(a)(2), "[e]ven a single [common] question" will do.¹⁴⁵

Although it is generally conceded that *Wal-Mart* raised the bar to establish commonality, that decision has not changed the observation that the "commonality test is easily met in most cases."¹⁴⁶

B. Predominance

Under Rule 23(b)(3), plaintiffs must also show that "questions of law or fact common to class members predominate over any questions affecting only individual members."¹⁴⁷ Determining whether common questions predominate over individual questions requires a "qualitative rather than [] quantitative" "pragmatic assessment of the entire action and all the issues involved."¹⁴⁸ As defined by the *Tyson* Court, "[a]n individual question is one

¹⁴⁴ MANUAL FOR COMPLEX LITIGATION (FOURTH) § 21.141, at 257 (2004).

¹⁴⁵ *Wal-Mart*, 564 U.S. at 376 n.9 (Ginsburg, J., concurring in part and dissenting in part) (quoting Richard A. Nagareda, *The Preexistence Principle and the Structure of the Class Action*, 103 COLUM. L. REV. 149, 176 n.110 (2003)); see FED. RULE CIV. P. 23 advisory committee's notes to 1937 amendment (citing with approval cases in which "there was only a question of law or fact common to" class members).

¹⁴⁶ See 1 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 3:24, at 251 (5th ed. 2011); *supra* note 24; see e.g., *Chi. Teachers Union, Local No. 1 v. Bd. of Educ.*, 797 F.3d 426, 438 (7th Cir. 2015) (whether "subjective, discretionary decisions" were based on company policy or practice); *Brown v. Nucor Corp.*, 785 F.3d 895, 903–17 (4th Cir. 2015) (whether promotions were denied because of racial discrimination); *Edwards v. Ford Motor Co.*, 603 F. App'x 538, 540 (9th Cir. 2015) (whether defect existed and whether defendant had duty to disclose defect); *Jimenez v. Allstate Ins.*, 765 F.3d 1161, 1164–66 (9th Cir. 2014) (whether claims adjusters must be paid for off-the-clock overtime); *Suchanek v. Sturm Foods, Inc.*, 764 F.3d 750, 755–58 (7th Cir. 2014) (whether seller's packaging likely to mislead reasonable consumer); *In re Deepwater Horizon*, 739 F.3d 790, 809–12 (5th Cir. 2014) (whether class members suffered same injury, even though injurious effects (damages) were diverse); *Abdullah v. U.S. Sec. Assocs., Inc.*, 731 F.3d 952, 957–63 (9th Cir. 2013) (whether security guards were required to work through meal periods); *Butler v. Sears, Roebuck & Co.*, 727 F.3d 796, 801 (7th Cir. 2013) (whether washing machines were defective); *In re Whirlpool Corp. Front-Loading Washer Prods. Liab. Litig.*, 722 F.3d 838, 858–61 (6th Cir. 2013) (whether product design defects caused mold, mildew, and bacteria); *Young v. Nationwide Mut. Ins.*, 693 F.3d 532, 542–43 (6th Cir. 2012) (whether insurers charged incorrect local tax on policyholders' premiums); *Ross v. RBS Citizens, N.A.*, 667 F.3d 900, 908–10 (7th Cir. 2012) (whether bank broadly enforced policy denying earned-overtime compensation), *cert. granted, vacated*, 133 S. Ct. 1722 (2013).

¹⁴⁷ FED. R. CIV. P. 23(b)(3). Rule 23(b)(3) was the "most adventuresome" addition to Rule 23 by the 1966 amendment to the Federal Rules of Civil Procedure. Benjamin Kaplan, *A Preparatory Note*, 10 B.C. INDUS. & COM. L. REV. 497, 497 (1969). "Rule 23(b)(3) added to the complex-litigation arsenal class actions for damages designed to secure judgments binding all class members save those who affirmatively elected to be excluded." *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 614–15 (1997).

¹⁴⁸ *Parko v. Shell Oil Co.*, 739 F.3d 1083, 1085 (7th Cir. 2014); DANIEL R. COQUILLETTE ET AL., 5 MOORE'S FEDERAL PRACTICE § 23.45[1], at 23–215 (3d ed. 2016); see, e.g., *Parko*, 739 F.3d at 1085; *Stillmock v. Weis Markets, Inc.*, 385 F. App'x 267, 273 (4th Cir. 2010); *Vinole v.*

where ‘members of a proposed class will need to present evidence that varies from member to member’ while a common question is one where ‘the same evidence will suffice for each member to make a prima facie showing [or] the issue is susceptible to generalized, class-wide proof.’”¹⁴⁹

As instructed by the Supreme Court in *Erica P. John Fund*, “[c]onsidering whether ‘questions of law or fact common to class members predominate’ begins, of course, with the elements of the underlying cause of action.”¹⁵⁰ A federal district court must identify the relevant factual and legal issues and “compare the issues subject to common proof against the issues subject solely to individualized proof.”¹⁵¹ As previously noted, the predominance inquiry thus “generally involves considerations that are enmeshed in the factual and legal issues comprising the plaintiff’s causes of action.”¹⁵²

After “careful scrutiny” of “the relation between common and individual questions,” a federal district court must decide “whether proposed classes are sufficiently cohesive to warrant adjudication by representation.”¹⁵³ “Even if Rule 23(a)’s commonality requirement may be satisfied by [class members’] shared experience, the predominance criterion is far more demanding.”¹⁵⁴ “The predominance inquiry ‘asks whether the common, aggregation-enabling issues in the case are more prevalent or important than the non-common, aggregation-defeating, individual issues.’”¹⁵⁵

There is no litmus test. Some federal courts consider whether issues “that are subject to generalized proof, and thus applicable to the class as a whole, . . . predominate over those issues that are subject only to

Countrywide Home Loans, Inc., 571 F.3d 935, 946 (9th Cir. 2009); *Williams v. Mohawk Indus., Inc.*, 568 F.3d 1350, 1357–58 (11th Cir. 2009); *In re Nassau Cty. Strip Search Cases*, 461 F.3d 219, 227–29 (2d Cir. 2006).

¹⁴⁹ *Tyson*, 136 S. Ct. at 1045 (quoting 2 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS § 4:50, at 196–97 (5th ed. 2012)).

¹⁵⁰ *Erica P. John Fund, Inc. v. Halliburton Co.*, 563 U.S. 804, 809 (2011) (quoting FED. R. CIV. P. 23(b)(3)); see *supra* note 13.

¹⁵¹ 2 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS § 4:50, at 197 (5th ed. 2012).

¹⁵² *Comcast Corp. v. Behrend*, 569 U.S. 27, 34 (2013) (citation omitted); see *supra* note 13; *Taha v. County of Bucks*, 862 F.3d 292, 308–09 (3d Cir. 2017) (“Clearly, the trier of fact should be able to determine whether a violation was ‘willful’ [“only remaining question of fact”] by considering common evidence regarding defendants’ actions and intent without taking into account information regarding the individual class members.”); see also *supra* note 140 and accompanying text.

¹⁵³ *Tyson*, 136 S. Ct. at 1045; *Amchem*, 521 U.S. at 623.

¹⁵⁴ *Amchem*, 521 U.S. at 623–24. Note that “[p]redominance is a test readily met in certain cases alleging consumer or securities fraud or violations of the antitrust laws.” *Id.* at 625 (citing FED. R. CIV. P. 23 advisory committee’s notes to 1937 amendment).

¹⁵⁵ *Tyson*, 136 S. Ct. at 1045 (quoting 2 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS § 4:49, at 195–96 (5th ed. 2012)).

individualized proof.”¹⁵⁶ Others focus on whether the substantive elements of class members’ claims require the same proof for each class member.¹⁵⁷ And others regard whether one or more common issues constitute significant elements of each class member’s individual claims or are a significant aspect of a case.¹⁵⁸ Still others consider whether common issues are “more substantial than the issues subject only to individualized proof.”¹⁵⁹ Some courts also assess judicial resources and the parties’ litigation expenses to decide whether it would be more efficient to determine one or more issues on a classwide basis rather than to adjudicate all issues in separate proceedings.¹⁶⁰

Rule 23(b)(3) “does *not* require a plaintiff seeking class certification to prove that each element of her claim is susceptible to classwide proof,” only that common questions “predominate over any questions affecting only individual class members.”¹⁶¹ If questions affecting individual class members do not “overwhelm common ones,” the class is “sufficiently cohesive to warrant adjudication by representation.”¹⁶²

Importantly, “Rule 23(b)(3) requires a showing that *questions* common to the class predominate, not that those questions will be answered, on the merits, in favor of the class.”¹⁶³ As the Supreme Court observed in *Amgen*, a plaintiff seeking Rule 23(b)(3) class certification does not need to “first establish that it will win the fray” because that would be “put[ting] the cart before the horse.”¹⁶⁴ The Court explained, “the office of a Rule 23(b)(3) certification ruling is not to adjudicate the case; rather, it is to select the ‘metho[d]’ best suited to adjudication of the controversy ‘fairly and efficiently.’”¹⁶⁵

Moreover, “[h]ow many (if any) of the class members have a valid claim is the issue to be determined *after* the class is certified.”¹⁶⁶ Indeed, if plaintiffs satisfy all the prerequisites of Rule 23(a) and Rule 23(b)(3), a class may be certified even if none of the putative class members ultimately recovers,

¹⁵⁶ *Brown v. Kelly*, 609 F.3d 467, 483 (2d Cir. 2010) (internal quotation marks and citations omitted).

¹⁵⁷ *See, e.g., Hayes v. Wal-Mart Stores, Inc.*, 725 F.3d 349, 359 (3d Cir. 2013); *Rosario v. Livaditis*, 963 F.2d 1013, 1018 (7th Cir. 1992).

¹⁵⁸ *See, e.g., Messner v. Northshore Univ. Health Sys.*, 669 F.3d 802, 815 (7th Cir. 2012); *Jenkins v. Raymark Indus., Inc.*, 782 F.2d 468, 470–71 (5th Cir. 1986).

¹⁵⁹ *Mazzei v. Money Store*, 829 F.3d 260, 272 (2d Cir. 2016) (internal quotation marks and citations omitted).

¹⁶⁰ *See, e.g., Butler*, 727 F.3d at 798.

¹⁶¹ *Amgen*, 568 U.S. at 469 (internal quotation marks and alterations omitted).

¹⁶² *Halliburton Co. v. Erica P. John Fund*, 134 S. Ct. 2398, 2412 (2014); *Amchem*, 521 U.S. at 623.

¹⁶³ *Amgen*, 568 U.S. at 459.

¹⁶⁴ *Id.* at 460.

¹⁶⁵ *Id.* (paraphrasing FED. R. CIV. P. 23(b)(3)).

¹⁶⁶ *Parko*, 739 F.3d at 1085.

either because class defendant prevails on the merits of a classwide dispositive issue, or because class members' claims ultimately fail for different reasons.¹⁶⁷

Our proposal to determine class members' individual damages by following a statistically-based sampling method would resolve a "common contention" "in one stroke" because "the same evidence will suffice for each member to make a prima facie showing" of individual damages and the damages issues would be "susceptible to generalized, class-wide proof."¹⁶⁸ Unless classwide liability issues are subject to individualized proof, and those issues "are more prevalent or important" than damages issues, the determination of individual damages will predominate.¹⁶⁹ Historically, however, many putative class actions fail because they cannot satisfy the predominance prerequisite.¹⁷⁰

¹⁶⁷ *Amgen*, 568 U.S. at 470 & n.5; see, e.g., *In re Urethane Antitrust Litig.*, 768 F.3d at 1245, 1254 (10th Cir. 2014); *DG ex rel. Stricklin v. Devaughn*, 594 F.3d 1188, 1198 (10th Cir. 2010); *Kohen v. Pac. Inv. Mgmt. Co.*, 571 F.3d 672, 676–77 (7th Cir. 2009); *Krell v. Prudential Ins. Co. of Am.*, 148 F.3d 283, 307 (3d Cir. 1998).

If a class is certified and it is thereafter discovered that some class members do not have valid claims against the class defendant, a federal district court can alter or amend its certification order to narrow or refine the definition of the class. See FED. R. CIV. P. 23(c)(1)(C) ("An order that grants or denies class certification may be altered or amended before final judgment."); cf. *Kohen*, 571 F.3d at 677 (suggesting random selection of class members to prove whether large numbers did not suffer from alleged securities laws violations and that "a class will often include persons who have not been injured by the defendant's conduct . . . [s]uch a possibility or indeed inevitability does not preclude class certification").

A federal district court can also grant summary judgment in favor of the class defendant on uninjured class members' claims, see *Tyson*, 136 S. Ct. at 1047 (noting that where there is "[an alleged] failure of proof as to an element of the plaintiffs' cause of action—courts should engage that question as a matter of summary judgment, not class certification" (quoting *Nagareda*, *supra* note 83, at 107)); *Vaquero v. Ashley Furniture Indus., Inc.*, 824 F.3d 1150, 1156 (9th Cir. 2016) (noting that "alleged failure of proof" should be decided through dispositive motion, "not class certification"), or instruct the jury not to base any damages awards on such claims, see *Tyson*, 136 S. Ct. at 1044–45.

¹⁶⁸ See *Tyson*, 136 S. Ct. at 1045 (quoting 2 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 4:50, at 196–97 (5th ed. 2012) (internal quotation marks omitted)); *Wal-Mart*, 564 U.S. at 349–50 (quoting *Falcon*, 457 U.S. at 157).

¹⁶⁹ *Tyson*, 136 S. Ct. at 1045 (internal quotation marks and citations omitted).

¹⁷⁰ See, e.g., *Arnold v. Direct TV, LLC*, Case No. 4:10-CV-352-JAR, 2017 WL 1251033, at *7, 12 (E.D. Mo. Mar. 31, 2017) ("Plaintiffs' proposed damages analysis cannot be performed on a classwide basis. . . . Plaintiffs here have insufficient records to establish their individualized unreported work time."); *Local 589 v. Mass. Bay Transp. Auth.*, No. 13-cv-11455-ADB, 2016 WL 5109508, at *4 (D. Mass. Sept. 20, 2016) ("individual liability and damages, which will be a highly individualized undertaking"); *Perez v. Wells Fargo & Co.*, No. 14-cv-0989-PJH, 2016 WL 4180190, at *8–11 (N.D. Cal. Aug. 8, 2016) (liability cannot "be established in the absence of individualized inquiries"); *Atkins v. United States*, No. 4:15 CV 933 CDP, 2016 WL 3878466, at *5 (E.D. Mo. July 18, 2016) (questioning use of "representative samplings" and finding no predominance of common issues "given the proposed development of adequate, individualized evidence"); *Benedict v. Hewlett-Packard Co.*, 314 F.R.D. 457, 477 (N.D. Cal. 2016) (finding there was no "common proof" to determine whether employees qualified for FLSA exemption).

C. Superiority

To certify a damages class action pursuant to Rule 23(b)(3), a federal district court must also find that “a class action is superior to other available methods for fairly and efficiently adjudicating the controversy.”¹⁷¹

Rule 23(b)(3) identifies four non-exclusive factors that pertain to a superiority finding: “the class members’ interests in individually controlling the prosecution or defense of separate actions”; “the extent and nature of any litigation concerning the controversy already begun by or against class members”; “the desirability or undesirability of concentrating the litigation of the claims in the particular forum”; and “the likely difficulties in managing a class action.”¹⁷²

The manageability factor is the one most heavily litigated and the one most germane to the use of sampling evidence.¹⁷³ *Ceteris paribus*, as the size of the putative class increases, the potential efficiency, and hence superiority, of a class action increases, especially if there are truly common questions.

The goals of the predominance and superiority prerequisites are to “achieve economies of time, effort, and expense, and promote . . . uniformity of decision as to persons similarly situated, without sacrificing procedural fairness or bringing about other undesirable results.”¹⁷⁴ There should be no doubt that our proposal would achieve the economy goals and promote uniform individual damages awards for similarly situated class members.

¹⁷¹ FED. R. CIV. P. 23(b)(3). “[O]ther available methods” may include individual actions, joinder of claims, or “test” or “bellwether” trials, among others. *See generally* 2 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS § 4:85–4:88, at 357–74 (5th ed. 2012). Typically, federal courts consider the manageability of a class action relative to alternative methods of adjudication and are usually reluctant to deny class certification on this ground. *See, e.g.,* Leyva v. Medline Indus., Inc., 716 F.3d 510, 514–15 (9th Cir. 2013); Klay v. Humana, Inc., 382 F.3d 1241, 1272–73 (11th Cir. 2004), *abrogated in part on other grounds by* Bridge v. Phx. Bond & Indem., Co., 553 U.S. 639 (2008).

Rule 23(b)(3) damages class actions are often superior to other available methods of adjudication in “negative value cases,” cases involving numerous individuals with small damages claims that, absent a class action or other form of aggregate adjudication, would have no remedies. *See, e.g., Amchem*, 521 U.S. at 616–17; *Phillips Petroleum Co. v. Shutts*, 472 U.S. 797, 809 (1985); *Deposit Guar. Nat’l Bank v. Roper*, 445 U.S. 326, 339 (1980); *Carnegie v. Household Int’l Inc.*, 376 F.3d 656, 661 (7th Cir. 2004). Rule 23(b)(3) damages class actions are also usually a more efficient and economical method to adjudicate numerous claims based on the same or similar causes of action, for example, mass torts. *See, e.g., Ortiz v. Fibreboard Corp.*, 527 U.S. 815, 860–61 (1999); *Califano v. Yamasaki*, 442 U.S. 682, 701 (1979); *Johnson v. Ry. Express Agency, Inc.*, 421 U.S. 454, 466–67 n.12 (1975); *Am. Pipe & Const. Co. v. Utah*, 414 U.S. 538, 550 (1974).

¹⁷² FED. R. CIV. P. 23(b)(3)(A)–(D); *see* *Sykes v. Mel S. Harris & Assocs. LLC*, 780 F.3d 70, 82 (2d Cir. 2015) (stating that, although the structure of Rule 23(b)(3) suggests that the factors apply to both the predominance and superiority prerequisites, the factors “more clearly implicate the superiority inquiry”).

¹⁷³ For discussions of all four factors, *see* COQUILLETTE ET AL., *supra* note 148, §§ 23.46[2][a]–3[c] at 23-272–23-294.

¹⁷⁴ *Amchem*, 521 U.S. at 615 (quoting FED. R. CIV. P. 23 advisory committee’s notes to 1966 amendment) (alterations in the original).

And, because we advocate following our method only when judgment variability, the randomness that each class member's damages claim is already exposed to, exceeds claim variability, the substantive differences among the class members' damages claims, no procedural fairness would be sacrificed.¹⁷⁵

One "matter[] pertinent to these findings [predominance and superiority] include[s] . . . the likely difficulties in managing a class action."¹⁷⁶ In opposing class certification, defendant will often argue that, even if liability to the class is a common question, the determinations of what, if any, damages are owed to the class as a whole or to its individual members are complex and fact-specific, rendering class proceedings unmanageable.¹⁷⁷

D. Aggregate Damages and Individual Damages

When considering Rule 23(b)(3) classwide damages issues, aggregate damages and individual damages should be distinguished. Aggregate damages are the total amount of damages that the class defendant, if found liable, pays the entire class. Individual damages are the separate portions of that total amount that each class member with a valid claim is awarded in damages.¹⁷⁸

Arguments about damages often arise at the class certification stage because plaintiffs must demonstrate that one or more common issues predominate over individual ones. To attempt to defeat certification of a Rule 23(b)(3) damages class, a defendant will often argue that individual damages calculations predominate over common liability issues. By citing and quoting approvingly a leading treatise, the *Tyson* Court put to rest an issue that had perplexed federal courts: whether the need to prove individual damages precludes Rule 23(b)(3) class certification. It ruled:

When "one or more of the central issues in the action are common to the class and can be said to predominate, the action may be

¹⁷⁵ See *infra* notes 249–262 and accompanying text.

¹⁷⁶ FED. R. CIV. P. 23(b)(3).

¹⁷⁷ See, e.g., *Murray v. GMAC Mortg. Corp.*, 434 F.3d 948, 953 (7th Cir. 2006) ("[A]n effort to determine a million consumers' individual losses would make the suit unmanageable."); *Newton v. Merrill Lynch, Pierce, Fenner & Smith, Inc.*, 259 F.3d 154, 192–93 (3d Cir. 2001) ("Because injury determinations must be made on an individual basis in this case, adjudicating the claims as a class will not reduce litigation or save scarce judicial resources."); *Rutstein v. Avis Rent-A-Car Sys., Inc.*, 211 F.3d 1228, 1239–40 (11th Cir. 2000) (damages claims "focus almost entirely on facts and issues specific to individuals rather than the class as a whole" (citation and quotation marks omitted)); *Allison v. Citgo Petroleum Corp.*, 151 F.3d 402, 419 (5th Cir. 1998) ("[I]ndividualized monetary damages determinations for more than a thousand potential plaintiffs would require multiple juries . . . [and] implicated significant . . . efficiency, and manageability problems.").

¹⁷⁸ See generally 4 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 12:1–12:5, at 89–120 (5th ed. 2014).

considered proper under Rule 23(b)(3) even though other important matters will have to be tried separately, such as damages . . . peculiar to some individual class members.”¹⁷⁹

Indeed, “[r]ecognition that individual damages calculations do not preclude class certification under Rule 23(b)(3) is well nigh universal.”¹⁸⁰ Thus, it is now clear that classwide liability may be adjudicated and individual damages determined, if necessary, in later proceedings.¹⁸¹ That said, before

¹⁷⁹ *Tyson*, 136 S. Ct. at 1045. (quoting CHARLES A. WRIGHT ET AL., 7AA FEDERAL PRACTICE AND PROCEDURE § 1778, at 123–24 (3d ed. 2005)); see, e.g., FED. R. CIV. P. 23 advisory committee’s notes to 1966 amendment (“[A] fraud perpetrated on numerous persons by the use of similar misrepresentations may be an appealing situation for a class action, and it may remain so despite the need, if liability is found, for separate determination of the damages suffered by individuals within the class.”); *Day v. Celadon Trucking Servs., Inc.*, 827 F.3d 817, 833 (8th Cir. 2016) (quoting *Tyson*).

¹⁸⁰ *Comcast*, 569 U.S. at 42 (Ginsburg & Breyer, JJ., dissenting) (citing 2 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS § 4:54, at 205 (5th ed. 2012)); see, e.g., *Roach* 778 F.3d at 405; *Levy*, 716 F.3d at 513–14; *In re Nassau Cty. Strip Search Cases*, 461 F.3d at 227; *Klay*, 382 F.3d at 1259–60; *Gunnells v. Healthplan Servs., Inc.*, 348 F.3d 417, 428–30 (4th Cir. 2003); *Smilow v. Sw. Bell Mobile Sys., Inc.*, 323 F.3d 32, 40 (1st Cir. 2003); *Sterling v. Velsicol Chem. Corp.*, 855 F.2d 1188, 1197 (6th Cir. 1988).

¹⁸¹ See, e.g., *McMahon v. LVNV Funding, LLC*, 807 F.3d 872, 876 (7th Cir. 2015) (“It is well established that, if a case requires determinations of individual issues of causation and damages, a court may bifurcate the case into a liability phase and a damages phase.” (citations omitted)); *Neale v. Volvo Cars of N. Am., LLC*, 794 F.3d 353, 374–75 (3d Cir. 2015) (“[I]t is ‘a misreading of *Comcast*’ to interpret it as ‘preclud[ing] certification under Rule 23(b)(3) in any case where the class members’ damages are not susceptible to a formula for classwide measurement.’” (quoting *In re Deepwater Horizon*, 739 F.3d at 815 & n.104)); *Roach*, 778 F.3d at 402 (“We hold that *Comcast* does not mandate that certification pursuant to Rule 23(b)(3) requires a finding that damages are capable of measurement on a classwide basis.”); *In re Urethane Antitrust Litig.*, 768 F.3d at 1257–58 (“[W]e know from the actual trial that individualized issues did not predominate.”); *In re Deepwater Horizon*, 739 F.3d 790, 817 (5th Cir. 2014) (“Even after *Comcast*, the predominance inquiry can still be satisfied under Rule 23(b)(3) if the proceedings are structured to establish ‘liability on a class-wide basis, with separate hearings to determine—if liability is established—the damages of individual class members.’” (citations omitted)); *In re Whirlpool Corp. Front-Loading Washer Prods. Liab. Litig.*, 722 F.3d 838, 860–61 (6th Cir. 2013) (“[D]istrict court certified only a liability class and reserved all issues concerning damages for individual determination.”); *In re New Motor Vehicles Canadian Export Antitrust Litig.*, 522 F.3d 6, 28–30 (1st Cir. 2008) (noting that “[p]redominance is not defeated by individual damages questions as long as liability is still subject to common proof,” and vacating and remanding class certification order, in part, to reconsider liability issues); *Chiang v. Veneman*, 385 F.3d 256, 273 & n.11 (3d Cir. 2004) (affirming Rule 23(c)(4)(A) issues class despite “think[ing] it unlikely that the calculation of damages will be suitable for class determination”); *Olden v. LaFarge Corp.*, 383 F.3d 495, 509 (6th Cir. 2004) (affirming Rule 23(b)(3) class certification for common liability issues and noting that federal district court can bifurcate liability issues from damages issues); *In re Visa Check/MasterMoney Antitrust Litig.*, 280 F.3d 124, 141 (2d Cir. 2001) (discussing “management tools available to a district court to address any individualized damages issues”); *In re Simply Orange Juice Mktg. & Sales Prac.*, Master Case No. 4:12-md-02361-FJG, 2017 WL 3142095, at *7-9 (W.D. Mo. July 24, 2017) (finding damages cannot be established through common evidence, but certifying liability issues class under Rule 23(c)(4)).

certifying a Rule 23(b)(3) class, a federal district court must still weigh whether individual damages issues will overwhelm common liability issues.¹⁸²

Plaintiff may argue that, even without a proposal regarding how to determine individual damages, the Rule 23(b)(3) predominance prerequisite is met by a trial plan to prove aggregate damages.¹⁸³ Defendant, while still

¹⁸² See, e.g., *McLaughlin v. Am. Tobacco Co.*, 522 F.3d 215, 231 (2d Cir. 2008) (citing *Eisen v. Carlisle & Jacquelin*, 479 F.2d 1005, 1008 (2d Cir. 1973)) (rejecting “fluid recovery” whereby “aggregate liability” would initially be determined and then distributed “usually through a simplified proof of claim procedure”); *Klay*, 382 F.3d at 1260 (affirming Rule 23(b)(3) class certification, but noting “[i]t is primarily when there are significant individualized questions going to liability that the need for individualized assessments of damages is enough to preclude [Rule] 23(b)(3) certification”).

¹⁸³ See, e.g., *Kleen Prods. LLC v. Int’l Paper Co.*, 831 F.3d 919, 929 (7th Cir. 2016) (rejecting as a matter of law that plaintiffs are obligated “to drill down and estimate each individual class member’s damages” at class certification stage); *In re Pharm. Indus. Average Wholesale Price Litig.*, 582 F.3d 156, 197–98 (1st Cir. 2009) (“The use of aggregate damages calculations is well established in federal court and implied by the very existence of the class action mechanism itself.”); see also *Hickory Sec. Ltd. v. Republic of Argentina*, 493 F. App’x 156, 160 (2d Cir. 2012) (remanding to federal district court to calculate “aggregate damage award [that] would ‘roughly reflect’ the loss to each [sub]class”); *Suchanek v. Sturm Foods, Inc.*, Case No. 11-CV-565-NJR-RJD, 2017 WL 3704206, at *14 (S.D. Ill. Aug. 28, 2017) (“[D]amages are capable of being measured on a class-wide basis using” models.); *In re Domestic Drywall Antitrust Litig.*, MDL No. 2437, 13-md-2437, 2017 WL 3623466, at *46 (E.D. Pa. Aug. 23, 2017) (noting that there is “no need to show individual damages for individual Plaintiffs” at class certification stage and finding proposed model was sufficient “for classwide proof of measurable damages”); *La. Mun. Police Emps. Ret. Sys. v. Green Mountain Coffee Roasters, Inc.*, Case No. 2:11-cv-289, 2017 WL 3149424, at *6–7 (D. Vt. July 21, 2017) (stock price “event study” “methodology . . . can be applied on a class-wide basis”); *Gordon v. Maxim Healthcare Servs., Inc.*, Civ. No. 13-7175, 2017 WL 3116153, at *8 (E.D. Pa. July 7, 2017) (damages measurable on classwide basis); *In re Chinese-Manufactured Drywall Prods. Liab. Litig.*, MDL No. 2047, 2017 WL 1421627, at *23 (E.D. La. Apr. 21, 2017) (“Plaintiffs have devised a reasonable and reliable solution to calculate remediation damages on a class-wide basis and accommodate individual class damage issues by shifting the individual damage components to subsequent adjudicative phases.”); *In re Dial Complete Mktg. & Sales Practices Litig.*, 320 F.R.D. 326, 337 (D.N.H. 2017) (proffered calculation capable of establishing “full extent of damages on a class-wide basis”); *Martin v. Monsanto Co.*, No. ED CV 16-2168-JFW (SPx), 2017 WL 1115167, at *8–9 (C.D. Cal. Mar. 24, 2017) (benefit-of-the-bargain model to determine total damages); *Naylor Farms, Inc. v. Chaparral Energy, LLC*, No. CIV-11-0634-HE, 2017 WL 187542, at *8 (W.D. Okla. Jan. 17, 2017) (expert’s model “can determine damages on a class wide basis”); *Fond Du Lac Bumper Exch., Inc. v. Jui Li Enter. Co.*, No. 09-CV-0852, 2016 WL 3579953, at *11 (E.D. Wis. June 24, 2016) (“[R]egression analysis serves as a sufficiently reliable method for establishing that damages can be proven on a class-wide basis.”).

Decisions holding that Rule 23(b)(3) prerequisites are satisfied by a trial plan to prove aggregate damages find support from dicta in *Comcast*. The damages model that plaintiffs proposed in *Comcast* did not purport to calculate damages suffered by each class member, and the model was not challenged or rejected on that basis. The Supreme Court rejected the model finding it did not satisfy the predominance standard of Rule 23(b)(3) because it “falls far short of establishing that damages are capable of measurement on a classwide basis.” *Comcast*, 569 U.S. at 34. That is, the damages model failed to measure aggregate damages. *Id.* The Court stated that “[w]ithout presenting another methodology, . . . [q]uestions of individual damage calculations will inevitably overwhelm questions common to the class.” *Id.* The negative implication of the Court’s statement is that a damages model that could determine aggregate damages would have satisfied the predominance prerequisite.

asserting that common questions do not predominate over individual ones, may counter that proving only aggregate damages is not sufficient to satisfy Rule 23(b)(3) and may raise the possibility of numerous mini-trials to calculate individual damages. Many federal courts have held that proof of aggregate damages is sufficient for Rule 23(b)(3) class certification purposes if plaintiff also proposes a common methodology to calculate individual damages.¹⁸⁴

Alternatively, if aggregate damages cannot be proven, plaintiff may argue that Rule 23(b)(3) prerequisites are still satisfied by proposing a common methodology to determine individual damages. A leading treatise states:

[T]he predominance inquiry is focused on . . . [whether] there must be a single or common *method* that can be used to measure and quantify the damages of each class member, lest individual damages calculations predominate over common questions of liability. The class proponents' task, therefore, is to demonstrate a *method* for quantifying *individual* damages that applies across the board and hence is *common* to the class: a common classwide method for calculating individual damages.¹⁸⁵

In many class actions, a common methodology to determine individual damages exists.¹⁸⁶ On the other hand, several federal courts have denied Rule

¹⁸⁴ See, e.g., *Waggoner v. Barclays PLC*, 875 F.3d 79, 106 (2d Cir. 2017) (damages for individual class members could be calculated by applying a method across the entire class); *In re Polyurethane Foam Antitrust Litig.*, No. 1:10 M D 2196-JZ, 2015 WL 4459636, at *8 (N.D. Ohio Jul. 21, 2015) ("A well-designed claims process can ensure each class member receives damage payments only if and as appropriate."); *In re Nexium (Esomeprazole) Antitrust Litig.*, 297 F.R.D. 168, 183 (D. Mass. 2013) ("[T]wo sets of common damages are calculated and presented under a reasonable and judicially acceptable methodology."); *In re Neurontin Antitrust Litig.*, Nos. 02-1830-FSH, 02-2731-FSH, 2011 WL 286118, at *10 (D.N.J. Jan. 25, 2011) ("[A]t class certification, Plaintiffs need only demonstrate that they have a 'viable method' for calculating damages that is common to the class."); *In re Cardizem CD Antitrust Litig.*, 200 F.R.D. 297, 323–24 (E.D. Mich. 2001) ("yardsticks" available to measure drug overprices).

¹⁸⁵ 4 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 12:4, at 110–11 (5th ed. 2014) (footnotes omitted); see *Lambert v. Nutraceutical Corp.*, 870 F.3d 1170, 1182 (9th Cir. 2017) ("Uncertainty regarding class members' damages does not prevent certification of a class as long as a valid method has been proposed for calculating those damages.").

¹⁸⁶ See, e.g., *Butler*, 727 F.3d at 801 (post-*Comcast* remand affirming certification of liability only class and noting "damages of individual class members can be readily determined in individual hearings, in settlement negotiations, or by creation of subclasses"); *Leyva*, 716 F.3d at 514 (reversing denial of Rule 23(b)(3) class certification because defendant's "computerized payroll and time-keeping database would enable the court to accurately calculate damages and related penalties for each claim"); *Johnson v. Meriter Health Servs. Emp. Ret. Plan*, 702 F.3d 364, 372 (7th Cir. 2012) (citing cases) ("Should it appear that the calculation of monetary relief will be mechanical, formulaic, a task not for a trier of fact but for a computer program . . . the district court can award that relief without terminating the class action and leaving the class members to their own devices . . ."); *Ward v. Dixie Nat'l Life Ins. Co.*, 595 F.3d 164, 180 (4th Cir. 2010)

(identical formula could be used for all class members); *Similow v. Sw. Bell Mobile Sys., Inc.*, 323 F.3d 32, 40–41 (1st Cir. 2003) (computer program could calculate damages for cellphone users allegedly overcharged for incoming calls); *Leon v. Diversified Concrete, LLC*, No. 15-6301-CJB, 2016 WL 6247674, at *7 (E.D. La. Oct. 26, 2016) (statutory damages awards); *Hurt v. Commerce Energy, Inc.*, No. 1:12-cv-00758-JSG, 2015 WL 1298674, at *2 & n.20 (N.D. Ohio 2015) (“[A]lthough the damages would be unique for each class member, they would all be determined by applying a consistent methodology—a mathematical formula based on the number of hours worked and the amount of wages already earned.”); *Barbosa v. Cargill Meat Sols. Corp.*, 297 F.R.D. 431, 456 n.2 (E.D. Cal. 2013) (Rule 23(b)(3) settlement with personnel and payroll records to determine each class member’s damages); *Parra v. Bashas’, Inc.*, 291 F.R.D. 360, 393 (D. Ariz. 2013), *amended in part sub nom. Estrada v. Bashas’, Inc.*, 2014 WL 1319189 (D. Ariz. 2014) (demonstrating model for calculating backpay “through a computer program, and relying upon ‘objective factors’ such as ‘the individual employee payroll record (dates of employment[,] job position, hours worked) and the wage scale,’ . . . plaintiffs will be able to calculate back pay losses for ‘each eligible class member’”); *In re Diamond Foods, Inc. Sec. Litig.*, 295 F.R.D. 240, 252 (N.D. Cal. 2013) (“[C]omputing individual damages will be virtually a mechanical task.” (quoting *Blackie v. Barrack*, 524 F.2d 891, 905 (9th Cir. 1975)); *In re Checking Account Overdraft Litig.*, 286 F.R.D. 645, 658 (S.D. Fla. 2012) (damages “ministerial[ly]” calculated using class defendant’s records); *Meijer, Inc. v. Warner Chilcott Holdings Co. III, Ltd.*, 246 F.R.D. 293, 311–13 (D.D.C. 2007) (overcharges “will be readily susceptible to formulaic analysis that does not require individualized inquiry as to each Class member”).

For post-*Tyson* decisions where federal district courts endorsed common methodologies to assess individual damages, see, for example, *Snipes v. Dollar Tree Distribution, Inc.*, No. 2:15-cv-00878-MCE-DB, 2017 WL 5754894, at *5 (E.D. Cal. Nov. 28, 2017) (valid statistical method for calculating classwide damages); *Luna v. Marvell Tech. Grp., Ltd.*, No. C 15-05447 WHA, 2017 WL 4865559, at *6 (N.D. Cal. Oct. 27, 2017) (event study on market efficiency to calculate classwide damages); *Johnson v. Hartford Cas. Ins. Co.*, Case No. 15-cv-04138-WHO, 2017 WL 2224828, at *16 (N.D. Cal. May 22, 2017) (damages measured based on defendant’s records); *Cazares v. AVA Rest. Corp.*, 15-CV-0477(KAM)(RML), 2017 WL 1229727, at *8 (E.D.N.Y. Mar. 31, 2017) (“[D]etermining damages . . . would be a matter of performing simple arithmetic using the data defendants have produced.”); *Kramer v. Am. Bank & Trust Co.*, 11 C 8758, 2017 WL 1196965, at *6 (N.D. Ill. Mar. 31, 2017) (“Plaintiffs propose a class-wide method of calculating damages, based on the hours that an individual class member worked”); *Herman v. Seaworld Parks & Entmt., Inc.*, Case No. 8:14-cv-3028-MSS-JSS, 2017 WL 1304302, at *18 (M.D. Fla. Mar. 10, 2017) (statutory damages “can be easily calculated utilizing class-wide data available from” defendant); *Balderrama-Baca v. Clarence Davids & Co.*, 318 F.R.D. 603, 613–14 (N.D. Ill. 2017) (“calculating individual damages will be a simple ministerial matter of gathering employee pay records” and applying “statutory damages formula”); *Durant v. State Farm Mut. Auto. Ins. Co.* No. 2-15-CV-01710-RAJ, 2017 WL 950588, at *6 (W.D. Wash. Mar. 9, 2017) (“able to calculate the [individualized] damages based on the submitted records”); *McCurdy v. Prof'l Credit Serv.*, No. 6:15-cv-01498-AA, 2016 WL 5853721, at *6 (D. Or. Oct. 3, 2016) (statutory damages); *Kempen v. Matheson Tri-Gas, Inc.*, No. 15-cv-00660-HSG, 2016 WL 4073336, at *2 (N.D. Cal. Aug. 1, 2016) (class members’ payments calculated formulaically); *Sibley v. Sprint Nextel Corp.*, 315 F.R.D. 642, 662–63 (D. Kan. 2016) (plaintiffs’ experts supplied “a ‘fancy’ computerized commission-calculation model to quantify individual damages for each class-member plaintiff”); *In re Volkswagen “Clean Diesel” Mktg., Sales Practices, & Prods. Liab. Litig.*, No. 2672 CRB (JSC), 2016 WL 4010049, at *4 (N.D. Cal. July 26, 2016) (remedies consisted of only two options); *Manouchehri v. Styles for Less, Inc.*, No. 14 cv 2521 NLS, 2016 WL 3387473, at *4 (S.D. Cal. June 20, 2016) (damages “can be calculated using a common methodology of calculation”); *Jimenez v. Menzies Aviation, Inc.*, No. 15-cv-02392-WHO, 2016 WL 3231106, at *7–8 (N.D. Cal. June 13, 2016) (“[U]niform pay stubs can serve as a common form of proof”); *Nitsch v. Dreamworks Animation SKG, Inc.*, 315 F.R.D. 270, 304 (N.D. Cal. 2016) (“[C]ommon evidence and a regression analysis could be used to create a model for quantifying the estimated cost to class members resulting from Defendants’ challenged conduct.”);

23(b)(3) class certification where individual damages could not be computed formulaically.¹⁸⁷ Others, however, have held that the failure to propose a model to determine individual damages does not defeat predominance.¹⁸⁸

The message to proponents of statistical evidence is that the evidence must demonstrate how and why the evidence is both relevant and reliable to prove either classwide liability, or aggregate damages or individual damages.

Harper v. Law Office of Harris & Zyde LLP, No. 15-cv-0114-HS6, 2016 WL 2344194, at *5 (N.D. Cal. May 4, 2016) (putative class seeking only statutory damages, therefore, there were no “individual questions in this action”); Chastain v. Cam, No. 3:13-cv-01802-SI, 2016 WL 1572542, at *9 (D. Or. Apr. 19, 2016) (“[O]bjective factors such as payroll records will allow for damage calculations on a class-wide basis.”); Dekeyser v. Thyssenkrupp Waupaca, Inc., 314 F.R.D. 449, 455 (E.D. Wisc. 2016) (damages would involve “mere mathematical computations”); cf. *In re Dial Complete Mktg. & Sales Practices Litig.*, 312 F.R.D. 36, 77 (D.N.H. 2015) (“[P]laintiffs have provided insufficient detail regarding their proposed methodologies for calculating classwide damages.”); Weidenhamer v. Expedia, Inc., No. C14-1239-RAJ, 2015 WL 7157282, at *15 (W.D. Wash. Nov. 13, 2015) (“[F]ailure to present a model or other method of evaluating class-wide injury weighs against a finding that common issues predominate.”).

¹⁸⁷ See, e.g., *Steering Comm. v. Exxon Mobil Corp.*, 461 F.3d 598, 602 (5th Cir. 2006) (noting that it was “clear from the record that the damages claims . . . are not subject to any sort of formulaic calculation”); *Blades v. Monsanto Co.*, 400 F.3d 562, 570 (8th Cir. 2005) (no “reliable methodology to determine the premiums paid”); *Bell Atlantic Corp. v. AT&T Corp.*, 339 F.3d 294, 307 (5th Cir. 2003) (“Class treatment, . . . may not be suitable where the calculation of damages is not susceptible to a mathematical or formulaic calculation”); *Sikes v. Teleline, Inc.*, 281 F.3d 1350, 1366 (11th Cir. 2002) (“extensive individualized inquiries on the issues of injury and damages”), *abrogated on other grounds by Bridge v. Phx. Bond & Indem. Co.*, 553 U.S. 639 (2008); *Windham v. Am. Brands, Inc.*, 565 F.2d 59, 66–67 (4th Cir. 1977) (en banc) (“The claims could not be proved by any set method of mathematical or formula calculation but would require individual proof and trial”); *Little v. Wash. Metro. Area Trans. Auth.*, Civil Action No. 14-1289 (RMC), 2017 WL 1403122, at *22 (D.D.C. Apr. 18, 2017) (no “formulaic approach for determining class-wide damages”).

In *Amchem*, the Supreme Court held that a federal district court improperly certified asbestos-related personal injury claims because the Rule 23(b)(3) predominance prerequisite was not satisfied. *Amchem*, 521 U.S. at 625. In *Ortiz v. Fibreboard Corp.*, 527 U.S. 815 (1999), the Court disallowed a proposed class action settlement in another asbestos case, purportedly brought under the limited fund provision of Rule 23(b)(1)(B). *Amchem* and *Ortiz* may indicate that mass tort cases involving damages claims for personal injuries are unlikely to be certified as class actions. Conversely, when a single accident gives rise to common liability and causation issues, those issues may predominate over individual damages issues. See, e.g., *Mullen v. Treasure Chest Casino, LLC*, 186 F.3d 620, 627 (5th Cir. 1999) (“[P]utative class members are all symptomatic by definition and claim injury from the same defective ventilation system over the same general period of time.”); cf. *In re Deepwater Horizon*, 739 F.3d at 815–20 (affirming settlement class certification for “economic loss and property damage”). But see FED. R. CIV. P. 23 advisory committee’s notes to 1966 amendment (“A ‘mass accident’ resulting in injuries to numerous persons is ordinarily not appropriate for a class action because of the likelihood that significant questions, not only of damages but of liability and defenses of liability, would be present, affecting the individuals in different ways. In these circumstances an action conducted nominally as a class action would degenerate in practice into multiple lawsuits separately tried.” (citations omitted)).

¹⁸⁸ See, e.g., *Roach*, 778 F.3d at 408–09.

Moreover, the proffered statistical evidence must tend to prove “common, aggregation-enabling, issues.”¹⁸⁹

As we will soon demonstrate, the sampling evidence we propose is both relevant and reliable and proves each class member’s individual damages.¹⁹⁰ As such, it is probative regarding “common, aggregation-enabling, issues.”

E. Bifurcation

Although there must be at least one common issue for a case to be certified as a class action, typically there are individual issues as well.¹⁹¹ In those rarer cases that only raise issues that are shared by all class members, there is no need for individual proofs, and a unitary trial is held.¹⁹² In many situations, however, a trial is bifurcated and common issues (often liability issues) are resolved in the trial’s first phase and individual issues (frequently damages issues) are adjudicated in the trial’s second phase.¹⁹³ As a popular treatise explains:

¹⁸⁹ *Tyson*, 136 S. Ct. at 1045 (quoting 2 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 4:50, at 196–97 (5th ed. 2012)).

¹⁹⁰ One federal district court has concluded that, “at a minimum, reliability under Rule 23 is a higher standard than reliability under *Daubert*.” *In re Rail Freight Fuel Surcharge Antitrust Litig.*, MDL Dock. No. 1869, Misc. No. 07-0489 (PLF), 2017 WL 5311533, at *50 (D.D.C. Nov. 13, 2017) (citations omitted). The court opined that, because it “must resolve expert disputes that bear on class certification. . . . [t]his may require evaluating the conclusions and results of competing experts, which goes beyond the scope of the *Daubert* inquiry.” *Id.* The court continued that a “‘rigorous analysis’ of whether plaintiffs have established predominance is certainly a more in-depth inquiry than required under *Daubert*.” *Id.*

¹⁹¹ As discussed previously in Part III, for a case to proceed as a money damages class action pursuant to Rule 23(b)(3), common issues must predominate over individual ones and the class action device must be superior to other available methods of adjudication. *See supra* notes 147–177 and accompanying text.

¹⁹² The prototypical case is one where liability is based on defendant’s conduct and the class is seeking injunctive or declaratory relief pursuant to Rule 23(b)(2) (“[T]he party opposing the class has acted or refused to act on grounds that apply generally to the class, so that final injunctive relief or corresponding declaratory relief is appropriate respecting the class as a whole.”). *See, e.g.*, FED. R. CIV. P. 23 advisory committee’s notes to 1966 amendment (“[i]llustrative are various actions in the civil-rights field where a party is charged with discriminating unlawfully against a class,” retailers against a seller based on illegal price discrimination, and purchasers or licensees against a patentee based on illegal “tying” conditions).

¹⁹³ *See, e.g., Vaquero*, 824 F.3d at 1153–55 (affirming Rule 23(b)(3) class certification on liability issues where plaintiff “proposed to resolve the damages phase . . . through use of a survey, sampling evidence, or a special master”); *In re Whirlpool Corp. Front-Loading Washer Prods. Liab. Litig.*, 722 F.3d at 860–61 (post-*Comcast* remand; affirming certification of liability only class and “reserv[ing] all issues concerning damages for individual determination”); *Sellers v. CRST Expedited, Inc.*, No. C15-117-LTS, 2017 WL 1193730, at *23–24 (N.D. Iowa Mar. 30, 2017) (bifurcation; “common issues as to liability will predominate over individual issues,” but “individualized questions will predominate” regarding “hostile work environment claim and damages as to both claims”); *St. Bernard Parish Gov’t v. United States*, No. 05-1119 L, 2016 WL 2641792, at *25–26 (Ct. Cl. May 4, 2016) (accepting “representative evidence” on liability issues); *Melgar v. Zicam LLC*, No. 2:14-CV-00160-MCE-AC, 2016 WL 1267870, at *5 (E.D. Cal. Mar. 31, 2016) (common issues “more prevalent or important” than individual issues

In class action cases, this familiar split serves a particularly important function: because the defendant's liability is an issue that likely applies to all class members' claims, while damages may be individualized, bifurcation often enables a case to proceed as a class action by ensuring that the common liability issues predominate while the individualized damages issues are addressed in some other manageable fashion.¹⁹⁴

Furthermore, bifurcation of common issues from individual issues "insulates a party from the possible prejudice of jointly trying certain issues."¹⁹⁵ For this and other reasons, bifurcation is ordered in a wide variety of contexts.¹⁹⁶

regarding damages or alleged misrepresentations); *Petersen v. Costco Wholesale Co.*, 312 F.R.D. 565, 579 (C.D. Cal. 2016) ("liability-only class, reserving for a second phase of trial the issue of damages"); *Rollins v. Taylor Bros., Inc.*, No. C14-1414-JCC, 2016 WL 258523, at *7-12 (W.D. Wash. Jan. 21, 2016) (bifurcated trial; first phase for disparate treatment and disparate impact claims and punitive damages and second phase for "backpay or particularized injunctive relief and compensatory damages, as well as the individual's share of any punitive damages"); *Lilly v. Jamba Juice Co.*, No. 13-cv-02998-JST, 2015 WL 1248027, at *6-8 (N.D. Cal. Mar. 18, 2015) (preliminary approval of class settlement of liability issues only).

Class actions may also be split into separate issues, FED. R. CIV. P. 23(c)(4) ("When appropriate, an action may be brought or maintained as a class action with respect to particular issues."); see 2 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* §§ 4:89-4:92, at 375-96 (5th ed. 2012), or subclasses, FED. R. CIV. P. 23(c)(5) ("When appropriate, a class may be divided into subclasses that are each treated as a class under this rule."); see 3 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS*, §§ 7.29-7:32, at 148-66 (5th ed. 2013).

¹⁹⁴ 4 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 11:3, at 9 (5th ed. 2014) (citations omitted); see *MANUAL FOR COMPLEX LITIGATION* (FOURTH) § 11.631, at 121-22 (2004) ("[T]he judge may consider severing for a joint trial those issues on which common evidence predominates, reserving noncommon issues for subsequent individual trials."). But see *Ebert v. Gen. Mills, Inc.*, 823 F.3d 472, 479 (8th Cir. 2016) (reversing Rule 23(b)(3) class certification because federal district court eliminated "any efficiencies gained by the class proceeding because many individual issues [e.g., "property-by-property assessment"] will require trial").

¹⁹⁵ 4 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 11:4, at 13 (5th ed. 2014). A federal district court can also address individual issues by "appointing a magistrate judge or special master to preside over individual damages proceedings," "decertifying the class after the liability trial and providing notice to class members concerning how they may proceed to prove damages," "creating subclasses," or "altering or amending the class." *In re Visa Check/MasterMoney Antitrust Litig.*, 280 F.3d 124, 141 (2d Cir. 2001) (Sotomayor, J.), *overruled on other grounds by In re Initial Pub. Offerings Sec. Litig.*, 471 F.3d 24 (2d Cir. 2006); see *Epenscheid v. DirectSat USA, LLC*, 705 F.3d 770, 775 (7th Cir. 2013).

¹⁹⁶ 4 WILLIAM B. RUBENSTEIN ET AL., *NEWBERG ON CLASS ACTIONS* § 11:4, at 15 (5th ed. 2014) (noting "trial bifurcation is widely accepted"); see, e.g., *Beattie v. CenturyTel, Inc.*, 511 F.3d 554, 564-66 (6th Cir. 2007) (Federal Communications Act); *Chiang v. Veneman*, 385 F.3d 256, 273 (3d Cir. 2004) (Equal Credit Opportunity Act); *Bertulli v. Indep. Ass'n of Cont'l Pilots*, 242 F.3d 290, 298 (5th Cir. 2001) (Labor Management Reporting and Disclosure Act and Railway Labor Act).

F. Sampling and Extrapolation

To address classwide damages issues raised at the class certification stage, an innovative plaintiff should consider proposing to select a random sample of class members' damages claims for adjudications and then extrapolate those results to the rest of the class.¹⁹⁷ As demonstrated in Part IV, a properly devised sampling methodology can achieve the goals of Rule 23(b)(3) predominance and superiority: "economies of time, effort, and expense," while promoting "uniformity of decision as to persons similarly situated."¹⁹⁸ Unquestionably, a federal district court has the authority to entertain such a proposal. Rule 23(d)(1)(A) grants the court broad discretion to "issue orders that . . . determine the course of [class action] proceedings or prescribe measures to prevent undue repetition or complication in presenting evidence or argument."¹⁹⁹ A sampling procedure has obvious efficiency benefits, as it obviates the need to expend resources trying numerous claims, thereby reducing the parties' transaction costs, as well as conserving valuable judicial resources.

Moreover, as also demonstrated in Part IV, if a proper statistical methodology is followed, sampling may also produce more accurate and more reliable results than those produced by separate adjudications of each class member's damages claim.²⁰⁰

¹⁹⁷ One or more "bellwether" trials may be held to inform the parties and the court. 4 WILLIAM B. RUBENSTEIN ET AL., NEWBERG ON CLASS ACTIONS § 11:11–11:20, at 41–70 (5th ed. 2014). "Bellwether" refers to:

[A] sheep that leads a flock, around whose neck a bell is hung. In a bellwether trial procedure, a random sample of cases large enough to yield reliable results is tried to a jury. A judge, jury, or participating lawyers use the resulting verdicts as a basis for resolving the remaining cases. Judges currently use bellwether trials informally in mass tort litigation to assist in valuing cases and to encourage settlement.

Alexandra D. Lahav, *Bellwether Trials*, 76 GEO. WASH. L. REV. 576, 577–78 (2008) (citations omitted). As the author points out, "[t]he central difficulty of bellwether trials is the reference class problem . . . caused by heterogeneity within the group of plaintiffs to whom bellwether verdicts are to be extrapolated." *Id.* at 605 (citing Mark Colyvan et al., *Is It a Crime to Belong to a Reference Class?*, 9 J. POL. PHIL. 168, 172 (2001)).

¹⁹⁸ *Amchem*, 521 U.S. at 615 (quoting FED. R. CIV. P. 23 advisory committee's notes to 1966 amendment); *cf.* *Barnes v. District of Columbia*, 278 F.R.D. 14, 20 (D.D.C. 2011) (selecting non-random sample of class members, up to fifteen selected by each side, to determine aggregate damages).

¹⁹⁹ FED. R. CIV. P. 23(d)(1)(A).

²⁰⁰ See Hillel J. Bavli, *Aggregating for Accuracy: A Closer Look at Sampling and Accuracy in Class Action Litigation*, 14 L. PROBABILITY & RISK 67, 69, 78–83 (2015) [hereinafter *Aggregating for Accuracy*] (deriving "conditions under which sampling will increase accuracy"); Alexandra D. Lahav, *The Case for "Trial by Formula"*, 90 TEX. L. REV. 571, 597 (2012) ("The practice of informal sampling in aggregate litigation demonstrates that in being forced to give reasons for how similarly situated people are treated, adjudicators produce a system that is fairer across the board than individual, decentralized litigation."); Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815, 833, 851 (1992) (arguing that "[a]ggregation, properly

As one commentator explained:

For statistical evidence, [“the reason why the sample average is allowed to affect or control outcomes”] is to achieve the best possible jury or judge determination of the relevant issue based on the facts of an individual case. Since the aim is to get as close to the right decision as possible, the focus is on the probative value of the sample average compared to other available evidence.²⁰¹

However, as the brief bleak history of sampling evidence in class actions reveals, *supra* Part I, few federal courts have allowed this form of statistical evidence to prove classwide issues. But, as suggested *supra* Part II, the fault often lies not with the reliability and fairness of the methodology, but rather with its improper application. By way of examples only, courts have properly rejected statistical evidence that does not conform to rigorous scientific standards.²⁰²

conducted, will provide awards that are more accurate, not less” and that aggregation “can systematically increase accuracy [and] reduce bias”); Lawrence Walker & John Monahan, *Sampling Damages*, 83 IOWA L. REV. 545, 567 (1998) (noting “[t]he availability of highly efficient survey techniques for proving damages would open the door to certification of many proposed mass tort class actions”); Lawrence Walker & John Monahan, *Sampling Evidence at the Crossroads*, 80 S. CAL. L. REV. 969, 988 (2007) (random sampling may produce “a more accurate picture of the facts at issue than the study of each case one-by-one”); see also David Rosenberg, *A Sampling-Based System of Civil Liability*, 15 THEORETICAL INQUIRIES L. 635, 649 (2014) (“[W]hen deterrence motivates our desire for knowledge about defendants’ liability and damages, then . . . ‘accuracy’ is as fully achieved by statistical, on-average aggregate assessments as by particularistic, individualized determinations.”).

²⁰¹ Bone, *supra* note 13, at 613.

²⁰² See, e.g., *DiCuio v. Brother Int’l Corp.*, 653 F. App’x 109, 112–14 (3d Cir. 2016) (comparing *Tyson* and affirming summary judgment in favor of defendant in putative consumer fraud class action where “a jury could [not] draw such an inference” based on plaintiffs’ “representative evidence”); *Espenscheid*, 705 F.3d at 774 (“There is no suggestion that sampling methods used in statistical analysis were employed to create a random sample of class members to be the witnesses, or more precisely random samples, each one composed of victims of a particular type of alleged violation.”); *Marlo v. UPS, Inc.*, 639 F.3d 942, 948–49 (9th Cir. 2011) (survey designer “stated in her deposition that she did not know whether the sample was representative”), for additional opinions, see 453 F. App’x 682 (9th Cir. 2011); *Senne v. Kansas City Royals Baseball Corp.*, 315 F.R.D. 523, 587–90 (N.D. Cal. 2016) (applying “‘tailored’ *Daubert* standard” and excluding expert evidence supporting motion to certify Rule 23(b)(3) class because methodology and survey results were “unreliable” due to “self-interest bias” among other things); *United States ex rel. Wall v. Vista Hospice Care, Inc.*, No. 3:07-cv-00604-M, 2016 WL 3449833, at *13–14 (N.D. Tex. June 20, 2016) (granting defendant summary judgment on all but one claim because relator’s “statistical evidence,” including extrapolation, was “unreliable” and expert’s analysis “deficient, because his methodology was fundamentally flawed” due to “failure to select a random sample or to account for relevant variables”); *In re Celexa & Lexapro Mktg.*, 315 F.R.D. 116, 127–28 (D. Mass. 2016) (denying Rule 23(b)(3) class certification in RICO case because of “fundamental flaw” in plaintiff’s expert’s approach); *Balasanyan v. Nordstrom, Inc.*, 294 F.R.D. 550, 568 (S.D. Cal. 2013) (“It is unclear whether the employees were randomly selected or even whether the population of employees was statistically significant (assuming that the selection was random, which, again, is unclear).”).

In addition to obvious economies, the primary argument to admit statistical evidence is that it assists the fact finder to determine a disputed issue based on the facts of the case.

The most common and salient argument against a sampling methodology is that extrapolating from adjudications of sample claims to non-adjudicated claims introduces error.²⁰³ For example, although sampling may enable all class members with valid claims to recover, it may result in over-recovery and under-recovery, and allow class members to recover who do not have valid claims. If all claims were separately adjudicated, the argument goes, extrapolation error would be eliminated. This argument, however, ignores, among other important things, that triers of fact, especially juries, introduce error in the form of unpredictability. With proper sampling, such error can be minimized. Thus, although sampling may introduce extrapolation error, it reduces judgment variability.²⁰⁴

In the end, “whether sampling or individual adjudication produces more accurate results depends on the homogeneity of the group of cases and an empirical question about jury behavior.”²⁰⁵ In addition:

[A] sample picked randomly from the correct reference class will yield fair results, so long as the extrapolation process is able to take into account objectively verifiable variables and does not systematically devalue certain categories of claims for socially undesirable and legally impermissible reasons.²⁰⁶

IV. ADMISSIBILITY OF SAMPLING EVIDENCE TO PROVE INDIVIDUAL DAMAGES

A. Revisiting Tyson

In *Tyson*, the Supreme Court unanimously ruled that “[o]ne way for [plaintiffs] to show . . . that the sample relied upon . . . is a permissible method of proving classwide liability is by showing that each class member could have relied on that sample to establish liability if he or she had brought an individual action.”²⁰⁷ The Court explained that, “[i]f the sample could have sustained a reasonable jury finding as to hours worked in each employee’s individual action, that sample is a permissible means of establishing the

²⁰³ Saks & Blanck, *supra* note 200, at 833.

²⁰⁴ See *infra* Part IV. See generally *Aggregating for Accuracy*, *supra* note 200, at 75–78; Edward K. Cheng, *When 10 Trials Are Better Than 1000: An Evidentiary Perspective on Trial Sampling*, 160 U. PA. L. REV. 955, 958–60 (2012); Saks & Blanck, *supra* note 200, at 833–37.

²⁰⁵ Cheng, *supra* note 204, at 960; see *Aggregating for Accuracy*, *supra* note 200, at 78–83 (deriving homogeneity conditions necessary and sufficient to improve accuracy).

²⁰⁶ Lahav, *supra* note 200, at 617.

²⁰⁷ *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1046 (2016); *id.* at 1050 (Roberts, C.J., concurring); *id.* at 1053 (Thomas, J. dissenting).

employees' hours worked in a class action."²⁰⁸ By parallel reasoning, sampling evidence is a permissible method of proving individual damages in a class action if each class member could rely on it to establish damages in an individual action. The admissibility of sampling evidence to prove individual damages in a class action therefore depends on whether it is relevant and reliable to prove damages in an individual action brought by any member of the class.²⁰⁹

The relevance inquiry can be quickly and positively answered. As long as sampling evidence is based on the facts of the case, that is, it "fit[s]" the case, and there is "a valid scientific connection to the pertinent inquiry," for example, individual damages, sampling evidence is relevant.²¹⁰ Assuming the relevance of sampling evidence, we build on recent articles regarding sampling and accuracy to show that, under appropriate circumstances, sampling evidence is admissible to prove individual damages at the Rule 23(b)(3) class certification stage, even where individual damages claims are known to be heterogeneous.

As discussed in Parts II and III, plaintiffs have attempted to introduce sampling evidence to establish individual damages to satisfy the Rule 23(b)(3) predominance prerequisite.²¹¹ Where individual damages cannot be calculated formulaically, federal district courts frequently find that individual issues predominate over common issues and deny Rule 23(b)(3) class certification.²¹² But, if plaintiffs can demonstrate that sampling evidence pertaining to the entire class can be reliably applied to determine individual damages, then they should prevail in convincing a federal district court that common issues predominate over individual ones.

Consider the circumstances in *Tyson*. Plaintiffs proposed a method to calculate classwide liability and aggregate damages. Tyson argued that plaintiffs' expert's study was "unrepresentative or inaccurate."²¹³ As explained below, in such circumstances, class plaintiffs can counter by arguing that sampling evidence (similar to that introduced in *Tyson*) can not only reliably prove classwide liability and aggregate damages, but can determine individual damages as well.

Assume that videotaped observations of a representative sample of workers in the class reveal that each sampled worker spends approximately twenty minutes "donning" and "doffing" safety gear and that "donning" and "doffing" times are the only factor that distinguishes one class member's

²⁰⁸ *Id.* at 1046-47.

²⁰⁹ See *id.*; *supra* notes 82-137 and accompanying text.

²¹⁰ See *Daubert v. Merrell Dow Pharms.*, 509 U.S. 579, 591-92 (1993).

²¹¹ See *supra* notes 178-189 and accompanying text.

²¹² See *supra* note 187 and accompanying text.

²¹³ *Tyson*, 136 S. Ct. at 1047.

damages claim from another.²¹⁴ Plaintiffs may persuasively argue, relying on *Tyson*, that expert testimony regarding the sampling evidence is reliable to establish damages for each class member and is admissible to prove individual damages. Therefore, the sampling evidence would also be admissible to satisfy the Rule 23(b)(3) predominance prerequisite.²¹⁵

Now, assume that the videotaped observations disclose more variability among the sampled workers' "donning" and "doffing" times. In that case, the pertinent question is: How variable is too variable? Or, in other words, what degree of homogeneity should be required to demonstrate that sampling evidence is reliable to prove individual damages? For example, what variance of "donning" and "doffing" times should be acceptable for a federal district court to admit the sampling evidence to prove individual damages?

It is important to keep in mind that a federal district court ruling on the admissibility of sampling evidence is not deciding whether the finder of fact must rely on it, or, for example, the sample mean in particular, to calculate individual damages for each class member. Rather, the court is only deciding whether the sampling evidence is sufficiently reliable to present to the trier of fact for consideration when awarding individual damages.²¹⁶ Ignoring, for now, the possibility that some class members do not have valid claims because they did not work in excess of forty hours in a week, that is, they cannot prove that the class defendant is liable to them, we posit that, at some level of homogeneity, the sampling evidence is reliable and therefore is admissible into evidence to prove individual damages.

We will demonstrate below that one method to satisfy Rule 23(b)(3) predominance, notwithstanding variable individual damages, is to show: (1) a sample of class members' claims is representative and provides reliable measures of central tendency and variability; and (2) individual damages are not too variable. We assume that plaintiffs will propose a sampling method "ground[ed] in the methods and procedures of science" to meet the first criterion.²¹⁷ Therefore, we now address the second criterion, and determine how variable is too variable.

²¹⁴ Alternatively, consider circumstances where "donning" and "doffing" times are found by a court, acting as the finder of fact, to be homogeneous. For example, if a federal district court finds, based on sampling evidence, that most workers perform "donning" and "doffing" activities in twenty minutes, and that that is also approximately the minimum "donning" and "doffing" time in the sample, the court could find that more than twenty minutes is unreasonable, and use twenty minutes as the basis for calculating individual damages.

²¹⁵ See *supra* notes 82–206 and accompanying text.

²¹⁶ See generally *Tyson*, 136 S. Ct. at 1036.

²¹⁷ *Daubert*, 509 U.S. at 590. The "principles and methods" of sampling are generally recognized "methods and procedures of science." FED. R. EVID. 702; *Daubert*, 509 U.S. at 590; see *supra* notes 138–206 and accompanying text.

B. Threshold for Homogeneity

We derive a threshold for determining whether class damages claims are sufficiently homogeneous to justify the admissibility of sampling evidence to prove individual damages. We begin by developing a framework for our analysis by defining the concepts of judgment variability, claim variability, and the accuracy of damages awards. We argue that, for purposes of satisfying the predominance prerequisite of Rule 23(b)(3), accuracy is an appropriate standard for evidentiary reliability. Then, we show that, when judgment variability exceeds claim variability, sampling evidence improves accuracy and thereby establishes evidentiary reliability.

C. Claim Variability and Judgment Variability

In a recent article, one of the authors described the concepts of claim variability and judgment variability. As defined, claim variability represents the substantive, that is, factual, differences among a set of claims. Judgment variability, on the other hand, is the randomness associated with the adjudication of a claim, or specifically, the differences in the outcomes that would result from repeated adjudications of the same claim.²¹⁸ Thus, for present purposes, we postulate that, if each class member's damages claim were adjudicated individually, the variability of the resulting damages awards could be attributed to both claim variability and judgment variability. To refine the inquiry addressed here: What threshold amount of claim variability is acceptable for sampling evidence to be admissible to prove individual damages?

D. Variability, Accuracy, and Evidentiary Reliability

As discussed in Part II, among the "general standards" that a federal district court "must use to assess the reliability . . . of proffered expert testimony," are whether the "testimony is the product of reliable principles and methods" and whether the "expert has reliably applied the principles and methods to the facts of the case."²¹⁹ According to the *Daubert* Court, evidentiary reliability means "trustworthiness" and, "[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific*

²¹⁸ *Aggregating for Accuracy*, *supra* note 200, at 75–78; see Hillel J. Bavli, *Sampling and Reliability in Class Action Litigation*, 2016 CARDOZO L. REV. DE NOVO 207, 209–10, http://cardozolawreview.com/content/denovo/BAVLI_2016_207.pdf [<https://perma.cc/QZ5F-GG7N>]; Saks & Blanck, *supra* note 200, at 833–37.

²¹⁹ FED. R. EVID. 702; FED. R. EVID. 702 advisory committee's notes to 2000 amendment; see *supra* notes 138–206 and accompanying text.

validity.”²²⁰ In turn, scientific validity is an affirmative answer to the question: “[D]oes the principle support what it purports to show?”²²¹

In the context of sampling to determine individual damages for Rule 23(b)(3) class members (and, likely, in most contexts), evidentiary reliability can be best understood in terms of accuracy: Does the sampling evidence produce a more accurate determination of individual damages?

More formally, assume that, for each class member’s claim, there is a “correct” amount of damages that would be awarded if the trier of fact were omniscient, that is, possessed complete information regarding the material facts of the claim and certainty about the governing law.²²² Further, for simplicity, assume that the “correct” damages award would be the arithmetic mean of the outcomes of an infinite number of adjudications of the claim under different circumstances (*e.g.*, different jurors, judges, lawyers, presentations of evidence, etc.).²²³ As explained in *Aggregating for Accuracy*, because no trier of fact is omniscient, the trier of fact must arrive at an estimate of the “correct” amount of damages.²²⁴ We can then define error and accuracy, respectively, in terms of the distance of an actual damages award from, and the proximity of an actual damages award to, the “correct” damages award.²²⁵

Specifically, we can define error in terms of mean squared error (“MSE”), which is a combination of bias and variance.²²⁶ Here, bias represents “systematic error,”²²⁷ or the difference between the “correct” award and the expected award (*i.e.*, the mean of the awards that would result from infinite separate adjudications). For example, sampling evidence

²²⁰ 509 U.S. at 590 n.9 (citation omitted).

²²¹ *Id.*

²²² *Aggregating for Accuracy*, *supra* note 200, at 74–75; Bavli, *supra* note 218, at 209–10; Saks & Blanck, *supra* note 200, at 833–34. See generally RONALD DWORKIN, *LAW’S EMPIRE* (1986) (describing “an imaginary judge [called Hercules] of superhuman intellectual power and patience” who seeks to identify “the best constructive interpretation of the community’s legal practice,” *id.* at 239, 225). Alternatively, it can be assumed that there is a distribution of “correct” damages awards that reflects, for example, incomplete information or uncertainty about the governing law.

²²³ *Aggregating for Accuracy*, *supra* note 200, at 74–75; Bavli, *supra* note 218, at 209–10; Saks & Blanck, *supra* note 200, at 833–34. Various measures of central tendency can be used to characterize the “correct” damages award. We use the arithmetic mean. Note that our assumption that an award is “correct” on average is conservative in the sense that all changes in the magnitude of a damages award caused by the introduction of sampling evidence are assumed to harm the accuracy of the award—regardless whether such changes are in fact harmful or helpful in practice.

²²⁴ *Aggregating for Accuracy*, *supra* note 200, at 74.

²²⁵ *Id.* at 74–75.

²²⁶ *Id.* at 81–83. More precisely, MSE is defined as the expectation of the squared differences between the “correct” awards and the actual awards, which is equal to a combination of the squared bias and variance. See Hillel J. Bavli, *The Logic of Comparable-Case Guidance in the Determination of Awards for Pain and Suffering and Punitive Damages*, 85 U. CIN. L. REV. 1, 15 (2017) [hereinafter *The Logic of CCG*] (defining MSE in the damages context).

²²⁷ See *Reference Guide on Statistics*, *supra* note 32, at 240.

introduces bias (under our assumptions) if it causes a change in the expected magnitude of an award. Variance represents “random error,” or the dispersion of awards around the mean award.²²⁸ For example, sampling evidence reduces the variance of an award if it would reduce the dispersion of repeated adjudications of the corresponding claim. Finally, we can define accuracy as the inverse of error: As error decreases, accuracy increases, and vice versa.

Consider again the factual circumstances in *Tyson*. Assume that we selected a subclass of workers, and that a representative sample of videotaped observations showed that their mean “donning” and “doffing” time was eighteen minutes. If the damages awarded to each selected class member were based on the assumption that each worker spent eighteen minutes each day “donning” and “doffing” safety gear, each award would involve error reflecting the difference between damages based on eighteen minutes and damages based on the worker’s actual “donning” and “doffing” times.²²⁹

The size of the error would depend, in part, on claim variability, that is, the variability in all the workers’ claims for “donning” and “doffing” times in excess of forty hours, because they took more or less than eighteen minutes to perform these tasks. In addition, the error would depend on judgment variability. For example, in *Tyson*, notwithstanding plaintiffs’ expert’s calculations that supported an aggregate damages award of approximately \$6.7 million, the jury awarded classwide damages of about \$2.9 million.²³⁰ A different jury could have awarded the amount requested (\$6.7 million), or \$1 million, or some other amount. Judgment variability notwithstanding, however, there are actual (“true”) “donning” and “doffing” times for each class member. If the trier of fact were omniscient, the “correct” damages would be awarded to each class member based on the “true” times. Therefore, the error associated with using eighteen minutes “donning” and “doffing” times to calculate individual damages results from both claim variability and judgment variability.²³¹

²²⁸ *The Logic of CCG*, *supra* note 226, at 14–15; *Reference Guide on Statistics*, *supra* note 32 at 239–40 (“The standard deviation [square root of variance] is a sort of mean deviation from the mean.”). Variance is defined more formally as the expectation of the squared differences between the actual damages awards and the expected damages award. See *The Logic of CCG*, *supra* note 226, at 14–15. Note that our analysis does not rely on mean squared error as the definition of error. We use mean squared error for computational and interpretational convenience, but alternative measures of error could be used.

²²⁹ Keep in mind that we are still assuming that all class members have valid claims and are entitled to damages awards because they worked more than forty hours in one or more weeks.

²³⁰ 136 S. Ct. at 1044.

²³¹ Note that, although we define the concept of a “correct” award for simplicity, our analysis does not rely on the existence of a single “correct” award. See *supra* notes 222–225 and accompanying text.

E. Implications of Sampling for Accuracy and Reliability

Sampling impacts the error resulting from both forms of variability, claim variability and judgment variability. For example, sampling can reduce error resulting from judgment variability by aggregating information across class members' claims.²³² To illustrate, consider the effect of replication, or multiple adjudications of the same individual claim:

Consider a costly procedure through which the outcome of a claim is determined by averaging the verdicts resulting from ten independent "replications" of the trial, or "repeated adjudications" (involving, for example, different judges, juries, attorneys, presentations of evidence, etc.). Assuming the outcome is relatively unbiased, it is easy to show that following this procedure results in an accurate outcome—an outcome that is close to the "correct" outcome. Similarly, this procedure will produce an accurate outcome for each claim of each member of a putative class (or subclass). Replication thus increases the reliability of legal outcomes by reducing the error caused by judgment variability.²³³

Without doubt, a replication procedure would be costly. But, in a class of identical, that is, homogeneous, claims, the full accuracy benefits of replication could be realized by extrapolating damages awards based on a representative sample of adjudicated claims to all non-adjudicated class claims.²³⁴ Moreover, the benefits of replication obtain whether the sampled units are damages awards, or a determinative variable, such as "donning" and "doffing" times.²³⁵

The benefits of replication (and resulting aggregation of information about class damages claims) can also be realized with heterogeneous class damages claims. But, with heterogeneous claims, the accuracy benefits of replication with respect to judgment variability must be balanced with the introduction of error that results from applying a single aggregated value—for

²³² See *Aggregating for Accuracy*, *supra* note 200, at 81–83; Saks & Blanck, *supra* note 200, at 833–34.

²³³ Bavli, *supra* note 218, at 211; see also *Aggregating for Accuracy*, *supra* note 200, at 77.

²³⁴ Bavli, *supra* note 218, at 211; see also *Aggregating for Accuracy*, *supra* note 200, at 77. Although a federal district court may extrapolate damages awards for non-sampled claims, it may not, due to constitutional constraints and other reasons, replace adjudicated damages awards with extrapolated damages awards. See *Aggregating for Accuracy*, *supra* note 200, at 78–81 (citing cases); *supra* notes 62–64.

²³⁵ See *infra* notes 254–259 and accompanying text for a more detailed discussion of these types of representative evidence (that is, representative evidence based on damages awards or determinative variables) and the relationship between representative adjudications and representative evidence.

example, the mean of the damages awards from repeated adjudications—to a class of variable claims that may have distinct “correct” outcomes.²³⁶

Returning to a federal district court’s “gatekeeping” role under *Daubert*, the admissibility of sampling evidence depends on whether the evidence is reliable. Restated for present purposes, admissibility depends on whether sampling evidence is “trustworthy” to prove individual damages.²³⁷ We argue here that evidentiary reliability is substantively equivalent to the concept of reliability explicated in *Sampling and Reliability in Class Action Litigation*: “[T]he reliability of a legal procedure [is] the accuracy of the legal outcome that can be expected by following the procedure.”²³⁸ In particular, the reliability of sampling evidence is based on its effect on the accuracy of the individual damages awards. There is substantial support for this argument. The *Daubert* Court emphasized that, “[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*.”²³⁹ And scientific validity “refers to the ability of a test to measure what it is supposed to measure—its accuracy.”²⁴⁰

In this regard, the *Reference Guide on Statistics* and the *Reference Manual on Scientific Evidence* more generally discuss two components of error—what can be considered as variance and bias—when assessing the validity of a measurement.

First, the *Reference Guide on Statistics* highlights the important role of scientific “reliability,” or consistency, in determining whether a measurement is scientifically “valid,”²⁴¹ as well as the harmful effects of “random error” on estimation, and statistical inference more generally.²⁴² It describes two different types of variability—“within-observer variability” and “between-observer variability”—as relevant to assessing whether a measurement is scientifically reliable, and therefore whether the measurement is accurate.²⁴³ Applying these concepts to our analysis, and viewing adjudications of class members’ claims as an estimation problem, they can be understood as elements of judgment variability. A trier of fact adjudicating the same

²³⁶ *Aggregating for Accuracy*, *supra* note 200, at 82–83; *see also* Bavli, *supra* note 218, at 212–14.

²³⁷ *See Daubert*, 509 U.S. at 590 n.9.

²³⁸ Bavli, *supra* note 218, at 210 (emphasis omitted).

²³⁹ 509 U.S. at 590 n.9; *see also* Paul C. Giannelli et al., *Reference Guide on Forensic Identification Expertise*, in *REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* 55, 71–72 (Federal Judicial Center, 3d ed. 2011) [hereinafter *Reference Guide on Forensic Identification Expertise*].

²⁴⁰ *Reference Guide on Forensic Identification Expertise*, *supra* note 239, at 71–72; *Reference Guide on Statistics*, *supra* note 32, at 228 (“A valid measuring instrument measures what it is supposed to.”); *see Daubert*, 509 U.S. at 590 n.9 (“[D]oes the principle support what it purports to show?”).

²⁴¹ *Reference Guide on Statistics*, *supra* note 32, at 227–28; *see also Reference Guide on Forensic Identification Expertise*, *supra* note 239, at 71–72.

²⁴² *Reference Guide on Statistics*, *supra* note 32, at 240–41.

²⁴³ *See id.* at 228.

individual class member's damages claim multiple times under different circumstances may award different amounts of damages, representing "within-observer variability." In addition, different triers of fact adjudicating the same claim under identical circumstances may determine different damages awards, representing "between-observer variability." Thus, when viewed in this light, the concept of accuracy, or "scientific validity," is based, in part, on judgment variability.²⁴⁴

Second, as explained in the *Reference Manual on Scientific Evidence*, "[v]alidity includes [scientific] reliability, but the converse is not necessarily true. Thus, a [scientifically] reliable, invalid technique will consistently yield inaccurate results."²⁴⁵ "In addition to reliability, validity is needed. A valid measuring instrument measures what it is supposed to."²⁴⁶ Therefore, the second component of validity, as described in the *Reference Guide on Statistics*, can be understood as bias. Contrasting this type of error from "random chance," or "random error," in the context of "inferences that may be drawn from a study," the *Reference Guide on Statistics* remarks, "[t]he data might not address the issue of interest, might be systematically in error, or might be difficult to interpret because of confounding. Statisticians would group these concerns together under the rubric of 'bias.'"²⁴⁷

In sum, reliability of damages evidence can be evaluated in terms of its effect on the accuracy of the damages award. By this standard, where sampling evidence can be expected to improve accuracy, the evidence should be considered reliable and therefore admissible.²⁴⁸

²⁴⁴ See *Reference Guide on Forensic Identification Expertise*, *supra* note 239, at 71–72; *Reference Guide on Statistics*, *supra* note 32, at 227–28 ("A reliable measuring instrument returns consistent measurements."). As noted, "'within-observer variability' of judgments should be small—the same evaluator should rate essentially identical cases in similar ways" and "'between-observer variability' should be small—different evaluators should rate the same cases in essentially the same way." *Reference Guide on Statistics*, *supra* note 32, at 228.

²⁴⁵ *Reference Guide on Forensic Identification Expertise*, *supra* note 239, at 71–72; *Reference Guide on Statistics*, *supra* note 32, at 228 ("Reliability is necessary but not sufficient to ensure accuracy.").

²⁴⁶ *Reference Guide on Statistics*, *supra* note 32, at 228.

²⁴⁷ *Id.* at 240. The *Reference Guide on Statistics* also describes validity in terms of correlation, which, similarly, can be stated in terms of bias and variance. See *id.* at 228 ("the polygraph is not valid as a lie detector unless the measurements it makes are well correlated with lying"; "[a] common measure of validity is the correlation coefficient between the predictor and the criterion (e.g., test scores and later performance)").

²⁴⁸ The expectation that sampling evidence improves accuracy is a sufficient but not a necessary condition for improving reliability. Where sampling evidence cannot be expected to improve accuracy, it does not necessarily follow that such evidence is not reliable. If sampling evidence cannot be expected to improve accuracy (and, moreover, if it cannot be shown to be expected to improve accuracy), this does not necessarily imply that such evidence is expected to reduce accuracy.

F. Threshold for Admissibility

Consider again circumstances similar to *Tyson*. Consider two scenarios: one, sampling evidence is admitted to prove individual damages, or two, Rule 23(b)(3) class certification is denied (at least for purposes of individual damages, but perhaps for liability and aggregate damages as well) and class members' individual damages claims must be separately adjudicated. In the second scenario, we do not require that each class member actually proceed with an individual adjudication. We only require that class certification is denied and claims will be resolved by individual adjudications, if at all.

Our immediate goal is to derive threshold heterogeneity conditions, that is, threshold claim variability conditions, under which sampling evidence is admissible to prove individual damages for all class members' damages claims. We focus on the admissibility of the average of damages awarded for sampled claims, or the average of other operative values, such as "donning" and "doffing" times in *Tyson*, in sampled claims. Specifically, we focus on the admissibility of the mean of such values, which we call the "sample mean." We are seeking conditions under which sampling evidence improves accuracy, notwithstanding the heterogeneity of the claims. In short, we want to know the values of claim variability for which sampling evidence can be said "to measure what it is supposed to measure," ensuring its scientific validity and thus its evidentiary reliability.²⁴⁹ Below, we show that this threshold is the value at which claim variability equals judgment variability. We conclude that using sampling evidence, and specifically the sample mean, to estimate the damages for each claim of a heterogeneous class improves accuracy when judgment variability exceeds claim variability.

G. Hierarchical Model

Let us consider the problem in more formal terms. Assume that the damages award for a claim i , call it X_i , is statistically distributed with mean α_i and variance σ^2 , for all $i = 1, 2, \dots, N$. Notationally, $X_i \sim (\alpha_i, \sigma^2)$.²⁵⁰ This simply means that each of the N claims will equal its "correct" award, α_i , on average, but that the award involves some degree of randomness or judgment variability, σ^2 .²⁵¹ Furthermore, to reflect the fact that the claims are related (as required to satisfy Rule 23(a)(2) commonality), assume that the "correct" awards for the claims, the α_i 's, are distributed with a "global" mean μ and

²⁴⁹ *Reference Guide on Forensic Identification Expertise*, *supra* note 239, at 71–72; *Reference Guide on Statistics*, *supra* note 32, at 228; see *Daubert*, 509 U.S. at 590 n.9.

²⁵⁰ See *Aggregating for Accuracy*, *supra* note 200, at 81–83; *The Logic of CCG*, *supra* note 226, at 15–18.

²⁵¹ See *Aggregating for Accuracy*, *supra* note 200, at 81–83; *The Logic of CCG*, *supra* note 226, at 15–18.

variance τ^2 . Notationally, $\alpha_i \sim (\mu, \tau^2)$.²⁵² In other words, the “correct” damages awards for the claims will equal the global mean, or μ , on average, but display some degree of variability, τ^2 , that is, claim variability. This recognizes the fact that the claims may differ substantively and involve distinct “correct” awards.²⁵³

This model, known in statistics as a standard two-level hierarchical (or multilevel) model, allows for a very general methodology with few assumptions regarding the damages awards. The statistical model is depicted in Figure 1 below.²⁵⁴

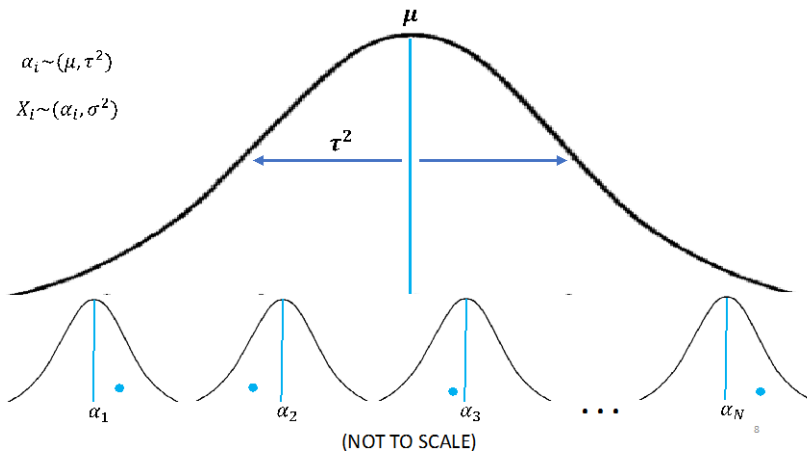


Figure 1. Image of hierarchical model with lower level $X_i \sim (\alpha_i, \sigma^2)$ and upper level $\alpha_i \sim (\mu, \tau^2)$. Note that the shape and size of the distributions are for illustrative purposes only and are not drawn to scale. For example, the distributions are not necessarily “normal,” or in the shape of a bell curve, and the dispersion, or variance, of the lower level distributions are not necessarily less than that of the upper level. The points at the bottom of the lower-level distributions represent possible values of X_i , reflecting the respective means of each claim i and judgment variability σ^2 .

Now, let us temporarily simplify the problem through abstraction for purposes of deriving an appropriate threshold. We are interested in measuring, or estimating, the α_i ’s (“correct” damages awards). We want to know for what values of claim variability, τ^2 , would applying the mean, \bar{X}_n , of a random sample of n claims from a total of N claims, rather than

²⁵² See *Aggregating for Accuracy*, *supra* note 200, at 81–83; *The Logic of CCG*, *supra* note 226, at 15–18.

²⁵³ See *Aggregating for Accuracy*, *supra* note 200, at 81–83; *The Logic of CCG*, *supra* note 226, at 15–18.

²⁵⁴ This graphic is available at <http://www.bc.edu/content/dam/bc1/schools/law/pdf/law-review-content/BCLR/59-2/bavli-felter-graphics-A1b.pdf> [<https://perma.cc/MJ63-4VTG>].

individual damages awards, improve accuracy, where the criterion for accuracy is MSE (summed over *N*).

Importantly, this framework applies whether the sampled units are damages awards—see, for example, discussions of *Cimino* (involving representative adjudications) and *Hilao* (involving representative evidence), *supra* Part I—or, as in *Tyson*, a variable that is determinative of the damages awards.²⁵⁵ In *Tyson*, for example, the sample “donning” and “doffing” times may be converted to damages awards (or vice versa) by applying a constant multiplier, that is, 150% of the hourly wage for each hour worked in excess of forty hours each week.

²⁵⁵ Bavli, *supra* note 218, at 217–18; see *Hilao v. Estate of Ferdinand Marcos*, 103 F.3d 767 (9th Cir. 1996); *Cimino v. Raymark Indus., Inc.*, 739 F. Supp. 328 (E.D. Tex. 1990); cf. *Tyson*, 136 S. Ct. at 1041. *Sampling and Reliability in Class Action Litigation* explains the relationship between representative evidence and representative adjudications as follows:

[I]n *Aggregating for Accuracy*, an award in a heterogeneous class is modeled hierarchically: the “correct” awards in the class are distributed around some global mean, whereas each actual award is “drawn” from a distribution around each claim’s “correct” award. The former distribution represents claim variability, whereas the latter distribution represents judgment variability. In a homogeneous class, replication offers accuracy benefits by providing additional information regarding the “correct” award associated with the replicated claim, which otherwise would be obscured by judgment variability. In a heterogeneous class, sampling offers accuracy benefits, with respect to a certain claim, not by providing information regarding that claim’s “correct” award directly, but by providing information regarding the global mean around which all of the “correct” awards are distributed, and thereby regarding the “correct” award for the subject claim indirectly. Similarly, representative evidence, such as the type in dispute in *Tyson Foods* (where, for example, the sample reflects variability of measured donning and doffing times rather than judgment variability), offers accuracy benefits, with respect to a certain claim, by providing information regarding the global mean around which the “correct” awards are distributed, and thereby regarding the “correct” award for that claim in particular. Another way of understanding this is through “comparable-case guidance” (CCG) methods, whereby a court uses information regarding awards in prior comparable cases as guidance for a fact-finder’s determination of damages. *The Logic of CCG* examines the statistical mechanism by which CCG affects awards, and the conditions under which such evidence will improve accuracy. In particular, the paper explains that, under certain mild behavioral assumptions, the risk that such evidence would reduce accuracy—that error resulting from claim variability and bias would outweigh the accuracy benefits of reducing judgment variability—is minimal. Like CCG, representative evidence provides information regarding the distribution of “correct” awards for comparable claims, including the global mean, and, in turn, about the “correct” award for the subject claim.

Bavli, *supra* note 218, at 218 n.67 (citations omitted). In addition, our model can be extended to representative evidence that is not immediately determinative of the claims’ damages awards.

Thus, let us derive the claim variability conditions, that is, the values of τ^2 , that yield greater accuracy when applying the sample mean, \bar{X}_n , to all claims, rather than an individual outcome, X_i , to each claim i .²⁵⁶

$$E \sum_{i=1}^N (X_i - \alpha_i)^2 > E \sum_{i=1}^N (\bar{X}_n - \alpha_i)^2 \quad (1)$$

$$N\sigma^2 > E \left(\sum_{i=1}^N (\bar{X}_n - \mu)^2 + \sum_{i=1}^N (\mu - \alpha_i)^2 \right) \quad (2)$$

$$\text{It follows that } \sigma^2 > \frac{n+1}{n-1} \tau^2 \quad (3)$$

$$\text{Therefore, } \sigma^2 > \tau^2, \text{ for large } n \quad (4)$$

The converse can be shown to be true as well. Thus, given a sufficiently large sample, the error resulting from individual outcomes is greater than the error resulting from the sample mean (*see* Line (1)), if and only if, the lower-level variance σ^2 (judgment variability) is greater than the upper-level variance τ^2 (claim variability) (*see* Line (4)).²⁵⁷ In other words, using the sample mean (the mean of awards for sampled class members' damages claims) rather than individual outcomes (separate adjudications) results in greater accuracy, if and only if, judgment variability is greater than claim variability. Furthermore, conversely, assuming a reasonable sample size—which we will assume for the remainder of Part IV—if judgment variability is greater than claim variability, using the sample mean, rather than individual outcomes, improves accuracy.

This result is intuitive. The accuracy benefit associated with the use of the sample mean arises from the sharing of information across claims. Information sharing reduces the error caused by judgment variability. On the other hand, using the sample mean, rather than individual outcomes, gives rise to error resulting from applying a single point estimate to variable claims. By applying one value, albeit a value that benefits from information sharing, to heterogeneous claims that involve distinct “correct” awards, necessarily introduces error.

²⁵⁶ This graphic is available at <http://www.bc.edu/content/dam/bc1/schools/law/pdf/law-review-content/BCLR/59-2/bavli-felter-graphics-A1b.pdf> [<https://perma.cc/MJ63-4VTG>].

²⁵⁷ As shown in the derivation (Line (3)), the necessary sample size will depend on the values of judgment variability and claim variability. However, if the estimated judgment variability is greater than the estimated claim variability, *see infra* notes 267–277 and accompanying text, a moderately-sized random sample would generally suffice to satisfy standards of reliability. Note, we could alternatively derive this result using certain additional assumptions.

Consider again the image in Figure 1, including the scale of the distributions, which would occur if judgment variability were far lower than claim variability. In that circumstance, there is little to gain from information sharing, because the individual outcomes are likely to approximate their respective “correct” awards. On the other hand, because claim variability is relatively high, using the sample mean would result in a high degree of error arising from the application of a single point estimate to very heterogeneous claims. However, if the lower-level distributions in Figure 1 were larger (*i.e.*, involving greater variance) than the upper-level distribution—that is, if judgment variability were greater than claim variability—then there would be significant gain from information sharing (via the sample mean), which could eliminate significant error caused by random variation, or judgment variability. And, because claim variability would be relatively low, there would be relatively little error arising from applying a single point estimate to the relatively low-variability (although still heterogeneous) claims.

To summarize, when judgment variability exceeds claim variability, applying sampling evidence, and particularly the sample mean, rather than individual outcomes (whether damages awards or, for example, “donning” and “doffing” times) to estimate the “correct” damages awards improves accuracy. Improving accuracy increases evidentiary reliability. In these circumstances, therefore, using sampling evidence improves evidentiary reliability.

It is axiomatic that evidence proving individual damages can be relied on to establish damages in an individual action. When sampling evidence—and a properly calculated sample mean in particular—improves accuracy, and hence reliability, it is admissible in an individual action. In short, the sample mean “could have been sufficient to sustain a jury finding . . . if it were introduced in . . . [an] individual action.”²⁵⁸ It is, therefore, admissible in a class action.

Our analysis does not rely on how a trier of fact will actually be influenced by expert sampling evidence. We simply adopt the standard for evidentiary reliability, which is based on scientific validity, to answer the questions: “[D]oes the principle support what it purports to show?”²⁵⁹ Does it “measure what it is supposed to measure?”²⁶⁰ The foregoing analysis answers each of these questions with a resounding “yes”—when judgment variability is greater than claim variability.

It is noteworthy that the above hierarchical model and related analysis involve minimal assumptions regarding class members’ damages awards and material evidence (*e.g.*, “donning” and “doffing” times), and are generally

²⁵⁸ *Tyson*, 135 S. Ct. at 1048.

²⁵⁹ *Daubert*, 509 U.S. at 590 n.9.

²⁶⁰ *Reference Guide on Forensic Identification Expertise*, *supra* note 239, at 71–72; *Reference Guide on Statistics*, *supra* note 32, at 228.

accepted statistical methods that easily satisfy *Daubert* standards and Federal Rule of Evidence 702 criteria. For example, the model and methodology have “general acceptance,” “can be (and ha[ve] been) tested,” and have been “subjected to peer review and publication.”²⁶¹ Moreover, expert testimony based on them would “help the trier of fact to understand the evidence” and would be “the product of reliable principles and methods,” in compliance with Rule 702.²⁶²

H. Correlation

The above analysis can be restated in terms of correlation, which is among the most generally accepted methods in statistics and the sciences generally.²⁶³

For example, in *Tyson*, consider the correlation between the “correct” damages awards for individual class members’ claims (independent variable), or, instead, the true “donning” and “doffing” times, and actual damages awards (dependent variable). Imagine a graph with the independent variable on the x-axis and the dependent variable on the y-axis, as depicted in the graphs in Figure 2 below. The variability of the data points would reflect the two types of variability we have been discussing—judgment variability and claim variability. The trier of fact does not know the “correct” damages awards but can estimate them based on the evidence introduced at trial. Therefore, there is generally a correlation between “correct” damages awards and actual damages awards (or, equivalently, “true” “donning” and “doffing” times and actual damages awards).

But there is also judgment variability, or random variation. If judgment variability is high relative to claim variability, then the correlation would not be apparent. In particular, the estimate of the correlation coefficient, the value that represents the strength of the relationship between the two variables, would be close to zero, and would likely be found “insignificant,” with a high *p*-value, in the context of a hypothesis test.²⁶⁴ If claim variability were large relative to judgment variability, however, the correlation would be apparent,

²⁶¹ *Daubert*, 509 U.S. at 593–94.

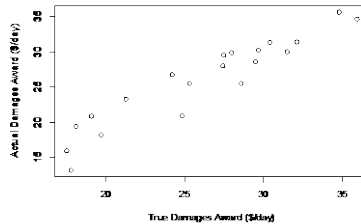
²⁶² FED. R. EVID. 702. Significantly, the hierarchical model and derivation of the foregoing threshold standard for admissibility are abstracted from any specific legal circumstances. This elucidates their general acceptance and tested reliability, without confusion caused by context. Furthermore, the concepts explained, such as judgment variability and claim variability, have strong foundations in the literature (not to mention case law), and meet the requirements of *Daubert* and Rule 702. See generally, Cheng, *supra* note 204, at 955; Saks & Blanck, *supra* note 200, at 833–37.

²⁶³ See, e.g., *Reference Guide on Statistics*, *supra* note 32, at 227–28 (“[g]iven independence, the correlation coefficient . . . between repeated measurements can be used as a measure of reliability”; “[a] common measure of validity is the correlation coefficient between the predictor and the criterion”).

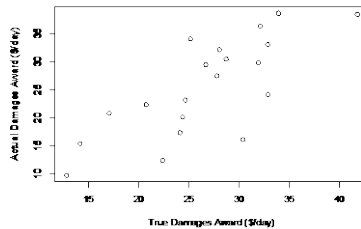
²⁶⁴ See *supra* notes 113–128 and accompanying text.

and the estimated correlation coefficient would likely be found “significant.”²⁶⁵

Graph A



Graph B



Graph C

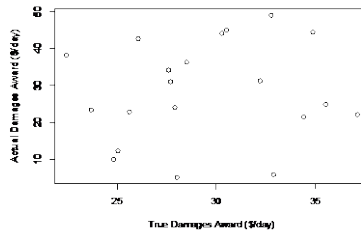


Figure 2. Comparison of correlation between “correct,” or “true,” damages awards and actual damages awards at different levels of judgment variability.

Consider, for example, the graphs in Figure 2.²⁶⁶ Graph A displays circumstances in which judgment variability is low relative to claim variability, and a correspondingly clear and strong correlation. Graph B displays circumstances in which judgment variability is closer to the value of

²⁶⁵ See *supra* notes 113–128 and accompanying text.

²⁶⁶ This graphic is available at <http://www.bc.edu/content/dam/bc1/schools/law/pdf/law-review-content/BCLR/59-2/bavli-felter-graphics-A1b.pdf> [<https://perma.cc/MJ63-4VTG>].

claim variability, and a correspondingly possible—although weak, if any—correlation. Graph C displays circumstances in which judgment variability exceeds claim variability, and a correspondingly zero or negligible correlation. The graphs in Figure 2 illustrate that it may be possible for putative class plaintiffs to argue, using hypothesis testing or confidence intervals, that, if there is no clear correlation between the dependent and independent variables, suggesting that judgment variability exceeds claim variability, then the sampling evidence improves accuracy and satisfies evidentiary criteria for reliability and admissibility.

This approach is effectively a form of statistical “model testing.” The model being tested involves “group means”—or individual outcomes—which would be advantageous (with respect to accuracy) if there were a significant correlation. The alternative model is based on the “sample mean,” where all values of the independent variable are assumed to be associated with the same value—the sample mean—of the dependent variable.

Of course, it is impossible to know the “correct” damages awards. But, reliable estimates can be derived through sampling. For example, in circumstances similar to those in *Tyson*, the independent variable on the x-axis could be replaced with estimated damages awards based on the “donning” and “doffing” times observed in the sample. Plaintiffs may be able to use such estimates to demonstrate their argument in terms of correlation.²⁶⁷

So far, we have not generally addressed how to estimate relevant model parameters, such as judgment variability and claim variability. Below, we offer recommendations.

I. Estimating Claim Variability and Judgment Variability

We have argued that, whether sampling evidence is reliable and therefore admissible to prove individual damages depends on claim variability and judgment variability. These variability parameters can be understood as having “true” fixed values. And, although it is impossible to ascertain their “true” values, it is possible to calculate reliable estimates.²⁶⁸

J. Claim Variability and Samples

Claim variability can often be estimated from the selected sample itself. For example, in wage and hour cases such as *Tyson*, the variability of class members’ damages claims frequently arises from a single variable, a single characteristic of the claims. This is true when the facts underlying one characteristic of the individual claims constitute the only facts that distinguish

²⁶⁷ Note that, for correlation-based methods, because putative class plaintiffs have the burden of proving that their proffered evidence is reliable, it may be insufficient to perform a standard hypothesis test using a significance level of .05.

²⁶⁸ Bavli, *supra* note 218, at 209.

one claim from another; or because the court (*e.g.*, by rulings or orders) or the parties (*e.g.*, by stipulations) have narrowed the disputed material facts to one case-dispositive variable.²⁶⁹ In *Tyson*, the only characteristic of the class members' claims that varied, with respect to damages, was the amount of time spent "donning" and "doffing" safety gear. In such "unidimensional" cases, a representative sample of claims can provide a reliable estimate of claim variability. For example, the claim variability of the damages claims among the workers in the "cut and retrim" departments could be estimated by measuring the variance of the "donning" and "doffing" times in the representative sample of workers in those departments. Or, more concretely (and to facilitate comparison to judgment variability), claim variability could be estimated by measuring the variance of statutory damages based on those times.

Admittedly, some cases are ill-suited for this type of variance measurement. For example, Title VII employment discrimination cases frequently require analyses of numerous characteristics of each class member's damages claim and involve too many variables (*e.g.*, race, sex, education, job classification, seniority, etc.) to permit a reliable estimate of claim variability.

On the other hand, it may be possible to compute reliable claim variability estimates even in cases where claim variability depends on more than one characteristic of class members' claims. In circumstances similar to those in *Tyson*, "donning" and "doffing" times of the sampled workers could be used to compute damages awards, and to estimate claim variability in terms of damages awards. But, *Tyson* is also an example of a case where claim variability depends on two variables. Specifically, if we now eliminate the assumption that all class members had valid claims and were entitled to damages, then damages, in a sense, depend on two variables: "donning" and "doffing" times and whether (for each additional incremental unit of time) a class member worked more than forty hours in the week.²⁷⁰ Claim variability can be estimated directly from the sample by aggregating these variables and considering the total amount of "donning" and "doffing" times exceeding forty hours of work each week.²⁷¹

²⁶⁹ We are still assuming that all class members have valid claims and are entitled to damages.

²⁷⁰ See *supra* notes 25–45 and accompanying text.

²⁷¹ Note, to compute damages, the times in excess of forty hours in a week could be multiplied by the statutorily-prescribed overtime rate (*i.e.*, 150% of hourly wages). One might think that, if it is possible to calculate damages based on a statute or other prescribed formula, judgment variability must be low or zero. This is not necessarily so. In *Tyson*, the jury awarded classwide damages of about \$2.9 million, notwithstanding plaintiffs' expert's formulaic calculations, based on the governing statute, supporting aggregate damages of approximately \$6.7 million. 136 S. Ct. at 1044. This disparity can likely be attributed to a number of factors. The point is, however, that multiple independent adjudications could result in variable outcomes.

By extension, it may be possible to apply standard statistical methods to compute a reliable estimate even in cases where claim variability depends on multiple characteristics of class members' claims.

K. Judgment Variability and "Test Cases"

Judgment variability, on the other hand, cannot be estimated from the sample directly. Although it may be theoretically possible to isolate and quantify the characteristics of class members' claims that would affect individual damages, judgment variability measures the dispersion of the individual damages awards, given the particular facts of each individual claim. Judgment variability reflects the variability of damages awards that would be observed by adjudicating a single claim multiple times, or the variability associated with the damages awards that is not attributable to substantive differences among the individual claims (*i.e.*, that is not attributable to claim variability). Therefore, an obvious approach to estimating judgment variability contemplates the use of "test cases," and specifically, replication.

For example, if a single claim is adjudicated multiple times, the resulting damages awards can be used to estimate judgment variability for that claim and other substantively similar claims.²⁷² Alternatively, if claim variability can be estimated, then a sample of claims can be adjudicated and claim variability subtracted from the total variability to estimate judgment variability. Thus, a robust method for estimating judgment variability is through the adjudication of "test cases," where a representative sample of class members' claims are adjudicated either by trials or alternative dispute resolution (ADR) procedures.

If judgment variability is unlikely to change significantly from claim to claim (which, as suggested, will often occur), then judgment variability can be determined relatively efficiently by adjudicating one "representative" claim before numerous triers of fact. Each trier of fact can observe the proceedings in a separate observation room, or together in a single room but then deliberate separately. In circumstances similar to those in *Tyson*, a "representative" claim based on average "donning" and "doffing" times can be selected for adjudication (perhaps following streamlined procedures presided over by magistrate judges or court-appointed special masters) before numerous triers of fact—say, five to ten separate juries of six. Each trier of

²⁷² It is not necessary that the claims be identical. It is likely that judgment variability is relatively similar across substantively similar claims, such as those in a class or subclass. For example, it is unlikely that the judgment variability associated with a claim in *Tyson* based on "donning" and "doffing" times of eighteen minutes is significantly different from that associated with a claim based on "donning" and "doffing" times of twenty-two minutes.

fact can arrive independently at a damages award, and the variability of the awards can be used to estimate judgment variability.

This “test case” procedure captures some, but not all, sources of judgment variability. Adopting the language of the *Reference Guide on Statistics*, this procedure focuses on “between-observer variability” rather than “within-observer variability,” because the claim is adjudicated only once by each “observer,” that is, each trier of fact.²⁷³ This procedure may capture certain aspects of “within-observer variability,” but, in any event, omits certain significant sources of judgment variability, such as variations in the presentations of evidence. Notwithstanding this shortcoming, the procedure is relatively inexpensive and convenient. Furthermore, to the extent that it does not capture certain sources of judgment variability, the calculation of judgment variability will be, if anything, conservative—that is, it will understate judgment variability.

Of course, the “test case” procedure can be modified to suit the circumstances of each case. A more accurate, but more costly, method to determine judgment variability is to adjudicate a selected claim multiple times before multiple triers of fact. In certain cases, it may be beneficial to adopt the “test case” procedure for subclasses where subclassification is capable of neutralizing a major source of claim variability and thereby stabilizing, or making relatively consistent, judgment variability within each subclass.²⁷⁴

A more robust procedure involves selecting a “representative” sample of claims for adjudications. Each sampled claim can be adjudicated multiple times (again, perhaps following streamlined procedures presided over by magistrate judges or court-appointed special masters) to determine judgment variability for each claim. These determinations can then be “aggregated” (e.g., by averaging) to estimate judgment variability for the claims in the entire class (or subclass). This procedure is costly, but it captures various sources of judgment variability and is more robust to variations in judgment variability across class claims.

In another alternative, each claim in a “representative” sample of claims can be adjudicated once, the variability of the resulting damages awards can be computed, and the claim variability subtracted from the result. This procedure may be less costly than adjudicating each claim multiple times, but it captures fewer sources of judgment variability and relies on the accuracy of the claim variability computation.

In any event, if class plaintiffs wish to present expert testimony regarding a correlation analysis similar to that described earlier, they may

²⁷³ See *Reference Guide on Statistics*, *supra* note 32, at 228.

²⁷⁴ See *supra* note 272 and accompanying text.

propose a procedure to adjudicate multiple claims multiple times.²⁷⁵ To determine whether judgment variability exceeds claim variability in a case similar to *Tyson*, a federal district court can adjudicate “representative” claims at different “donning” and “doffing” times, each multiple times. For example, the court can try a “typical” claim involving five minutes “donning” and “doffing” times, ten minutes “donning” and “doffing” times, fifteen minutes “donning” and “doffing” times, and so on, each adjudication using multiple (e.g., four to six) juries of six. This way, the resulting damages awards would capture both claim variability (reflecting different times) and judgment variability (reflecting different outcomes by different triers of fact); and plaintiffs can use the resulting data to test hypotheses or derive confidence intervals related to the relevant correlation coefficients. In Figure 2, sample “donning” and “doffing” times can be used as estimates of “true” “donning” and “doffing” times for the independent variable on the x-axis and damages awards can be used as the dependent variable on the y-axis, thus allowing the possibility of correlation-based arguments.

The basic point is that “test cases,” and specifically replication, can be used (with varying levels of costs) to arrive at reliable estimates of judgment variability and claim variability, as well as other values, such as relevant correlation coefficients. The procedures are not mutually exclusive and can be tailored to the circumstances of each class action.

Furthermore, a federal district court can reduce the costs of determining damages awards for sampled class members by lowering the required judgment variability threshold so that relatively inexpensive methods (such as adjudicating a single claim before numerous triers of fact) can be used notwithstanding less accurate, but conservative (if anything), results. Under certain circumstances, a court can subclassify, or “stratify,” a class to neutralize, or “control for,” sources of claim variability, and then sample—and eventually award individual damages—within each subclass.²⁷⁶ These procedures can be employed to obtain relatively (or at least moderately) homogeneous subclasses, so that detection of a relatively low level of judgment variability would be sufficient to confidently conclude that judgment variability exceeds claim variability within a subclass, notwithstanding a less costly estimation procedure.

Additionally, in some cases, even where class members’ damages claims are heterogeneous, a federal district court, acting as the finder of fact, could find that certain variability is immaterial, rendering the claims less variable (or even homogeneous). In circumstances similar to those in *Tyson*, for example, a court can find that a large number of workers in the “cut and retrim” departments had “donning” and “doffing” times of approximately twenty

²⁷⁵ See *supra* notes 263–267 and accompanying text.

²⁷⁶ See FED. R. CIV. P. 23(c)(5). See generally *supra* notes 60–64 and accompanying text.

minutes, and that workers who spent more time “donning” and “doffing” safety gear were not deserving of additional compensation. In such cases—and particularly in cases in which the court’s findings reduce claim variability significantly—less costly estimation procedures are even more feasible.

CONCLUSION

The Supreme Court decision in *Tyson* revitalized the role of sampling in class actions. The *Tyson* Court analyzed sampling as an evidentiary issue and unanimously ruled that sampling evidence is a permissible method to prove classwide claims if it is shown “that each class member could have relied on” the evidence in an individual action.²⁷⁷ To paraphrase the *Tyson* Court, when class members are “similarly situated,” there is a “role for representative evidence,” in both individual and class adjudications.²⁷⁸ When class members are “similarly situated,” “experiences of a subset of [class members] can be probative as to the experiences of all of them.”²⁷⁹ And, in such cases, a “just and reasonable” inference can be drawn from the sampling evidence to prove a claim of each class member.²⁸⁰

After *Tyson*, federal courts have understandably been more receptive to plaintiffs’ efforts to introduce sampling evidence to prove classwide liability, and occasionally aggregate damages, in Rule 23(b)(3) class actions. This is an important development because sampling is often the only practicable method to prove classwide liability and damages and still satisfy the predominance prerequisite of the Rule. Yet, federal courts continue to refuse to certify individual damages issues because they find that individual issues, especially the calculations of individual damages, defeat the predominance of common issues.

In this Article, after examining a brief bleak history of sampling evidence in federal court class actions, we analyzed the law governing the admissibility of expert evidence and the certification of Rule 23(b)(3) damages class actions. As we explained, the admissibility of sampling evidence depends on its relevance and, most importantly, its reliability. We demonstrated how sampling evidence can satisfy *Daubert* and Rule 702 standards and the prerequisites for Rule 23(b)(3) classes.

While being careful to comply with these standards and prerequisites, we developed a method and derived a threshold to determine whether class damages claims are sufficiently homogeneous to justify the admissibility of sampling evidence to prove individual damages.

²⁷⁷ 136 S. Ct. at 1046–47; see *supra* notes 10–82 and accompanying text.

²⁷⁸ 136 S. Ct. at 1048.

²⁷⁹ *Id.*

²⁸⁰ *Id.* (quoting *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687 (1946)).

We argued that accuracy is an appropriate standard for evidentiary reliability. We then introduced two defined terms, judgment variability and claim variability, and, employing generally accepted statistical methods, we demonstrated that, when judgment variability exceeds claim variability, sampling evidence improves accuracy and evidentiary reliability, and therefore is admissible to prove individual damages. Finally, we recommended several procedures to evaluate whether damages claims of a putative class satisfy the derived threshold.

In conclusion, we argue that our proposed method to prove individual damages achieves the goals of Rule 23(b)(3) class actions, as stated by the Supreme Court. Federal courts and parties that follow the proposed method will realize “economies of time, effort, and expense.”²⁸¹ And, the proposed method promotes “uniformity of decision as to persons similarly situated, without sacrificing procedural fairness or bringing about other undesirable results.”²⁸²

²⁸¹ *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 615 (1997) (quoting FED. R. CIV. P. 23 advisory committee’s notes to 1966 amendment).

²⁸² *Id.*