Open Access

Hillel J. Bavli* and Yang Chen

Shrinkage Estimation in the Adjudication of Civil Damage Claims

DOI 10.1515/rle-2015-0010

Abstract: Recent papers have highlighted the use of claim aggregation as a tool for reducing the unpredictability of legal outcomes. Specifically, it has been argued that sampling methods can be used in the class action context, and comparable-case guidance – information regarding awards in comparable cases as guidance for determining damage awards – can be used in the individual-claim context, to reduce variability and improve the accuracy of awards. In this paper, we examine a third form of claim aggregation based on a statistical method called “shrinkage estimation,” which is used to aggregate information and thereby improve estimation. We examine the conditions under which “shrinkage” can improve the accuracy of damage awards, and we apply it to gain a deeper understanding of the benefits and limitations of claim aggregation in the sampling and comparable-case guidance contexts with respect to accuracy.

Keywords: accuracy, judgment variability, class action, claim aggregation, shrinkage estimation, sampling, pain and suffering, punitive damages, comparable-case guidance, jury awards, damages

1 Introduction

A legal proceeding can be understood as a procedure for generating an outcome that serves as an estimate of the “correct” outcome associated with a legal claim. In this sense, a criterion for measuring the strength of a legal procedure is the degree to which the procedure can be expected to generate “accurate” outcomes, outcomes that are close in proximity to the “correct” outcome (Bavli 2015, 2016). In two recent articles written by one of the authors, it is argued that certain claim aggregation methods, methods in

***Corresponding author: Hillel J. Bavli**, Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA, E-mail: hbavli@fas.harvard.edu

Yang Chen, Department of Statistics, Harvard University, Cambridge, MA, USA

which the outcome of a claim is based not only on the characteristics of the claim itself, but also on the outcomes of other claims, can improve the accuracy of claim outcomes. The first article examines the conditions under which sampling procedures can improve accuracy in the class action context (Bavli 2015), while the second article examines the use of comparable-case guidance (CCG), or “prior-award information” – information regarding awards in comparable cases as guidance for determining damage awards – to improve accuracy in the individual-claim context (Bavli 2017). Although sampling and comparable-case guidance are distinct in practice, and arise in different contexts, the underlying mechanisms by which they affect accuracy are similar.

Sampling procedures involve adjudicating a proportion of claims (the claims in the “sample group”) in a class action and extrapolating damage awards for the remaining claims (the claims in the “extrapolation group”). CCG methods involve incorporating information regarding awards in prior comparable cases in the adjudication of a damages award in a present case. Sampling allows for the sharing of information across claims in a class, whereas CCG allows for the sharing of information across individual claims. But both methods aggregate and use information regarding awards in comparable claims to influence awards of other claims.

In this article, we examine a third but closely related – and, in a sense, unifying – form of claim aggregation that integrates such influence explicitly. This form of claim aggregation is based on a statistical method called “shrinkage estimation” (or “shrinkage”), which is used to aggregate information and thereby improve estimation. Specifically, shrinkage involves adjusting an estimate of some value to account for information derived from the population of units from which that value is drawn (Casella 1985; Efron & Morris 1975; James & Stein 1961). CCG, which uses information regarding comparable claims to influence the subject claim, can be understood as a form of shrinkage. Similarly, sampling constitutes a special case of shrinkage where the population of units is the class of claims, and awards are “adjusted” to account for information derived from the population of claims either entirely or not at all, depending on whether a claim is in the extrapolation group or the sample group, respectively.

Our objectives in this article are to examine the conditions under which shrinkage can increase the accuracy of damage awards in the class action and individual-claim contexts, and to apply shrinkage to gain a deeper understanding of the benefits and limitations of the foregoing methods with respect to accuracy.

We begin in Section 2 by reviewing the sampling framework developed in Bavli (2015) (hereinafter “*Aggregating for Accuracy*”). In Section 3, we build on this framework to examine the benefits of shrinkage in the class action context, and to reexamine the benefits of sampling in light of shrinkage. We consider alternative methodologies under various assumptions regarding cost and legal constraints. In Section 4, we examine conditions under which shrinkage can be used to increase the accuracy of damage awards in the individual-claim context. In particular, we consider shrinkage in the CCG context, and we derive and illustrate the conditions under which CCG improves accuracy. In Section 5, we conclude.

2 A framework for examining sampling and accuracy in class action litigation

In this section, we summarize the framework developed in *Aggregating for Accuracy* and a number of central results related to the use of sampling to improve accuracy in class action litigation. We begin by discussing sampling in a class of homogeneous claims and then extend our discussion to classes of heterogeneous claims.

2.1 Sampling in a class of homogeneous claims

Aggregating for Accuracy builds on previous literature to develop a framework for examining the effect of sampling on accuracy in class action litigation. The article examines a procedure by which 1) a number of claims are sampled from a class of claims for individualized adjudication (and individualized damage awards), and 2) the mean of the awards adjudicated in the sample group is applied as the award for all remaining claims, the claims in the extrapolation group.

The article’s analysis is intended to respond to arguments that such procedures increase efficiency (by allowing putative class members to proceed as a class rather than as individual claimants), but only at the cost of reducing accuracy. It builds on assertions by Professors Michael Saks and Peter Blanck in a 1992 Stanford Law Review article to argue that, under certain conditions, sampling can increase accuracy by reducing error associated with *judgment variability* – that is, uncertainty in the adjudication of an award resulting, for example, from variability in the composition of a jury, the presentation of evidence, and the selection of a judge. (Bavli 2015; Saks & Blanck 1992).

To illustrate, consider how *replication* may be used to reduce judgment variability:

[I]magine now a (costly) hypothetical procedure in which each and every claim [in a class] were litigated ten times independently, and in which the outcome associated with each claim were computed by taking the average of the ten verdicts associated with that claim. That is, start by taking the first claim and litigate it before ten independent juries to obtain ten independent verdicts. Then assign the average of the ten verdicts as the outcome of the first case. By applying this aggregated outcome, rather than any single verdict, we may reduce the error resulting from judgment variability to nearly nothing (Bavli 2015).

Aggregating for Accuracy shows that sampling in a class of homogeneous claims, through its use of replication, can improve accuracy by reducing judgment variability (Bavli 2015; Saks & Blanck 1992). In particular, the article concludes that, given a class of N homogeneous claims, and legal restrictions that can be described by “reductive sampling” – where a court will not *replace* an individually adjudicated award with an award extrapolated from other claims – accuracy is maximized by randomly selecting a sample of $n^* = \sqrt{N}$ claims for individualized adjudication, assigning the individualized awards as the outcomes of the claims in the sample group, respectively, and then applying the sample mean of the sample-group awards as the outcome of all remaining $(N - \sqrt{N})$ claims in the class (Bavli 2015).

To be more precise: Assume we have a class of N homogeneous claims from which we sample n claims. Let us define a “correct” outcome associated with a particular claim as the average award that would emerge from repeated adjudication of the claim under different conditions, such as different jury combinations, different lawyers, different judges, and different presentations of evidence (Bavli 2015; Saks & Blanck 1992). The “correct” outcome can be defined using various measures of central tendency, such as the mean or median. Throughout this paper, we adopt the measure used in *Aggregating for Accuracy* and Saks & Blanck (1992) – the mean. Thus, let μ be the correct award in each of the N homogeneous claims – the mean of the awards that would result from repeated adjudications of any (or all, since the claims are homogeneous) of the N claims. And let X_i be a random variable defined by the actual award in the i^{th} claim, for $i = 1, 2, 3, \dots, N$, where the X_i are independent and identically distributed (i.i.d.) with mean μ and variance σ^2 , and the sample mean of the X_i , for $i = 1, 2, 3, \dots, n$, is defined as $\bar{X}_n$, which is distributed with mean μ and variance $\frac{\sigma^2}{n}$. Notationally: $X_i \sim (\mu, \sigma^2)$ and $\bar{X}_n \sim (\mu, \frac{\sigma^2}{n})$. Note, in *Aggregating for Accuracy*, it is assumed that the X_i are distributed normally; but this distributional assumption is not necessary for the results derived in that paper or in the current paper. We therefore drop the normality assumption.

Thus, using the sum of square residuals

$$r = \sum_{i=1}^n (X_i - \mu)^2 + (N - n)(\bar{X}_n - \mu)^2$$

as the criterion for measuring the error associated with all N claims, *Aggregating for Accuracy* concludes that total error, or “risk” (R) – defined as the expectation of r above – is minimized, and accuracy is maximized, *not* by adjudicating each claim individually (a procedure often viewed by courts and scholars as the *ideal*, with respect to accuracy), but rather by sampling and individually adjudicating

$$n_{hom}^* = \sqrt{N}$$

claims, and applying the sample mean $\bar{X}_n$ as the outcome for all remaining $N - \sqrt{N}$ claims (Bavli 2015).

Thus, in the context of a homogeneous class of claims, sampling may improve accuracy as well as efficiency. But, as explained in *Aggregating for Accuracy*, homogeneity is not necessary for sampling to increase accuracy. First, a court may stratify a heterogeneous class to obtain relatively homogeneous subclasses. For example, the District Court for the Eastern District of Texas, in *Cimino v. Raymark*, 751 F. Supp. 649 (E.D. Tex. 1990), used such a procedure when it divided a heterogeneous class of asbestos claims into five disease categories. Second, although homogeneity is helpful, it is not necessary – the error-reducing benefits of sampling apply even to a class of heterogeneous claims.

2.2 Sampling in a class of heterogeneous claims

Homogeneity allows a court to maximize accuracy by sampling $\sqrt{N}$ claims. However, as a class becomes more heterogeneous, the utility of sampling is reduced, since the benefits of reducing judgment variability must now be balanced with the error introduced by applying a single point estimate – the sample mean – to a class of heterogeneous claims. Thus, for a class of heterogeneous claims, the optimal sample size will fall between $\sqrt{N}$ and N (Bavli 2015). Simply stated, sampling can increase accuracy as long as the heterogeneity of the claims is not too large.

Aggregating for Accuracy models heterogeneous awards as draws from normal distributions with means μ_i and variance σ^2 . Contrary to a homogeneous class, where all claims have the same correct award μ , in a heterogeneous class, each claim i has a correct award μ_i , where the *correct* awards are distributed with mean μ_0 and variance τ^2 . Thus, $X_i \sim (\mu_i, \sigma^2)$, independent, where $\mu_i \sim (\mu_0, \tau^2)$, i.i.d. Note, here again we relax the normality assumption used in *Aggregating for Accuracy*.

Thus, again using the sum of square residuals, we have

$$r = \sum_{i=1}^n (X_i - \mu_i)^2 + \sum_{j=n+1}^N (\bar{X}_n - \mu_j)^2$$

as the criterion for measuring the error associated with all N claims. *Aggregating for Accuracy* minimizes the expectation of this expression and derives the optimal sample size for heterogeneous claims to be

$$n^* = \sqrt{N} \sqrt{\frac{\sigma^2 + \tau^2}{\sigma^2 - \tau^2}}.$$

Thus, when τ^2 , or “claim variability” (i.e., the heterogeneity of the class), is dominated by σ^2 , or judgment variability, n^* falls between $\sqrt{N}$ and N . If claim variability is greater than judgment variability ($\tau^2 \geq \sigma^2$), then $n^* = N$. And when claim variability is zero ($\tau^2 = 0$), then $n^* = \sqrt{N}$, the result obtained for a homogeneous class (Bavli 2015).

3 Shrinkage estimation in the class action context

Determining an appropriate aggregation method, with respect to accuracy, depends on relevant legal and cost constraints. As mentioned, the framework and conclusions described above assume that a court may not replace an adjudicated award with an extrapolated award (an assumption referred to in *Aggregating for Accuracy* as “reductive sampling”). *Aggregating for Accuracy* argues that, while there is clear precedent for extrapolating awards for non-adjudicated claims, replacing individually adjudicated awards with extrapolated awards raises major constitutional, and other, problems. In the absence of this constraint, however, other aggregation methods may be more beneficial with respect to accuracy. For example, assuming no legal or cost constraints, a court may adjudicate *all* claims individually and then replace *all* individual awards with extrapolated awards, such as with the mean of the individual awards.

In the current section, we relax the legal constraints assumed in *Aggregating for Accuracy* in order to consider the effect of shrinkage – which involves replacing an adjudicated award with one that is influenced by awards in comparable claims – on accuracy in the class action context. Relaxing these constraints is useful for at least two reasons. First, there may be contexts in which such procedures are permissible. For example, parties may opt for them in settlement or alternative dispute resolution contexts. Second, examining the effects of shrinkage permits a more complete understanding of claim aggregation in light of relevant legal and cost constraints.

Thus, in the current section, we begin by showing that, for a class of claims, shrinkage can achieve greater accuracy than classical case-by-case adjudication. Our point of comparison is case-by-case adjudication (as it is in *Aggregating for Accuracy*), rather than a typical class action, because the former procedure is often viewed as the ideal with respect to accuracy, and is used as the primary alternative to class certification if putative class representatives are unable to show that class treatment is appropriate. We then apply shrinkage to reexamine the sampling results derived in *Aggregating for Accuracy*, and show that relaxing the reductive sampling constraint and applying shrinkage leads to greater accuracy than even the sampling method examined in that paper.

3.1 Comparison to individual adjudications

Our objective in this subsection is to show that, for a class of claims, replacing an adjudicated damages award with an award based on shrinkage increases accuracy, in expectation, for each adjudicated claim, and therefore, in the aggregate for all sampled claims.

As above, assume we have a class of N claims and that each claim i is associated with a correct award μ_i , for $i = 1, 2, 3, \dots, N$. Assume the μ_i 's are equal (i.e., the claims are homogeneous) or that they arise from a common distribution with mean μ_0 and variance τ^2 . Denote the awards of the n sampled claims by $X_1, \dots, X_n$, and assume that they are distributed around their correct awards μ_i , respectively, with variance σ^2 . Thus, $X_i \sim (\mu_i, \sigma^2)$, independent, where $\mu_i \sim (\mu_0, \tau^2)$, i.i.d., and σ^2 and τ^2 are known and represent judgment variability and claim variability, respectively. (See *Aggregating for Accuracy* for recommendations for estimating relevant parameters). As above, we do not rely on distributional assumptions.

Our objective is to impute all of the missing correct outcomes $\{\mu_i\}_{i=1}^N$, which are not directly observable, using estimated values $\{\hat{\mu}_i\}_{i=1}^N$, and to do so in a way that minimizes error, represented by the (standard) risk function,

$$R = \mathbb{E} \sum_{i=1}^N (\hat{\mu}_i - \mu_i)^2 = \mathbb{E} \sum_{i=1}^n (\hat{\mu}_i - \mu_i)^2 + \mathbb{E} \sum_{i=n+1}^N (\hat{\mu}_i - \mu_i)^2.$$

We thus replace the classical estimator – an adjudicated award (X_i) – with a shrinkage estimator ($\hat{\mu}_i^s$), which combines the adjudicated award with additional information obtained from the other claims in the class. We define the shrinkage estimator as:

$$\hat{\mu}_i^s = \frac{X_i/\sigma^2 + \mu_0/\tau^2}{1/\sigma^2 + 1/\tau^2}.$$

This estimator ($\hat{\mu}_i^s$) for the correct outcome in the i^{th} claim thus differs from the classical estimator, the adjudicated award (X_i), in that $\hat{\mu}_i^s$ is a weighted average of the adjudicated award (X_i) and the mean (μ_0) of the μ_i , weighted by the inverse of the variability of each, respectively. Thus, X_i and μ_0 are weighted by the inverse of the judgment variability (σ^2) and the inverse of the claim variability (τ^2), respectively. Intuitively, smaller variability implies greater information, and therefore heavier weight (Casella 1985).

Note, if we assume that the distributions of X_i and μ_i are Gaussian (or “Normal”) – i.e., if we regard $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ as the likelihood and $\mu_i \sim \mathcal{N}(\mu_0, \tau^2)$ as the prior – we indeed have an explicit statistical justification for this estimator. It is the Bayes estimator, which is admissible. It is also the maximum likelihood estimator in the hierarchical model.

It follows that the risk of the shrinkage estimator is

$$R_i^s := \mathbb{E}(\hat{\mu}_i^s - \mu_i)^2 = \mathbb{E}\left(\frac{X_i/\sigma^2 + \mu_0/\tau^2}{1/\sigma^2 + 1/\tau^2} - \mu_i\right)^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}, \quad [1]$$

which is smaller than $R_i^c := \mathbb{E}(X_i - \mu_i)^2 = \sigma^2$, the risk associated with the classical estimator X_i of μ_i .

Therefore, for each individual claim, using the shrinkage estimator to compute damages yields greater accuracy on average – that is, lower risk – as compared to an adjudicated award. Furthermore, this result implies that applying an individualized shrinkage award $\hat{\mu}_i^s$ to each claim in the class reduces total risk, since risk is reduced for each individual claim.

In general, we will not know the value of μ_0 . However, we can substitute the unbiased estimator $\hat{\mu}_0 = \sum_{i=1}^n X_i/n$ for μ_0 in the shrinkage estimator, yielding the “empirical” shrinkage estimator

$$\hat{\mu}_i^{se} = \frac{X_i/\sigma^2 + \hat{\mu}_0/\tau^2}{1/\sigma^2 + 1/\tau^2}, \quad [2]$$

where $\hat{\mu}_0 \sim (\mu_0, \frac{\sigma^2 + \tau^2}{n})$. Note that in the Gaussian case, where the X_i are independent with distribution $\mathcal{N}(\mu_i, \sigma^2)$, and μ_i has prior distribution $\mathcal{N}(\mu_0, \tau^2)$, this estimator is the “empirical Bayes estimator” (Efron & Morris 1973; Robbins 1955), which replaces the hyperparameter μ_0 with its maximum likelihood estimate. This empirical Bayes estimator, which converges to the Bayes estimator as $n \rightarrow \infty$, is asymptotically admissible. Therefore, statistically it is a justified estimator.

Thus, let us confirm that the risk associated with this estimator is less than the risk associated with the classical estimator, an adjudicated award. Letting $A = \frac{\sigma^{-2}}{\sigma^{-2} + \tau^{-2}}$, we can write this “empirical” shrinkage estimator $\hat{\mu}_i$ (note that, for simplicity, we drop the “se” in the superscript) as $\hat{\mu}_i = AX_i + (1-A)\hat{\mu}_0$. The risk of this estimator is

$$R_i^{se} := \mathbb{E} (\hat{\mu}_i - \mu_i)^2 = \mathbb{E} [A(X_i - \mu_i) + (1 - A)(\hat{\mu}_0 - \mu_i)]^2 \quad [3]$$

$$= A^2 \sigma^2 + 2A(1 - A) \mathbb{E}(X_i - \mu_i)(\hat{\mu}_0 - \mu_i) + (1 - A)^2 \mathbb{E}(\hat{\mu}_0 - \mu_i)^2 \quad [4]$$

$$= \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} + \frac{1}{n} \frac{\sigma^4}{\sigma^2 + \tau^2} \rightarrow R_i^s \quad (n \rightarrow \infty). \quad [5]$$

This risk is somewhat larger than R_i^s with $R_i^{se} - R_i^s = n^{-1} \sigma^4 / (\sigma^2 + \tau^2) > 0$. Intuitively, because we have less information – we do not know the true value of μ_0 – we have greater risk. However, as the sample size increases, this risk converges to that in the case in which we know the true μ_0 . That is, $R_i^{se} \rightarrow R_i^s$ as $n \rightarrow \infty$.

Most significantly, however, R_i^{se} , the risk associated with the empirical shrinkage estimator, is substantially smaller than R_i^c , the risk associated with the classical estimator (*i.e.*, the adjudicated award). That is, $R_i^{se} - R_i^c = (-1 + n^{-1}) \sigma^4 / (\sigma^2 + \tau^2) < 0$. Thus, under the specified criterion, the *empirical* shrinkage estimator is also (in addition to the shrinkage estimator) a better estimator than an adjudicated award.

Note that if it is not feasible to estimate claim variability (a possibility discussed in *Aggregating for Accuracy*), we can nevertheless rely on the James-Stein estimator (Efron & Morris 1975; James & Stein 1961), which does not depend on claim variability:

$$\hat{\mu}_i^{JS} = \hat{\mu}_0 + \left(1 - \frac{n-3}{S} \sigma^2 \right) (X_i - \hat{\mu}_0),$$

where $S = \sum_{i=1}^n (X_i - \hat{\mu}_0)^2$ is an unbiased estimator for $(n-1)(\sigma^2 + \tau^2)$. If X_i and μ_i are Gaussian, $(n-3)/S$ is an unbiased estimator for $(\sigma^2 + \tau^2)^{-1}$. The risk associated with the James-Stein estimator can be derived under Gaussian assumptions (Efron & Morris 1975); and it can be shown that this risk is also less than the risk associated with the classical estimator. Indeed, the James-Stein estimator converges to the shrinkage estimator as $n \rightarrow \infty$; and the risk associated with the James-Stein estimator and the risk associated with the shrinkage estimator above are asymptotically equal.

3.2 Sampling with shrinkage estimation

In this subsection, we reexamine the accuracy benefits of sampling, but now using the empirical shrinkage estimator ($\hat{\mu}_i$), rather than an adjudicated award (X_i), for the claims in the sample group. As before, the awards for the remaining claims (the claims in the extrapolation group) are extrapolated using the estimated global mean $\hat{\mu}_0$.

Let us begin by deriving the risk associated with the empirical shrinkage estimator in the sampling framework discussed above. The total risk for sampled and non-sampled claims is:

$$\begin{aligned}
 R_S &= R_{\text{sampled}} + R_{\text{non-sampled}} \\
 &= \sum_{i=1}^n \mathbb{E}(\hat{\mu}_i - \mu_i)^2 + \sum_{i=n+1}^N \mathbb{E}(\hat{\mu}_0 - \mu_i)^2 \\
 &= \sum_{i=1}^n R_i^{\text{se}} + (N-n) \left(\tau^2 + \frac{\sigma^2 + \tau^2}{n} \right) \\
 &= n \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} + \frac{\sigma^4}{\sigma^2 + \tau^2} + (N-n) \left[\frac{\sigma^2 + \tau^2}{n} + \tau^2 \right],
 \end{aligned}$$

where R_i^{se} (eq. [3]) is the risk associated with a single sampled claim i using the empirical shrinkage estimator (eq. [2]).

R_S is thus a monotone decreasing function of n :

$$\frac{\partial R_S}{\partial n} = -\frac{\tau^4}{\tau^2 + \sigma^2} - \frac{N}{n^2}(\sigma^2 + \tau^2) < 0. \quad [6]$$

This means that risk will continue to decrease as we increase the sample size n .

On the other hand, using the classical estimator, an adjudicated award X_i , to estimate μ_i results in the risk function,

$$\begin{aligned}
 R_C &= \mathbb{E} \sum_{i=1}^N (X_i - \mu_i)^2 = \mathbb{E} \sum_{i=1}^n (X_i - \mu_i)^2 + \mathbb{E} \sum_{j=n+1}^N (\hat{\mu}_0 - \mu_j)^2 \\
 &= n\sigma^2 + (N-n) \left(\frac{\sigma^2}{n} + \frac{n+1}{n} \tau^2 \right),
 \end{aligned}$$

which, as derived in *Aggregating for Accuracy* (and reviewed above), is minimized at

$$n^* = \sqrt{N \frac{\sigma^2 + \tau^2}{\sigma^2 - \tau^2}}. \quad [7]$$

Thus, Figure 1 illustrates the divergence of the risk of the empirical shrinkage estimator from the risk of the classical estimator. Specifically, Figure 1 plots the risk associated with a class of 1,000 claims against the number of claims, n , sampled for individual adjudication. The figure shows the risk R_C associated with the classical estimator (adjudicated awards) and the risk R_S associated with the shrinkage estimator for sample sizes between 10 and 200. As Figure 1 illustrates, R_C is minimized based on eq. [7] – where, here, n^* is approximately

Figure 1: Illustration of risk, plotted against the number of claims, n , sampled for adjudication, comparing the risk, R_C , associated with the classical estimator (adjudicated awards), and the risk, R_S , associated with the shrinkage estimator, given a class of 1,000 claims and sample sizes between 10 and 200.

33 (with $\sigma/\tau = 5$) – whereas R_S is monotonically decreasing in n . Therefore, if we relax the reductive sampling constraint in *Aggregating for Accuracy*, and instead assume the permissibility of shrinkage estimation, the sample size that minimizes risk is N , the total number of claims in the class. (Note, if claim variability is unknown, the court can first sample a small number of claims for adjudication and use $S = \sum_{i=1}^n (X_i - \hat{\mu}_0)^2$ to estimate $(n-1)(\sigma^2 + \tau^2)$ (Bavli 2015). In this case, $n^* = \sqrt{N} \left(\frac{2(n-1)\sigma^2}{S} - 1 \right)^{-\frac{1}{2}}$. We can then obtain the asymptotic interval for this new estimated number of claims using the central limit theorem and the delta method. We can therefore obtain a range for the optimal sample size and determine our choice based on some extraneous criteria.)

Importantly, the foregoing results should not be interpreted as supporting an argument against sampling; the accuracy benefits of sampling are substantial under the conditions, including the legal constraints, described in *Aggregating for Accuracy*. Rather, these results demonstrate the benefits of shrinkage. Indeed, shrinkage does not detract from the accuracy benefits of sampling; rather, it sufficiently enhances the accuracy of individual estimates – i.e., resulting from individualized adjudication and the replacement of individually adjudicated awards with shrinkage estimates – that, in a sense, it reduces the need for (i.e., the relative benefits of) sampling.

Furthermore, it is significant that in circumstances in which judgment variability dominates claim variability ($\sigma^2 > \tau^2$), shrinkage estimation reduces risk at a relatively high rate when $n < n^*$ and at a relatively low rate when $n > n^*$ (since, as n approaches N , the second term on the right side of eq. [6] approaches 0 and the first term dominates). Therefore, if the reductive sampling constraint is relaxed – and, in particular, if courts are willing to adjust individual adjudications to incorporate aggregate information – then, to balance concerns regarding accuracy with concerns regarding *litigation costs*, a court may apply a sampling-based method consistent with the results derived in *Aggregating for Accuracy*, but add the step of replacing the individually adjudicated awards with shrinkage estimates. That is, the court may sample $n^* = \sqrt{N \frac{\sigma^2 + \tau^2}{\sigma^2 - \tau^2}}$ claims for individualized adjudication; assign individualized *shrinkage estimates* to the n^* sample-group claims (based on their respective individualized awards as well as the mean of the sampled claims); and apply the mean of the awards in the sample group as the outcome of all non-sampled claims. This procedure is identical to the procedure derived in *Aggregating for Accuracy*, with one difference: here, the court would apply individualized *shrinkage estimates*, rather than individualized classical estimates (i.e., individually adjudicated awards), as the outcomes of the claims in the sample group.

In concluding this section, we note that we do not intend to make normative statements regarding the appropriate use of shrinkage in litigation. For example, it is beyond the scope of this paper to address the constitutionality of shrinkage or related policy concerns. Instead, we aim to develop a more complete understanding of aggregation, with respect to accuracy, and to examine a number of key results regarding the accuracy benefits of shrinkage.

In light of the results above, choosing an aggregation approach to maximize accuracy depends on the applicable legal and cost constraints. For example, if there is no concern for the reductive sampling constraint or cost constraints, repeated adjudications of each claim in a class would maximize accuracy. If constrained by litigation costs but not by reductive sampling, then following a method such as the method described above involving both sampling and shrinkage may maximize accuracy. If constrained by reductive sampling (whether or not there is also concern regarding litigation costs), then a sampling method without shrinkage, such as the procedure derived in *Aggregating for Accuracy*, would maximize accuracy.

In the following section, we extend our analysis to the individual-claim context, which, unlike the class action context, involves no predefined set of claims on which to base aggregation.

4 Shrinkage estimation in the individual-claim context

In Section 3, we examined the accuracy benefits of shrinkage estimation in the class action context. We showed that shrinkage may be beneficial even under conditions of high claim variability. The class action context provides a convenient starting point for examining the benefits of shrinkage and aggregation procedures generally, since we are *given* a population of claims (presumably with relatively low claim variability) over which to aggregate. However, a heterogeneous *class* of claims bound together by common facts or issues is not far different, for purposes of shrinkage, from a population of individual, but “comparable,” claims that are similarly bound together by common facts or issues. Therefore, in the current section, we extend our discussion of shrinkage to the individual-claim context.

For purposes of this section, there are two major challenges to applying shrinkage in the individual-claim context. First, as highlighted above, it is generally impermissible to replace an adjudicated damages award with an award extrapolated formulaically. Second, applying shrinkage in the individual-claim context first requires identifying a suitable set of prior comparable cases.

It is beyond the scope of this article to examine the legality of replacing an adjudicated award with a shrinkage award. Rather, we apply shrinkage to examine methods that aim to reduce the judgment variability of certain types of (particularly unpredictable) damage awards by informing a trier of fact of awards in prior comparable cases (Bavli 2017). In a recent article written by one of the authors, shrinkage estimation is used to explain the accuracy benefits of comparable-case guidance (CCG), and to address the primary challenges to CCG methods (Bavli 2017) (hereinafter “*The Logic of CCG*”). Providing a trier of fact with prior-award information, or CCG, may serve as an innovative way to use shrinkage to improve certain types of damage awards. After all, a trier of fact may choose (explicitly or implicitly) to incorporate prior-award information in its adjudication just as a shrinkage estimator would incorporate such information formulaically.

As described in *The Logic of CCG*, there is substantial evidence that providing jurors with prior-award information is effective in reducing judgment variability and influencing damage awards generally; but whether a juror incorporates prior-award information as a shrinkage estimator would (e.g., by weighting prior-award information in proportion to the inverse variability of such information) is currently being studied in a series of experiments.

In the current section, we address the second challenge – the problem of identifying a set of prior comparable cases – by assuming that a trier of fact incorporates prior-award information as a shrinkage estimator would, and examining the conditions under which prior-award information increases accuracy. Our aim is to answer the following question: assuming the trier of fact acts “rationally,” in the sense of incorporating prior-award information as a shrinkage estimator would, what choices of prior cases increase accuracy? For example, how “wrong” can a set of prior awards be before prior-award information *reduces* accuracy?

This concern is essential for determining policy surrounding CCG methods. Although we do not yet know whether triers of fact act as predicted, applying shrinkage explicitly in this context enables an understanding of the potential benefits, and some of the potential risks, associated with the use of CCG, including the robustness of such benefits to “incorrect” sets of prior awards.

4.1 Background: the use of comparable-case guidance to reduce the variability of damage awards

The problem addressed in *The Logic of CCG* is the unpredictability (*i.e.*, judgment variability) of awards for pain and suffering and punitive damages – two types of awards for which the jury receives very little guidance from the court. The Supreme Court and lower courts have repeatedly emphasized the importance of reducing the variability of such awards. *See, e.g., Exxon Shipping Co. v. Baker*, 554 U.S. 471 (2008). (Bavli 2017)

The Logic of CCG highlights problems associated with existing methods, such as additur and remittitur, tools used by courts to increase or decrease the amount of an award found to be inadequate or excessive. Although these tools can be useful, and can be used to incorporate prior-award information in various ways (Kadane 2009), and in conjunction with other methods, alone they address extreme awards only, rather than variability generally, and (in practice) they ordinarily address only excessive awards and not inadequate ones (Bavli 2017). Additionally, widespread use of such methods arguably replaces the discretion of the trier of fact with that of the court, raising constitutional and policy issues. Other methods, such as caps, arbitrarily draw cutoff points, leading to bias and perverse outcomes (Bavli 2017).

As mentioned above, there is empirical evidence that providing the trier of fact with information regarding awards in prior cases is effective in reducing variability. But such studies do not address the effect of prior-award information on *accuracy* – that is, bias and variability.

The Logic of CCG develops a framework for examining the benefits and limitations of prior-award information in terms of accuracy; and it addresses a number of major challenges to the use of prior-award information to reduce variability, including the possibility of using award information from an “incorrect” set of prior cases (Bavli 2017). In the current section, we apply shrinkage estimation to analyze the effect of prior-award information on accuracy, and to derive a number of important results regarding this latter challenge in particular.

4.2 Identifying prior cases

As a preliminary matter, it is important to realize that there is no “correct” or “incorrect” set of prior cases. As discussed in *The Logic of CCG*, the effect of prior-award information on accuracy depends on 1) the alignment of the mean of the correct awards in the prior cases with the correct award in the subject case (or, in practice, the alignment of the material facts and issues in the prior cases with those in the subject case); 2) the substantive breadth of the prior cases; and 3) the number of prior cases, or the “sample size.” For example, the alignment of material facts and issues (or, for short, the alignment of the prior cases or prior awards) affects the bias introduced by the prior awards. We would like for the average correct award in the prior cases to align with, or be equal to, the correct award in the subject case. The *breadth* of the prior awards (or cases) affects, for example, the influence of the prior-award information on the subject award; but a set of prior cases that contains only identical, or almost identical, material facts and issues may result in a sample size of one or two, or even zero, prior awards. Thus, in identifying a set of prior cases, a court must balance its interests in maintaining a reasonable sample size, a reasonable breadth, and cases that involve facts and issues that are relatively aligned with those in the subject case (Bavli 2017).

Consider the example of the Seventh Circuit case, *Jutzi-Johnson v. United States*, 263 F.3d 753 (7th Cir. 2001), described in *The Logic of CCG*. That case involved an award for pain and suffering arising from circumstances in which a jail inmate committed suicide by hanging, due to a failure of the jail to supervise him appropriately. A court considering prior awards (as Judge Posner did in *Jutzi-Johnson*) would decide, for example, whether to use only cases involving inmates who hung themselves, individuals who hung themselves from the general population, individuals who committed suicide from the general population, individuals who suffered from asphyxiation (e.g., drowning) from the general population, etc. See *Jutzi-Johnson*, 263 F.3d at 760-61. If a court were to restrict its consideration to cases involving inmates who hung themselves, it would potentially obtain a very poor

sample size; if the court were to use a wider breadth of cases, the prior awards would have less influence and a higher risk of introducing bias (Bavli 2017).

The important point, for purposes of the current analysis, is that there are tradeoffs among breadth, alignment, and sample size; and combinations of these factors correspond to various levels of bias and variance, and therefore accuracy.

Thus, consider an individual claim that receives award Y . Y is centered at the correct award μ_y with judgement variability σ_y^2 , which is assumed to be known. The correct award μ_y is centered at λ_0 with variance η_0^2 , and can be understood as a single award from a distribution representing a population of awards from comparable claims. The correct awards in the comparable claims, as well as the correct award in the subject claim, are distributed around a global mean λ_0 with variability η_0^2 . Thus,

$$Y \sim (\mu_y, \sigma_y^2), \mu_y \sim (\lambda_0, \eta_0^2).$$

To be clear, in statistical terms, by “comparable” claims or cases, we mean to suggest that their awards somehow arise from the same distribution.

Now, if we know the global mean (λ_0) and variability (η_0^2) associated with this population, the shrinkage estimator is

$$\hat{\mu}_y^s = \frac{Y/\sigma_y^2 + \lambda_0/\eta_0^2}{1/\sigma_y^2 + 1/\eta_0^2}.$$

Similar to eq. [1], we know that the risk of this estimator is $(\sigma^{-2} + \eta_0^{-2})^{-1}$, which is smaller than σ^2 , the risk associated with the classical estimator Y to estimate μ_y . Notice, however, that this shrinkage estimator requires knowledge of the values of λ_0 and η_0 . A more realistic scenario is one in which we do not know these values and instead need to estimate them based on prior awards. Assume that the set of prior cases identified involves awards $X_1, \dots, X_N$ centered at the correct awards $\mu_1, \dots, \mu_N$, respectively, with judgement variability σ^2 , which can be estimated, but for simplicity is assumed to be known. That is,

$$X_i \sim (\mu_i, \sigma^2), \mu_i \sim (\lambda_0, \eta_0^2); i = 1, \dots, N.$$

We can then use the following unbiased estimators to estimate λ_0 and η_0^2 :

$$\hat{\lambda}_0 = \frac{\sum_{i=1}^N X_i}{N}, \hat{\eta}_0^2 = \frac{\sum_{i=1}^N (X_i - \hat{\lambda}_0)^2}{N-1} - \sigma^2.$$

And we can use the “empirical” shrinkage estimator,

$$\hat{\mu}_y^{se} = \frac{Y/\sigma_y^2 + \hat{\lambda}_0/\hat{\eta}_0^2}{1/\sigma_y^2 + 1/\hat{\eta}_0^2},$$

which converges to the (non-empirical) shrinkage estimator as $N \rightarrow \infty$. As in Section 3, replacing λ_0 and η_0 with unbiased estimators introduces some

uncertainty, and therefore some risk. However, as $N \rightarrow \infty$, the risk converges to that of the (non-empirical) shrinkage estimator. Therefore, in terms of risk, we benefit from increasing the number of prior cases. If the set of prior cases is too small – in the sense that the risk resulting from the estimated values of λ_0 and η_0 dominates the benefits of the empirical shrinkage estimator – a court can increase sample size by expanding the substantive breadth of the prior cases.

In the following subsection, we discuss the breadth and alignment of prior cases. Although sample size is an important consideration, the accuracy benefits of shrinkage are fairly robust to sample size. Specifically, because the shrinkage estimator is influenced by sample size only through estimation of the hyperparameters (i.e., the mean and variance of the correct awards in the prior cases) and because these quantities can be estimated reasonably well with a small sample size, a sample of 5 to 15 often suffices. Furthermore, as discussed in detail below, the accuracy benefits of shrinkage are robust to misalignment; therefore, even if the hyperparameter estimation turns out to be relatively poor, shrinkage can generally be expected to improve accuracy. This is especially true in light of procedures courts use to prevent outlying awards, such as remittitur and appellate review. For these reasons, and for simplicity, although concern for sample size is implicit in our analysis, it does not play a central role in our examples and illustrations below.

4.3 Breadth and alignment of prior cases

We are interested in examining two considerations: breadth and alignment. Consider Judge Posner's opinion in *Jutzi-Johnson*. Judge Posner disagreed with the prior cases identified by both the plaintiff and the defendant. He explained:

The plaintiff cites three cases in which damages for pain and suffering ranging from \$600,000 to \$1 million were awarded, but in each one the pain and suffering continued for hours, not minutes. The defendant confined its search for comparable cases to other prison suicide cases, implying that prisoners experience pain and suffering differently from other persons, so that it makes more sense to compare Johnson's pain and suffering to that of a prisoner who suffered a toothache than to that of a free person who was strangled, and concluding absurdly that any award for pain and suffering in this case that exceeded \$5,000 would be excessive.

Jutzi-Johnson, 263 F.3d at 760. Judge Posner ultimately concluded that “[t]he parties should have looked at awards in other cases involving asphyxiation, for example cases of drowning, which are numerous.” *Id.*

In the language of the current section, Judge Posner disagreed with the alignment of the plaintiff's cases, implying that awards corresponding to cases involving hours, rather than minutes, of pain and suffering would be inappropriately high. He disagreed with the breadth (and the alignment) of the defendant's cases, suggesting that a set of cases involving the pain and suffering of inmates, rather than the general population, is too narrow, and that the defendant's focus on inmates led to alignment issues that resulted in "absurd" conclusions. Additionally, Judge Posner seems to suggest that the sets of cases identified by the parties suffered from small sample sizes as well, indicating that broadening the prior cases to include other cases involving asphyxiation in the general population would have led to "numerous" cases. *Id.* (Bavli 2017)

Thus, consider again a claim that receives an award Y "drawn from" a distribution centered at the correct award μ_y with judgement variability σ_y^2 (known), where μ_y is centered at λ_0 with variance η_0^2 . That is,

$$Y \sim (\mu_y, \sigma_y^2), \mu_y \sim (\lambda_0, \eta_0^2).$$

Assume that the court identifies a set of prior cases involving awards $X_1, \dots, X_N$ with correct outcomes $\mu_1, \dots, \mu_N$ centered at μ_0 with variance τ^2 . Thus,

$$X_i \sim (\mu_i, \sigma^2), \mu_i \sim (\mu_0, \tau^2); i = 1, \dots, N.$$

Essentially, this means $X_i \sim (\mu_0, \psi^2)$ for $1 \leq i \leq N$, where $\psi^2 = \sigma^2 + \tau^2$. Now, consider an estimator of the form

$$\hat{\mu}_y^s = \frac{\frac{Y}{\sigma_y^2} + \frac{\mu_0}{\psi^2}}{\frac{1}{\sigma_y^2} + \frac{1}{\psi^2}}, \quad [8]$$

which we can approximate by plugging in unbiased estimators

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} \text{ and } S = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$$

for the unknown parameters μ_0 and $\psi^2 = \sigma^2 + \tau^2$, respectively, to obtain:

$$\hat{\mu}_y^a = \frac{\frac{Y}{\sigma_y^2} + \frac{\bar{X}}{S}}{\frac{1}{\sigma_y^2} + \frac{1}{S}}. \quad [9]$$

Again, as $N \rightarrow \infty$, $\hat{\mu}_y^a \rightarrow \hat{\mu}_y^s$. And the risk of the shrinkage estimator $\hat{\mu}_y^s$ (which is roughly equal to the approximation, assuming a reasonable sample size) is

$$\begin{aligned}
 R_y^s &= \mathbb{E}(\hat{\mu}_y^s - \mu_y)^2 = \mathbb{E}\left(\frac{\frac{Y}{\sigma_y^2} + \frac{\mu_0}{\psi^2}}{\frac{1}{\sigma_y^2} + \frac{1}{\psi^2}} - \mu_y\right)^2 = \mathbb{E}\left(\frac{\frac{Y - \mu_y}{\sigma_y^2} + \frac{\mu_0 - \mu_y}{\psi^2}}{\frac{1}{\sigma_y^2} + \frac{1}{\psi^2}}\right)^2 \\
 &= \left(\frac{1}{\sigma_y^2} + \frac{1}{\psi^2}\right)^{-2} \left[\frac{1}{\sigma_y^2} + \frac{(\mu_0 - \lambda_0)^2 + \eta_0^2}{\psi^4}\right],
 \end{aligned}$$

which is smaller than σ_y^2 , the risk of the classical estimator, when

$$(\mu_0 - \lambda_0)^2 + \eta_0^2 < 2\psi^2 + \sigma_y^2. \quad [10]$$

Let us consider the meaning of this condition and then examine a number of numerical examples to gain a deeper understanding of the circumstances necessary to improve accuracy. On the right side of eq. [10], $\psi^2 = \sigma^2 + \tau^2$ is the total variance (the sum of claim variability and judgment variability) of the prior awards; and σ_y^2 is the judgment variability of the subject case. On the left side of eq. [10], $(\mu_0 - \lambda_0)^2$ is the square of the misalignment, the square difference between the expected correct award in the subject case and the mean of the correct awards in the prior cases; and η_0^2 can be understood as the claim variability of the hypothetical population to which the correct award in the subject case belongs. For simplicity, we can set $\eta_0^2 = 0$ and view the correct award in the subject case as its own population or as a realization of μ_y . Note that, although we employ this assumption throughout this section, our conclusions and illustrations herein are robust to reasonable alternatives and potentially substantial values of η such as τ . The condition is satisfied, for example, if (1) the judgment variability of the subject award is greater than the hypothetical claim variability (e.g., where $\eta_0^2 = 0$), and (2) the breadth of the prior awards is greater than the misalignment of the prior awards. Of course, the effects of one factor can be offset by the effects of the other. For example, the effects of extreme misalignment can be offset by the effects of extreme judgment variability.

Furthermore, in general, the greater the dispersion of the prior awards, the more “tolerance” there is for misalignment. On the other hand, higher prior-award concentration requires greater alignment. Thus, it may be beneficial for the breadth of the prior awards to reflect the court’s confidence in their alignment with respect to the subject award.

Let us consider an example based on data obtained from Saks et al. (1997), which tested the effects, with respect to variability, of providing mock jurors with certain information regarding prior awards. In one set of control conditions in which mock jurors were provided with a fact pattern (based on actual personal injury cases) involving a “high-severity injury,” a broken back, the

mean and standard deviation of the award amounts determined by participants were approximately \$3 million and \$4 million, respectively (Saks et al. 1997).

Note that these values are based on amounts determined by mock jurors rather than mock juries. Also, however, “[b]ecause the distribution for the raw dollar awards was highly variable and positively skewed, awards greater than two standard deviations above the mean were recoded to the amount at two standard deviations.” *Id.* The authors thereby limited the variability of the data.

Based on these data we construct Figure 2, which assumes a correct award (μ_y) of \$3 million and judgment variability (σ_y) of \$4 million. It is intuitive to imagine an approximately “normal” distribution with almost all awards falling between 0 and \$11 million (that is, the mean ± 2 standard deviations). We assume also that $\mu_y = \lambda_0 = 3$ million and $\eta_0 = 0$. Thus, the figure illustrates risk as a function of the mean of the prior awards (indicating alignment), and displays different curves corresponding to different levels of prior-award variability and a shaded horizontal line corresponding to the risk of the classical estimator (which is not dependent on the mean or variability of the prior awards). We can see, for example, that if the prior awards are centered at \$4 million with standard deviation equal to \$2 million, we reduce risk by 92% by using the shrinkage estimator

Figure 2: Comparison of the risk corresponding to the shrinkage estimator (black curves) and the risk corresponding to the classical estimator (gray horizontal line) plotted against the mean of the prior awards when the correct award is \$3 million (vertical black line), and assuming $\eta_0 = 0$ and the judgment variability of the subject case, σ_y , is equal to \$4 million. The black curves correspond to different values of prior-award variability, which ranges from \$0.2 million to \$3 million.

rather than the classical estimator; and we can expect the award to fall within the interval \$74,000 to \$5.26 million (that is, the mean ± 2 standard deviations), rather than \$0 to \$11 million. If the prior awards are centered at \$5 million with standard deviation equal to \$3 million, we reduce risk by 76.8% relative to the classical estimator; and we can expect the award to fall within the interval \$0 to \$6.85 million. Although not shown in the figure, it can be shown that, for prior award mean and standard deviation equal to \$8 million and \$4 million, respectively, we reduce risk by 36% relative to the classical estimator. Finally, if the distribution of prior awards has a mean and standard deviation equal to \$3 million (the correct award) and \$2 million, respectively, the variability (in terms of standard deviation) of the estimator is \$0.8 million rather than \$4 million, and we reduce risk by 96% relative to the classical estimator.

Note that, although Saks et al. (1997) used mock jurors rather than juries, our choice of judgment variability – an important factor for whether prior-award information causes accuracy to increase or decrease – is likely conservative, since our choice (\$4 million) reflects the methodology in that study whereby all award amounts above two standard deviations above the mean were reduced to the amount of two standard deviations above the mean. To be sure, however, let us illustrate an example in which we set judgment variability to *half* the standard deviation used above. Thus, Figure 3 assumes a correct award (μ_y) of \$3 million and judgment variability (σ_y) of \$2 million. In this example, if prior

Figure 3: Comparison of the risk corresponding to the shrinkage estimator (black curves) and the risk corresponding to the classical estimator (gray horizontal line) plotted against the mean of the prior awards when the correct award is \$3 million (vertical black line), and assuming $\eta_0 = 0$ and the judgment variability of the subject case, σ_y , is equal to \$2 million. The black curves correspond to different values of prior-award variability, which ranges from \$0.2 million to \$3 million.

awards are centered at \$4 million with a standard deviation of \$2 million, we reduce risk by 68.75 % relative to the classical estimator. If the distribution of prior awards is centered at \$1 million with a standard deviation of \$400,000, a distribution that is concentrated around a significantly incorrect award, we nevertheless reduce risk by 7.5 % relative to the classical estimator. Finally, if the distribution of prior awards has a mean and standard deviation equal to \$3 million (the correct award) and \$2 million, respectively, we reduce risk by 75 % relative to the classical estimator. Furthermore, if the variability in this final scenario were \$200,000, the risk associated with the shrinkage estimator is even smaller – representing a risk reduction of 99.99 % (corresponding to a reduction in standard deviation from \$4 million for the classical estimator to \$20,000 for the shrinkage estimator).

Lastly, we construct a third example using data from Bovbjerg et al. (1988), which examined real award data by severity of injury to analyze the variability of awards for pain and suffering. The data presented in this example are arguably conservative as well, since 1) the authors excluded the 5 % of award values farthest from the median; 2) the data include reported incidents of additur and remittitur; and 3) the data reflect the value of dollars in 1987. Thus, in Figure 4, we consider the example of severity level 7 (out of 9, representing severe, but not maximum-severity, injuries), with mean and standard deviation values of approximately \$2 million and

Figure 4: Comparison of the risk corresponding to the shrinkage estimator (black curves) and the risk corresponding to the classical estimator (gray horizontal line) plotted against the mean of the prior awards when the correct award is \$2 million (vertical black line), and assuming $\eta_0 = 0$ and the judgment variability of the subject case, σ_y , is equal to \$2 million. The black curves correspond to different values of prior-award variability, which ranges from \$0.2 million to \$4 million.

\$2 million. The graph again displays different curves corresponding to different levels of prior-award variability and a shaded horizontal line representing the risk associated with the classical estimator at \$4 million squared, which is equal to the standard deviation squared. First, we see that if prior awards are centered at \$2 million (assumed to be the correct award) with a standard deviation of \$500,000, we reduce risk by 99.65 % relative to the classical estimator. If the prior awards have a mean and standard deviation equal to \$500,000 and \$500,000, we reduce risk by 49.83 % – a milder reduction, due to an introduction of bias (but a reduction nevertheless). If the prior awards have a mean and standard deviation equal to \$500,000 and \$1.5 million, respectively, we reduce risk by 64 % – an improvement relative to the former scenario, due to the increase in breadth, which reduces the impact of the bias. If the prior awards have a mean and standard deviation equal to \$4.2 million and \$200,000, respectively, we *increase* risk by 18.63 %, since we have a tightly bound distribution centered at a significantly incorrect award value. On the other hand, if the prior awards have a mean and standard deviation equal to \$4.2 million and \$1.5 million, we reduce risk by 37.48 %, since, now, the introduction of bias is reduced due to high prior-award breadth, and the beneficial effect of the prior awards on award variability dominates.

Thus, using the derivations and illustrations above, we state the following conclusions regarding the alignment and breadth of prior awards:

1. Prior awards that are relatively aligned with the correct award in the subject case can lead to large accuracy benefits. These benefits are robust to changes in the alignment and breadth of the prior awards.
2. Prior awards that are misaligned – even significantly misaligned – but have relatively high breadth can lead to accuracy benefits, but benefits that are small relative to those that result from prior awards that are aligned and have lower breadth.
3. Increasing only the breadth of the prior awards (without affecting alignment) will generally not harm accuracy, but will reduce the influence of the prior awards, and therefore reduce their benefits with respect to accuracy. Considerations for determining an appropriate breadth include the sample size and the court's confidence in the alignment of the prior awards.
4. Prior awards that are significantly misaligned and have low breadth can lead to harmful effects on accuracy. However, such effects generally require the unusual circumstance of tightly bound prior awards that are significantly misaligned.

In short, under relatively mild conditions, the shrinkage estimator outperforms the classical estimator, an adjudicated award.

5 Conclusion

Claim aggregation may enable a court to improve the accuracy of damage awards by allowing for the sharing of information across claims. Recent papers have argued as such in the contexts of 1) sampling a proportion of claims in a class action for purposes of extrapolating awards for unsampled class claims; and 2) providing a trier of fact with prior-award information as guidance for determining awards for pain and suffering or punitive damages.

Our goal in this paper was to examine certain implications of a third, but closely related, form of claim aggregation called shrinkage estimation. We analyzed the accuracy benefits of shrinkage in the contexts of sampling and comparable-case guidance; and we applied it to gain a deeper understanding of the benefits and limitations of claim aggregation generally. We began our analysis by applying shrinkage in the class action context, and by building on the results obtained in *Aggregating for Accuracy*. We found that shrinkage leads to accuracy improvements relative to individual adjudications, and also relative to the sampling methods examined in *Aggregating for Accuracy*. But shrinkage also requires relaxing certain legal constraints that are applicable in many legal contexts. Indeed, the optimal aggregation method depends on legal and cost constraints.

We then extended our analysis to the individual-claim context, and applied shrinkage to gain a deeper understanding of the potential benefits and limitations of comparable-case guidance. Applying certain behavioral assumptions, we derived the precise conditions, in terms of the alignment and breadth of a set of prior awards, under which comparable-case guidance leads to an increase or decrease in accuracy. We then used our analysis to draw conclusions regarding the robustness of the accuracy benefits of comparable-case guidance to variations in the set of prior awards identified, and to illustrate them using a number of figures and examples.

Shrinkage is an important concept in statistics. Although it has (unsurprisingly) received little attention in law, it has many applications. By allowing for the sharing of information across claims, shrinkage has the potential to play an important role (explicitly or implicitly) to improve the accuracy of damage awards.

Acknowledgments: The views expressed in this paper are those of the authors, and not those of any organization with which they are affiliated. In writing this paper, the authors benefitted from the guidance of Professors Donald B. Rubin and Jun S. Liu.

References

- Bavli, H.J. 2015. "Aggregating for Accuracy: A Closer Look at Sampling and Accuracy in Class Action Litigation," 14(1) *Law, Probability and Risk* 67–90.
- Bavli, H.J. 2016. "Sampling and Reliability in Class Action Litigation," 2016 *Cardozo Law Review de novo* 207–219.
- Bavli, H.J. 2017. "The Logic of Comparable-Case Guidance in the Determination of Awards for Pain and Suffering and Punitive Damages," 85 *University of Cincinnati Law Review* 1–31.
- Bovbjerg, R.R., F.A. Sloan and J.F. Blumstein. 1989. "Valuing Life and Limb in Tort: Scheduling 'Pain and Suffering'," 83 *Northwestern University Law Review* 908–976.
- Casella, G. 1985. "An Introduction to Empirical Bayes Data Analysis," 39(2) *Journal of the American Statistical Association* 83–87.
- Efron, B. and C. Morris. 1973. "Stein's Estimation Rule and Its Competitors – an Empirical Bayes Approach," 68 *Journal of the American Statistical Association* 117–130.
- Efron, B. and C. Morris. 1975. "Data Analysis Using Stein's Estimator and Its Generalizations," 70 *Journal of the American Statistical Association* 311–319.
- James, W. and C. Stein. 1961. "Estimation with Quadratic Loss," 1 *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 361–379.
- Kadane, J.B. 2009. "Calculating Remittiturs," 8(2) *Law, Probability and Risk* 125–131.
- Robbins, H. 1956. "An Empirical Bayes Approach to Statistics," 1 *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 157–163.
- Saks, M.J. and P.D. Blanck. 1992. "Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts," 44 *Stanford Law Review* 815–851.
- Saks, M.J., L.A. Hollinger, R.L. Wissler, D.L. Evans and A.J. Hart. 1997. "Reducing Variability in Civil Jury Awards," 21(3) *Law and Human Behavior* 243–256.