TUESDAY, JULY 6, 2021

— PERSPECTIVE —

Using data sets to settle cases

**By Daniel Garrie
and Gail Andler**

Suppose you have a consumer class action involving thousands, or millions, of consumers who ingested a tainted supplement. How do you settle a case of that scope? Or a data breach incident that exposed the personal identifying information of thousands of consumers? How can you calculate unpaid wages for hundreds or thousands of employees without examining the payroll records of each individual employee? And if you are not able to settle these cases, how do you present proof at trial of these large data sets?

The discovery and use of large data sets for mediation and trial often take place through surveying, sampling and extrapolating. However, trial courts have been cautioned that statistical methods alone "cannot entirely substitute for common proof." California Judges Benchbook: Civil Proceedings Before Trial Section 11.29.

Extrapolating from existing data to produce new data is common in science and in law.

Extrapolation is the process by which information that is already known (the "sample") is used to predict the outcome for a larger group. That is, a sample of data is used to make inferences about the larger, general group. For such inferences to be properly drawn from the known facts of the subset to the larger relevant population, it is essential that the sample be statistically valid. This means that the underlying methodology must be designed to yield a representative result. However, this does not necessarily mean that the sample itself has to be representative.

The California Supreme Court discussed the concepts of sampling, extrapolation, and the need for sound methodology in *Duran v. U.S. Bank Nat'l Assn.*, 59 Cal. 4th 1 (2014). In Duran, the trial court was found to have improperly extrapolated the amount of overtime pay from a sample to the class as a whole, where the sample

was devised without expert input which allowed the parties to "impeach the model or otherwise show its liability is reduced." Although the Supreme Court recognized the appropriateness of the use of sampling and surveys for proof of liability of damages, it found problems with the methodology employed.

Subsequent cases at both the state and federal levels, including the U.S. Supreme Court, have addressed the issue of proof of liability through a representative sample. For example, in *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442 (2016), the court noted that "[w]hether a representative sample may be used to establish classwide liability will depend on the purpose for which the sample is being introduced and on the underlying cause of action." Specifically, Justice Anthony Kennedy wrote that [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."

In the discovery phase or informally pre-mediation, counsel will often negotiate criteria for the sample to be produced. Negotiated criteria could include, among others: (1) the time period to be reviewed; (2) the size of the sample to be used; (3) what questions are to be answered through the sample, etc.

During this process, counsel for plaintiffs should always bear in mind the need for the sample to be representative, otherwise, the court will not find that the inferences are justified. As stated in the Benchbook: "Several considerations determine whether a sample is sufficiently representative to fairly support inferences about the underlying population, including the variability in the population, whether the sample size is appropriate, whether the sample is random or infected by selection

bias, and whether the margin of error in the statistical analysis is reasonable."

In general, to ensure that a court will find extrapolations of a sample to be justified, counsel for the party relying on this information should take certain steps. First, the relevant population and relevant units of measure and analysis should be clearly defined. Next, there should be a clear procedure set forth to select the sample from the general population (e.g., probability sampling versus non-probability sampling). Additionally, the relevant characteristics for each element or unit should be measured using a reliable protocol for the measure. Finally, the elements examined in the sample should be reliable and relevant to the cause of action in the case at bar.

While bearing in mind the stringent standards for using extrapolation and statistical sampling at trial, for purposes of mediation counsel may wish to "start the information flowing" in order to preliminarily evaluate the settlement value of the case, subject to later con-

firmatory discovery as an aspect of presenting the class action settlement to the court for approval. Counsel should consider working with a consultant to determine the size and other criteria for a random sample to be requested of defense counsel for purposes of mediation. For some cases counsel may wish to consider "hits" on a particular website over a specified period of time; for other cases, the criteria may need to consider a sample per product distributor, employee worksite, geographic location, age of the consumer, etc. It is critical that the data set used for mediation be the same data set each side considered in evaluating potential value and exposure. It will also be helpful for any consultants used by counsel to be available by phone for the mediation.

There are a number of software products available to help counsel manage and navigate through large data sets. The complexities of these important cases demonstrate a perfect example of the value of partnering with other counsel. Have fun and settle. ■

Daniel B. Garrie *is the co-founder of Law & Forensics, a global forensic, e-discovery and cybersecurity engineering firm. He is also a neutral, discovery referee, arbitrator, technical special master and forensic neutral with JAMS.*

Gail A. Andler *joined JAMS as a fulltime neutral after more than 21 years on the Orange County Superior Court where she served from 2007-2017 on the Complex Civil Litigation Panel.*