Bigger Isn't Always Better: The Use of Statistical Samples

Legal practitioners involved in the dispute resolution process are increasingly confronted with ever larger sets of data and documents. What role can statistical sampling play in dealing with the challenges of making sense of this volume of material? What does the sampling process look like? Can this be done in a precise, pragmatic and proportionate way? How can samples be used to their full potential whilst avoiding the common pitfalls? The following key lessons can be learned:

— **Lesson 1: Always establish the purpose of a sample.** Agreeing the purpose up-front can mean greater clarity. Be wary about using a sample for a purpose for which it was not originally designed, as this can lead to unreliability.

— **Lesson 2: Look out for sample selection biases.** Biases are a persistent concern for statisticians, and can lead to inaccurate and unreliable estimates.

— **Lesson 3: Beware non-statistical samples**. Whilst 'convenience' samples are tempting to use, they are unlikely to be representative and cannot be relied upon for extrapolation.

— **Lesson 4: Bigger isn't always better.** When deciding the size of a sample, there is a trade-off between precision and confidence versus time and cost.

The Availability of Data: A Double-Edged Sword

The growing availability of large, detailed and complex sets of data and documents in disputes is a mixed blessing for legal practitioners. On the one hand, these datasets can be used to address complex questions of legal liability and compensatory damages, with assistance from experts using specialist tools and techniques. On the other hand, legal practitioners can easily find themselves overwhelmed by the sheer volume of material that requires review. This can result in them struggling to find the proverbial needle in the haystack, or being more likely to encounter unusual findings that are a result of chance rather than representative of the truth. There is a solution to how legal practitioners can deal with this problem in a precise, pragmatic and proportionate way – statistical sampling. It's not new, but it does work.

A sample is simply a subset of a population, used to investigate the population in circumstances where it is impractical or too costly to investigate directly. Statistical sampling is not a new technique – one of the earliest recorded uses of a sample was in the 1600's where the population of London was estimated using data on the number of burials per year in a sample of parishes.[1]

Statistical samples are a well-established, intuitive and versatile tool used in many different fields, and they have found a new lease of life in modern day commercial disputes, with large volumes of data and documents in evidence. Legal practitioners are increasingly turning to

FTI CONSULTING™

samples to help provide an effective and cost-efficient alternative to analysing all of the data. When these samples are properly designed, implemented and analysed, they can help to draw compelling conclusions about large volumes of data, with a high and precisely quantified level of confidence, within the tight time frames and cost constraints of the dispute resolution process. However, samples that are inappropriately designed, poorly implemented or incorrectly analysed can have the opposite effect – imprecise and unreliable evidence, misleading conclusions, and costly mistakes.

What do legal practitioners need to know about sampling? How can they use samples to their full potential, whilst avoiding the common pitfalls?

What Is a Sample?

A sample is defined as *"a selected subset of a population chosen by some process usually with the objective of investigating particular properties of the parent population."*[2] Samples are used in a wide range of contexts and purposes.

— To take the pulse of public opinion in the run-up to elections, polling organisations conduct regular surveys of samples of voters.[3]

— To understand consumer preferences and inform product development, businesses conduct research on samples of potential consumers.

— To ensure that products meet quality and safety standards, manufacturers subject samples of units coming off a production line to stringent testing.

— To inform conclusions as to whether the financial statements of a company are fairly presented, auditors routinely examine samples of transactions to identify the prevalence and extent of misstatements in the accounts.[4]

Samples are also increasingly being used across a broad range of commercial disputes, for determining legal liability and assessing compensatory damages.

— **In product liability claims in the electronics industry,** samples of allegedly defective products can be drawn for testing to assess whether the overall product line meets warranted standards, and if not, what proportion of the products are – or will be by a certain time – defective or in breach of warranty, and ought to be remedied.

— **In breach of contract disputes in the insurance industry,** samples of insurance claims can be audited to gauge whether a portfolio of claims has been assessed correctly and managed in line with the terms of the insurance, whether the settlement amounts agreed on those claims are appropriate and, if not, the quantum of any overpayment (or 'leakage' as it is known in the industry).

— **In intellectual property disputes around the value of patent portfolios**, where standardised technologies are covered by thousands of patents, and patent holders are required to license their technology on fair, reasonable and non-discriminatory ('FRAND') terms, patent portfolios are commonly analysed using a sampling approach to determine what proportion of patents claimed to be essential are in fact essential to the technology.

— **In pre-action fraud investigations**, parties and legal teams often consider a sampling approach to assess the extent of a suspected fraud, and the likely scale of losses so as to inform and substantiate their pre-action correspondence, and to evaluate the likely costs and benefits of formalising a claim.

There are many types of samples, and entire statistics textbooks devoted to the theory and practice of designing, implementing, analysing and extrapolating from these samples. However, the sampling process generally follows the following steps:[5]

— **Step One:** Define the relevant population, the unit of analysis, and the purpose of the exercise.

— **Step Two:** Identify the sampling frame.

— **Step Three:** Determine the sampling method, which is the process that will be used to choose units from the sampling frame.

— **Step Four:** Draw the sample, and measure the relevant characteristics of the units selected.

— **Step Five:** Conduct analysis of the sample, and extrapolate it to reach a conclusion about the broader population.

The first step is to define the relevant population, the unit of analysis, and the purpose of the exercise

For example, if you are interested in the voting preferences of the UK public, then registered voters comprise the 'relevant population', the voters are the 'unit of analysis', and the purpose might be to estimate the proportion of voters who will vote for a particular candidate, or favour a particular policy. In the context of a dispute, perhaps a product liability claim, the relevant population may be defined as all units of an allegedly defective product that were purchased by the claimant.

The second step is to identify the sampling frame

This is the list of units from which the sample can be selected in practice, and it may differ from the relevant population. For example, if the political poll is to be run using a social media survey, the 'sampling frame' will exclude some registered voters who do not use social media (so called 'under-coverage') and may also include some other social media users who are not registered voters ('over-coverage').[6] In a product liability claim, the sampling frame may be restricted to those units of the product that are still in use, as the claimant may already have discarded certain products that stopped working.

The third step is to determine the sampling method, which is the process that will be used to choose units from the sampling frame

There are many different sampling methods available, and the so-called 'simple random sample' method is the simplest and most widely used. In a simple random sample, each unit in the population has an equal chance of being selected. For example, in an insurance dispute, individual insurance claims could each be assigned a random number, and then the 100 smallest random numbers selected for the sample. The number of units to be included in the sample (the 'sample size') is an important consideration at this stage, and is usually the topic of much deliberation. Although larger samples are generally better from a statistical perspective, since they can allow more precise and confident conclusions to be drawn, they are also more costly and time consuming to obtain and analyse, especially when detailed and specialist work is required to examine or inspect each unit. There is therefore a trade-off between statistical precision and confidence on the one hand, and time and cost on the other.

The fourth step is to draw the sample, and measure the relevant characteristics of the units selected

For example, in an insurance dispute, the parties' legal teams or an independent insurance auditor may be instructed to pore through the documentation relating to each sampled claim, and determine whether it was handled correctly or not. In a product liability dispute, engineering experts may be instructed to examine and test each sampled product, to determine whether it was defective or not.

The fifth step is to conduct analysis of the sample, and extrapolate it to reach a conclusion about the broader population

For example, in an intellectual property dispute about the value of a 500-strong patent portfolio (i.e. too many to assess individually), you may draw a sample of 80 patents and find that only 20 of those patents (i.e. 25% of the sample) are in fact essential to the standardised technology in dispute. Under certain conditions, and depending on the design of the sample, this 25% finding can be extrapolated to the broader population of 500 patents, to estimate that 125 of those will in in fact be essential. Complex but well-established statistical formulae can also be used to quantify how precise this estimate is and how much confidence one can have in it, by reference to 'margins of error' and 'confidence intervals'.[7]

Statistical Samples in Practice: Four Lessons

This process can seem straightforward on the face of it, but complications can and do arise in practice. Four key lessons should be kept in mind.

Lesson One: Always Establish the Purpose of a Sample

The purpose of the sample is of paramount importance to its proper design and analysis. It should be established and documented early, as a matter of priority, and then considered at every stage of the sampling process. Legal practitioners faced with designing a new sample for the purpose of a dispute should ideally seek to agree the purpose of the sample between the parties and with the court or tribunal, then design it to meet this purpose.

Where there is already an existing sample, perhaps designed by one of the parties at an earlier date, it's essential to clarify what the *original* purpose of the sample was, how and why it was designed, and reach an objective and dispassionate view of its suitability for the current purpose. Sometimes, it may be necessary to start again, with a new sample.

We have seen the benefits of following this lesson, and the dangers of not. For example:

— In a recent UK High Court (Business and Properties Courts) litigation in the insurance industry, concerning a claim for damages in relation to allegedly substandard claims management services, the parties agreed the purpose of the sample up-front, and then jointly instructed us to design a sample that would be used by the court to determine both liability and damages. The parties' legal advisors recommended an early investment in expert advice, and avoided the additional time, cost and complications that might have arisen if the parties had instead sought to analyse all claims, designed their own separate samples in isolation, or worse, 'cherry picked' insurance claims that best supported their respective cases. The dispute was subsequently settled.

— Of course, disputes do not always settle early. In another recent dispute in the electronics industry, the parties initially worked together amicably to design and test multiple samples of an allegedly defective product, but relations subsequently soured and the samples were then put to use for forecasting product failure rates to substantiate a multi-million dollar claim for damages – a very different purpose to that for which the samples were first defined. In the arbitration proceedings that followed, the purpose and suitability of the samples was the subject of intense and expensive argument, with multiple rounds of expert reports and much airtime during the hearing. Whilst it is tempting to 'make do' with sample data that already exists, this can sometimes be a false economy.

— One final example comes from the published judgment in a recent case between an English County Council (Cumbria) and a highways maintenance and services company, Amey, held before the High Court (Technology and Construction Court).[8] Cumbria alleged that road patching work completed by Amey was defective, and sought to substantiate its claim for liability and damages using a sample of road patches. The court determined that the sample was not sufficiently reliable, in part because *"…the sample is being used for a purpose for which it was not originally designed, with no or insufficient attempt being made to address these difficulties, whether at the outset or during the later stages."*[9]

Lesson Two: Look Out for Sample Selection Biases

'Sample selection bias' occurs when the units that are selected for a sample are not representative of the target population,[10] leading to inaccurate and unreliable estimates of the characteristics of that population. Sample selection biases are a persistent concern for statisticians. They are tricky to prevent or detect, and can have serious consequences. An infamous example is from the 1936 US general election, when the Literary Digest magazine sent out over 10 million straw vote ballots, and used the responses to predict a 55% majority for Presidential candidate Alf Landon. The prediction was totally wrong: the election was in fact a landslide victory for Franklin D. Roosevelt, who won 61% of the vote (compared to only 37% by Landon). The poll failed because there were serious sample selection biases 'baked in' to its design.[11] First, the sample frame was biased, as the sample was drawn primarily from automobile registration lists and telephone books, which underrepresented the supposed core of Roosevelt's support (the poor). Second, the response rates were also much higher among Landon supporters than Roosevelt supporters, compounding this bias.

Selection biases are not unique to political polls – they can also plague commercial disputes. In *Amey v Cumbria*, the court found that the sampling frame was a tiny and unrepresentative portion of the relevant population, leading to deliberate and clear bias.[12] The court determined that because of these failings, it was not safe to extrapolate the sample, meaning the sample was not sufficiently reliable to substantiate the claimant's case on liability and damages.[13]

Lesson Three: Beware Non-Statistical Samples

Statistical samples involve randomly selecting units, and then using probability theory to evaluate the sample results, whereas non-statistical sample units use subjective judgment instead to select the units.

A financial fraud investigator may scrutinise a small number of transactions that they consider to be the most suspicious, based on their understanding of the size of the transaction, the description provided, the account numbers involved, their past experience and any "hunches" or personal unconscious biases they might have. Such non-statistical samples can be useful in general investigations, or when the purpose of the exercise is to uncover problems. However, their results can rarely be extrapolated reliably to the population, and it is not possible to calculate confidence intervals and margins of error. If the fraud investigator were to find that 50% of the selected transactions were fraudulent, they could not assume that half of all transactions on the account where fraudulent, since their sample is biased by design towards the more suspicious transactions.

The distinction between statistical and non-statistical samples is therefore very important for legal practitioners to bear in mind when designing and evaluating a sample.

— We were recently involved in a UK High Court (Commercial Court) litigation in the car insurance industry, in which the defendants were accused of misrepresenting information relating to a large number of individual car insurance claims, causing the claimants to incur additional costs for which they now sought compensation. Since it was not feasible to assess every single insurance claim in turn, the court instead ordered that the parties select a 'trial sample' of 200 insurance claims. The parties selected their claims in a non-statistical manner, with the claimants selecting those claims which in their subjective judgement demonstrated the gravest and largest misstatements, and the defendants did the opposite.

Whilst this 'trial sample' might have been sufficient for the courts' initial purposes, it was later deemed insufficient for the purpose of assessing any damages due, since the results could not be reliably extrapolated to all relevant claims.

— In *Amey v Cumbria*, Cumbria's statistical expert accepted that it did not have a statistical sample but sought to argue that it was still representative and therefore safe to extrapolate. The court did not accept these arguments and determined that Cumbria's reliance on the non-statistical sample was *"misplaced"*.[14] This example shows that whilst it is theoretically possible for a non-statistical sample to be representative, this cannot be assumed and is not straightforward to establish.

Lesson Four: Bigger Isn't Always Better

It is tempting to think that bigger samples are better – after all, they lead to more precise extrapolation and more confidence in the results, and leave the door open to more sophisticated analyses in the future. This can lead parties to seek out as large a sample as possible. However, there is invariably a trade-off to be made between statistical precision and confidence versus time and cost. Irrespective of the benefits, litigations and arbitrations operate to specific timetables, and costs must be considered. Additionally, the statistical benefits of a larger sample diminish as the sample grows larger.

To illustrate this, suppose we need to determine how many products in an order of 10,000 are defective and in breach of warranty. We decide to draw a simple random sample of 50 products for inspection, 25 of which are found to be defective (i.e. 50 %). Using statistical theory, we could extrapolate from this finding, with 95% confidence,[15] that the number of defective products in the entire order of 10,000 is between 3,600 and 6,400.[16]

The range of uncertainty here is quite wide because the initial sample used is quite small. If instead we had increased our initial sample by 50 products (bringing the total to 100), and again found half of the products in the sample to be defective, we would have be able to make a more precise statement – with 95% confidence, that the number of defective products in the entire order is between 4,000 and 6,000 products and we would have narrowed the range of uncertainty by 800 products. If a further 50 products were added to the sample, our estimates would be yet more precise, but the improvement itself would diminish. This time, the confidence interval would be only slightly narrower, being 4,200 to 5,800 products. Clearly, there will come a point at which the benefits of having a larger sample no longer outweigh the costs of collecting and processing it. Finding this 'optimal point' requires an understanding of statistics, commercial reality, dispute resolution processes, and, in some cases, a more creative and sophisticated approach to sampling.

Looking to a real world example, we recently assisted a client operating in the water distribution industry, to conduct a pre-claim investigation into the extent to which the client had been defrauded by customers systematically underreporting their true water usage and underpaying their water bills. Due to the geographical spread of the customers, it would have been prohibitively expensive to draw a simple random sample to provide the level of confidence and precision the client desired – put simply, it would have taken months to drive across the country to sample readings from randomly chosen addresses. Instead, we developed a more complex sample design using 'clustering' and 'stratification' to take into account the geography of the country and the types of customers, whilst still producing a sample that met the purpose.

Final Thoughts

Whilst the availability of large sets of data and documents can be a double-edged sword for legal practitioners involved in the dispute resolution process, statistical samples are a well-established, intuitive and versatile tool that can be used in a precise, pragmatic and proportionate way. Though sample design and analysis is deceptively simple, it can sometimes be quite unintuitive. Ultimately, statistical sampling done well can be very effective, but when done badly and it can lead to misleading conclusions. Understanding the key steps in the sampling process can be helpful, and seeking expert advice and input at an early stage can be vital.

MELORIA MESCHI
Senior Managing Director
+44 20 3727 1362
meloria.meschi@fticonsulting.com

RAVI KANABAR
Managing Director
+44 20 3727 1280
ravi.kanabar@fticonsulting.com

FTI CONSULTING

.3002-1124

Endnotes

1   John Graunt's calculation uses data from various other sources too, but is based primarily on extrapolating from a sample. See: Anders Hald, *'History of Probability and Statistics and Their Applications before 1750'* (1990), at pages 81-105, https://onlinelibrary.wiley.com/doi/book/10.1002/0471725161.

2   B. S. Everitt and A. Skrondal, *'The Cambridge Dictionary of Statistics'* (2010).

3   See for example, The British Polling Council, About the BPC, https://www.britishpollingcouncil.org/.

4   For example, the Financial Reporting Council, the UK regulator for auditors, has established an International Standard on Audit Sampling, here: https://www.frc.org.uk/library/standards-codes-policy/audit-assurance-and-ethics/auditing-standards/isa-uk-530/

5   These steps are consistent with some general principles set out in a recent High Court judgment, based on a joint statement agreed between the parties' statistical experts. See Amey LG Ltd v Cumbria County Council [2016] EWHC 2856 (TCC) (11 November 2016) (bailii.org), from paragraph 25.99.

6   In circumstances where advanced sampling methods are used (for example, cluster sampling), one might adjust the sampling frame to first identify clusters and then sample from within each cluster.

7   Whenever a sample is used to draw inference about a population, there is always uncertainty associated with that inference. This "sampling uncertainty" arises precisely because the sample is chosen randomly: if a second sample were drawn using the same design, a different set of units would be randomly selected, and the estimate drawn from that second sample might therefore differ. Statisticians measure such uncertainty by reference to "margins of error" and "confidence intervals". If the confidence level is 95% and the margin of error is 9%, then this indicates a 95% confidence interval of 25% ± 9%, or 16% to 34%. This confidence interval means that there is a 95% chance that the true proportion of essential patents in the broader portfolio is between 16% and 34%.

8   *Amey LG Limited v Cumbria County Council* (2016) England and Wales High Court (Technology and Construction Court), case 3MA500110. Available at https://www.bailii.org/ew/cases/EWHC/TCC/2016/2856.html.

9   *Amey v Cumbria*, at 25.110.

10 See "selection bias" in B. S. Everitt and A. Skrondal, *'The Cambridge Dictionary of Statistics'* (2010).

11 See: Peverill Squire, *'Why the 1936 Literary Digest Poll Failed'* in The Public Opinion Quarterly (Spring, 1988), https://www.jstor.org/stable/2749114?seq=1#page_scan_tab_contents

12 The judge stated that *"I am satisfied that there were a number of errors in the development of the process for choosing the samples in this case. In summary, although there were 1,706 separate works instructions involving patching issued during the course of the contract only 544 works instructions were identified and only 116 works instructions were available for selection.... There was an initial bias in the selection of the initial samples, both by year and by area. Worse than this, was the decision to focus on the patches laid in the first 3 years in heavily trafficked roads. This is an example of deliberate clear bias."* Amey v Cumbria, at 25.143 and 24.145.

13 The judge stated that *"This raises the question as to whether it is safe to extrapolate at all … In conclusion, in my view Cumbria has failed to demonstrate that the sampling exercise undertaken on its behalf in this case is a sufficiently reliable exercise to justify the court in making the finding as against Amey…"* Amey v Cumbria, at 25.153 and 25.167.

14 *"In his report and in his evidence [Mr Hodgen, Cumbria's statistical expert] sought to justify Cumbria's case on extrapolation on the basis that the sample, although not statistically random, could nonetheless be justified as being statistically representative… In so doing, he placed significant reliance upon his assessment of PTS as a company, and Mr O'Farrell as an individual, as having significant knowledge and experience in sampling…. Unfortunately for him, the evidence demonstrates quite clearly in my view… that this reliance was misplaced. Although he strove gallantly in cross-examination to support his opinions, he faced a very difficult task and, ultimately, was unsuccessful, for reasons I give in detail later."* Amey v Cumbria, at 3.76 and 3.77.

15 95% confidence levels are often used.

16 This sort of extrapolation can be useful for assessing damages. In other circumstances, the relevant question might be one of liability, such as whether or not the proportion of defective products exceeds a maximum warranted defect rate (of, for example, 20%), and therefore whether the defendant is in breach of warranty, or not. Statistical samples can be used to test such hypotheses, explicitly and quantitatively.

FTI CONSULTING ™