### FGAR <sup>Global</sup> Arbitration Review

# The European Arbitration Review

2023

The Use of Statistical Samples in Commercial Disputes

## The European Arbitration Review

2023

The European Arbitration Review 2023 contains insight and thought leadership from 15 preeminent practitioners from the region. It provides an invaluable retrospective on what has been happening in some of Europe's more interesting seats.

This edition also contains think pieces on human rights in arbitration, the mechanism's growing use in technology disputes, damages in investment disputes and the use of statistical samples.

All articles come complete with footnotes and relevant statistics.

#### Generated: March 7, 2024

The information contained in this report is indicative only. Law Business Research is not responsible for any actions (or lack thereof) taken as a result of relying on or in any way using information contained in this report and in no event shall be liable for any damages resulting from reliance on or use of this information. Copyright 2006 - 2024 Law Business Research



## The Use of Statistical Samples in Commercial Disputes

#### Meloria Meschi and Ravi Kanabar

FTI Consulting

Summary

IN SUMMARY

**DISCUSSION POINTS** 

**REFERENCED IN THIS ARTICLE** 

INTRODUCTION

WHAT IS A SAMPLE?

LESSONS FOR LEGAL PRACTITIONERS

CONCLUSIONS

ENDNOTES

#### In summary

Legal practitioners involved in the dispute resolution process are increasingly confronted with ever larger sets of data and documents. In this article, we discuss how statistical sampling – which is a well-established, intuitive and versatile tool from the field of statistics – can be used to deal with the challenge of making sense of this volume of material, in a precise, pragmatic and proportionate way, to meet the demands of the arbitration process.

#### Discussion points

- · What do legal practitioners need to know about sampling?
- How are statistical samples designed and assessed?
- How can samples be used to their full potential, while avoiding common pitfalls?

#### Referenced in this article

Amey LG Ltd v Cumbria County Council

#### Introduction

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 borrowed from economics, statistics and data science.<sup>[1]</sup> On the other hand, legal practitioners can easily find themselves inundated by the sheer volume of material that requires review, having to find the proverbial needle in the haystack, or be more likely to encounter unusual findings that are a result of chance rather than representative of the truth. In this article, we discuss how to deal with this problem in a precise, pragmatic and proportionate way – using statistical sampling.

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 it directly. Statistical sampling is not a new technique: one of the earliest recorded uses of a sample was by John Graunt (regarded as one of the founders of demography), who estimated the population of London over 400 years ago using data on the number of burials per year in a sample of parishes.<sup>[2]</sup> 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 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 assist legal practitioners in forming 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? And how can they use samples to their full potential, while avoiding the common pitfalls? We answer these questions in the rest of this article.

#### 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.'<sup>[3]</sup> Samples are used in a wide range of contexts and for different purposes:

- to take the pulse of public opinion in the run-up to elections, polling organisations conduct regular surveys of samples of voters;<sup>[4]</sup>
- 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; and
- 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.<sup>[5]</sup>

Samples are also increasingly being used across a broad range of commercial disputes to determine legal liability and assess compensatory damages. For example:

- 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 breaches of contract disputes in the insurance industry, samples of insurance claims can be audited to assess 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 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; and
- in pre-action fraud investigations, parties and legal teams often consider a sampling approach to assessing the extent of 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.<sup>[6]</sup>

Step one: define the relevant population, unit of analysis and 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 that 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.

#### Step two: 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 (under-coverage) and may also include some other social media users who are not registered voters (over-coverage).<sup>[7]</sup> 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.

#### Step three: determine the sampling method

There are many different sampling methods available, and the '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 because, although larger samples are generally better from a statistical perspective as 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. Therefore, there is a trade-off between statistical precision and confidence on the one hand, and time and cost on the other.

Step four: draw the sample and measure the characteristics of 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. In a product liability dispute, engineering

experts may be instructed to examine and test each sampled product to determine whether it was defective.

#### Step five: conduct analysis of the sample and extrapolate

For example, in an intellectual property dispute around the value of a 500-strong patent portfolio (ie, too many to assess individually), you may draw a sample of 80 patents and find that only 20 of those patents (ie, 25 per cent 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 per cent finding can be extrapolated to the broader population of 500 patents, to estimate that 125 of those will in fact be essential. Complex but well-established statistical formulae can also be used to quantity how precise this estimate is and how much confidence one can have in it by reference to 'margins of error' and 'confidence intervals'.<sup>[8]</sup>

#### Lessons for legal practitioners

This process can seem straightforward on the face of it, but complications can and do arise in practice. From our experience in providing advice and expert evidence on matters of sample design and analysis in recent litigation, arbitration and investigations, we have identified four lessons for legal practitioners.

#### 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, and then design it to meet this purpose. Likewise, legal practitioners confronted with an existing sample (perhaps designed by one of the parties at an earlier date) should seek to clarify what the original purpose of the sample was, clarify how and why it was designed, and reach an objective and dispassionate view on 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 upfront, 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 advisers had the foresight to recommend an early investment in expert advice, and thereby 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 multimillion 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 were the subject of intense and expensive argument, with multiple rounds of expert reports and much airtime during the hearing. This example shows that while 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). <sup>[9]</sup> 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'.

Lesson two: look out for sample selection biases

Sample selection bias occurs when the units that are selected for a sample are (for whatever reason) not representative of the target population,<sup>[11]</sup> leading to inaccurate and unreliable estimates of the characteristics of that population. Sample selection biases are a perennial 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*, a magazine, sent out over 10 million straw vote ballots and used the responses to predict a 55 per cent majority for presidential candidate Landon. The prediction was totally wrong: the election was in fact a landslide victory for President Roosevelt, who won 61 per cent of the vote (compared to only 37 per cent by Landon). The poll failed because there were serious sample selection biases baked in to its design.<sup>[12]</sup> First, the sample frame was biased, as the sample was drawn primarily from automobile registration lists and telephone books, which under-represented 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.<sup>[13]</sup> The Court determined that because of these (and other) failings of the sample, it was not safe to extrapolate from it, and the sample was not sufficiently reliable to substantiate the claimant's case on liability and damages.<sup>[14]</sup>

Lesson three: beware non-statistical samples

Statistical samples (sometimes called probability samples) involve randomly selecting units and using probability theory to evaluate the sample results, whereas non-statistical sample units instead use subjective judgement to select the units. For example, 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 (perhaps 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 per cent of their selected transactions were fraudulent, they could not assume that half of all transactions on the account where fraudulent, since their sample is biased (entirely by design) towards the more suspicious transactions. The distinction between statistical and non-statistical samples is, therefore, important to bear in mind when designing and evaluating a sample.

As an example of this, 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 on a large number of individual car insurance claims, causing the claimants to incur additional costs for which they sought compensation. As 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 that in their subjective judgement demonstrated the gravest and largest misstatements, and the defendants did the opposite. While this trial sample might have been sufficient for the Court's initial purposes, it was later deemed insufficient for the purpose of assessing any damages due, as the results could not be reliably extrapolated to all relevant claims.

Further, in *Amey v Cumbria*, Cumbria 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'.<sup>[15]</sup> This example shows that while 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 they leave the door open to more sophisticated analyses in the future, which would not be possible with a small sample. This notion can lead parties (usually the defendant) to seek out as large a sample as possible. However, as we explain above, there is invariably a trade-off to be made between statistical precision and confidence on the one hand, and time and cost on the other. Irrespective of the benefits, litigation and arbitrations operate on specific timetables, and costs must be borne in mind. Further, 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 (ie, 50 per cent). Using statistical theory, we could extrapolate from this finding that, with 95 per cent confidence, the number of defective products in the entire order of 10,000 is between<sup>[16]</sup> 3,600 and 6,400. 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 that, with 95 per cent confidence, the number of defective products in the entire order is between 4,000 and 6,000 products (ie, 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 more precise, but the improvement itself would diminish: this time, the confidence interval would only be slightly narrower, being 42 to 58 per cent, or 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 and dispute resolution processes, and in some cases, a more creative and sophisticated approach to sampling.

For example, we recently assisted a client operating in the water distribution industry to conduct a preliminary (ie, pre-claim) investigation into the extent to which the client had been defrauded by customers systematically under-reporting their true water usage and underpaying their water bills. Owing 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, while still producing a sample that met the purpose.

#### Conclusions

While the availability of large sets of data and documents is a mixed blessing for legal practitioners involved in the dispute resolution process, statistical samples are a well-established, intuitive and versatile tool that can be used to deal with this problem in a precise, pragmatic and proportionate way. However, sample design and analysis is deceptively simple and sometimes quite unintuitive. Legal practitioners faced with considering an existing sample or developing a new one may find it helpful to understand the key steps in the sampling process, to bear in mind the lessons we have highlighted from our experience, and to seek expert advice and input at an early stage.

#### Footnotes

[1] For a discussion of how these techniques can be used to provide compelling evidence on factual causation, see: Meloria Meschi, David Eastwood and Ravi Kanabar, *Connecting Cause and Effect* in Global Arbitration Review (2020), https://globalarbitrationreview.com/review/the-european-arbitration-review/ 2021/article/connecting-cause-and-effect. [2] 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), pages 81-105, <u>https://onlinelibrary.wiley.com/doi/book/10.1002/0471725161</u>.

[3] B. S. Everitt and A. Skrondal, The Cambridge Dictionary of Statistics (2010).

[4] See for example, The British Polling Council, About the BPC, <u>https://www.britishpollingcouncil.org/</u>.

[5] For example, the Financial Reporting Council, the UK regulator for auditors, has established an International Standard on Audit Sampling, here: <u>https://www.frc.org.uk/getattachment/d4de8d94-03d9-49b9-8a6d-045864b75494</u> /ISA-(UK)-530.pdf.

[6] 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 <u>Amey LG Ltd v Cumbria County Council [2016] EWHC 2856</u> (TCC) (11 November 2016) (-bailii.org). From paragraph 25.99.

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

[8] 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 was to be drawn using the same design, then a different set of units would be randomly selected, and therefore the estimate drawn from that second sample may differ. Statisticians measure such uncertainty by reference to margins of error and confidence intervals. If the confidence level is 95 per cent and the margin of error is 9 per cent, then this indicates a 95 per cent confidence interval of 25 per cent  $\pm$  9 per cent, or 16 to 34 per cent. This confidence interval means that there is a 95 per cent chance that the true proportion of essential patents in the broader portfolio is between 16 and 34 per cent.

[9] '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. From here on, referred to as 'Amey v Cumbria' for brevity.

[10] Amey v Cumbria, 25.110.

[11] See "selection bias" in B. S. Everitt and A. Skrondal, The Cambridge Dictionary of Statistics (2010).

[12] See: 1936 Peverill Squire, Why the Literary Digest Poll Failed The Public Opinion 1988), in Quarterly (Spring, https://www.jstor.org/stable/2749114?seg=1#page\_scan\_tab\_contents.

[13] 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*, 25.143 and 24.145.

[14] 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*, 25.153 and 25.167.

[15] '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*, 3.76 and 3.77.

[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 per cent), and therefore whether the defendant is in breach of warranty, or not. Statistical samples can also be used to test such hypotheses, explicitly and quantitatively.

#### IN SUMMARY

Legal practitioners involved in the dispute resolution process are increasingly confronted with ever larger sets of data and documents. In this article, we discuss how statistical sampling – which is a well-established, intuitive and versatile tool from the field of statistics – can be used to deal with the challenge of making sense of this volume of material, in a precise, pragmatic and proportionate way, to meet the demands of the arbitration process.

#### **DISCUSSION POINTS**

- · What do legal practitioners need to know about sampling?
- How are statistical samples designed and assessed?
- · How can samples be used to their full potential, while avoiding common pitfalls?

#### **REFERENCED IN THIS ARTICLE**

Amey LG Ltd v Cumbria County Council

#### INTRODUCTION

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 borrowed from economics, statistics and data science.<sup>[11]</sup> On the other hand, legal practitioners can easily find themselves inundated by the sheer volume of material that requires review, having to find

the proverbial needle in the haystack, or be more likely to encounter unusual findings that are a result of chance rather than representative of the truth. In this article, we discuss how to deal with this problem in a precise, pragmatic and proportionate way – using statistical sampling.

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 it directly. Statistical sampling is not a new technique: one of the earliest recorded uses of a sample was by John Graunt (regarded as one of the founders of demography), who estimated the population of London over 400 years ago using data on the number of burials per year in a sample of parishes.<sup>[2]</sup> 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 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 assist legal practitioners in forming 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? And how can they use samples to their full potential, while avoiding the common pitfalls? We answer these questions in the rest of this article.

#### 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.<sup>[3]</sup> Samples are used in a wide range of contexts and for different purposes:

- to take the pulse of public opinion in the run-up to elections, polling organisations conduct regular surveys of samples of voters;<sup>[4]</sup>
- 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; and
- 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.<sup>[5]</sup>

Samples are also increasingly being used across a broad range of commercial disputes to determine legal liability and assess compensatory damages. For example:

- 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 breaches of contract disputes in the insurance industry, samples of insurance claims can be audited to assess 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 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; and
- in pre-action fraud investigations, parties and legal teams often consider a sampling approach to assessing the extent of 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.<sup>[6]</sup>

#### Step One: Define The Relevant Population, Unit Of Analysis And 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 that 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.

#### Step Two: 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 (under-coverage) and may also include some other social media users who are not registered voters (over-coverage).<sup>[7]</sup> 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.

#### Step Three: Determine The Sampling Method

There are many different sampling methods available, and the '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 because, although larger samples are generally better from a statistical perspective as 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. Therefore, there is a trade-off between statistical precision and confidence on the one hand, and time and cost on the other.

#### Step Four: Draw The Sample And Measure The Characteristics Of 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. In a product liability dispute, engineering experts may be instructed to examine and test each sampled product to determine whether it was defective.

#### Step Five: Conduct Analysis Of The Sample And Extrapolate

For example, in an intellectual property dispute around the value of a 500-strong patent portfolio (ie, too many to assess individually), you may draw a sample of 80 patents and find that only 20 of those patents (ie, 25 per cent 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 per cent finding can be extrapolated to the broader population of 500 patents, to estimate that 125 of those will in fact be essential. Complex but well-established statistical formulae can also be used to quantity how precise this estimate is and how much confidence one can have in it by reference to 'margins of error' and 'confidence intervals'.<sup>[8]</sup>

#### LESSONS FOR LEGAL PRACTITIONERS

This process can seem straightforward on the face of it, but complications can and do arise in practice. From our experience in providing advice and expert evidence on matters of sample design and analysis in recent litigation, arbitration and investigations, we have identified four lessons for legal practitioners.

#### 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, and then design it to meet this purpose. Likewise, legal practitioners confronted with an existing sample (perhaps designed by one of the parties at an earlier date) should seek to clarify what the original purpose of the sample was, clarify how and why it was designed, and reach an objective and dispassionate view on 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 upfront, 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 advisers had the foresight to recommend an early investment in expert advice, and thereby 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 multimillion 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 were the subject of intense and expensive argument, with multiple rounds of expert reports and much airtime during the hearing. This example shows that while 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).-<sup>[9]</sup> 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'.<sup>[10]</sup>

#### Lesson Two: Look Out For Sample Selection Biases

Sample selection bias occurs when the units that are selected for a sample are (for whatever reason) not representative of the target population,<sup>[11]</sup> leading to inaccurate and unreliable estimates of the characteristics of that population. Sample selection biases are a perennial 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*, a magazine, sent out over 10 million straw vote ballots and used the responses to predict a 55 per cent majority for presidential candidate Landon. The prediction was totally wrong: the election was in fact a landslide victory for President Roosevelt, who won 61 per cent of the vote (compared to only 37 per cent by Landon). The poll failed because there were serious sample selection biases baked in to its design.<sup>[12]</sup> First, the sample frame was biased, as the sample was drawn primarily from automobile registration lists and telephone books, which under-represented 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.<sup>[13]</sup> The Court determined that because of these (and other) failings of the sample, it was not safe to extrapolate from it, and the sample was not sufficiently reliable to substantiate the claimant's case on liability and damages.<sup>[14]</sup>

#### Lesson Three: Beware Non-statistical Samples

Statistical samples (sometimes called probability samples) involve randomly selecting units and using probability theory to evaluate the sample results, whereas non-statistical sample units instead use subjective judgement to select the units. For example, 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 (perhaps 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 per cent of their selected transactions were fraudulent, they could not assume that half of all transactions on the account where fraudulent, since their sample is biased (entirely by design) towards the more suspicious transactions. The distinction between statistical and non-statistical samples is, therefore, important to bear in mind when designing and evaluating a sample.

As an example of this, 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 on a large number of individual car insurance claims, causing the claimants to incur additional costs for which they sought compensation. As 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 that in their subjective judgement demonstrated the gravest and largest misstatements, and the defendants did the opposite. While this trial sample might have been sufficient for the Court's initial purposes, it was later deemed insufficient for the purpose of assessing any damages due, as the results could not be reliably extrapolated to all relevant claims.

Further, in *Amey v Cumbria*, Cumbria 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'.<sup>[15]</sup> This example shows that while 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 they leave the door open to more sophisticated analyses in the future, which would not be possible with a small sample. This notion can lead parties (usually the defendant) to seek out as large a sample as possible. However, as we explain above, there is invariably a trade-off to be made between statistical precision and confidence on the one hand, and time and cost on the other. Irrespective of the benefits, litigation and arbitrations operate on specific timetables, and costs must be borne in mind. Further, 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 (ie, 50 per cent). Using statistical theory, we could extrapolate from this finding that, with 95 per cent confidence, the number of defective products in the entire order of 10,000 is between <sup>[16]</sup> 3,600 and 6,400. 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 that, with 95 per cent confidence, the number of defective products in the entire order is between 4,000 and 6,000 products (ie, 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 more precise, but the improvement itself would diminish: this time, the confidence interval would only be slightly narrower, being 42 to 58 per cent, or 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 and dispute resolution processes, and in some cases, a more creative and sophisticated approach to sampling.

For example, we recently assisted a client operating in the water distribution industry to conduct a preliminary (ie, pre-claim) investigation into the extent to which the client had been defrauded by customers systematically under-reporting their true water usage and underpaying their water bills. Owing 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, while still producing a sample that met the purpose.

#### CONCLUSIONS

While the availability of large sets of data and documents is a mixed blessing for legal practitioners involved in the dispute resolution process, statistical samples are a well-established, intuitive and versatile tool that can be used to deal with this problem in a precise, pragmatic and proportionate way. However, sample design and analysis is deceptively simple and sometimes quite unintuitive. Legal practitioners faced with considering an existing sample or developing a new one may find it helpful to understand the key steps in the sampling process, to bear in mind the lessons we have highlighted from our experience, and to seek expert advice and input at an early stage.

#### Endnotes

- 1 For a discussion of how these techniques can be used to provide compelling evidence on factual causation, see: Meloria Meschi, David Eastwood and Ravi Kanabar, Connecting Cause and Effect in Global Arbitration Review (2020), <u>https://globalarbitrationreview.com/review/the-european-arbitration-review/</u> 2021/article/connecting-cause-and-effect. <u>ABack to section</u>
- 2 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), pages 81-105, <u>https://onlinelibrary.wiley.com/doi/book/10.1002/0471725161</u>. <u>> Back to section</u>
- 3 B. S. Everitt and A. Skrondal, The Cambridge Dictionary of Statistics (2010). <u>A Back to section</u>

- 4 See for example, The British Polling Council, About the BPC, <u>https://www.britishpollingcouncil.org/</u>. <u>About the BPC</u>, <u>Back to section</u>
- 5 For example, the Financial Reporting Council, the UK regulator for auditors, has established an International Standard on Audit Sampling, here: <u>I</u> <u>https://www.frc.org.uk/getattachment/d4de8d94-03d9-49b9-8a6d-045864b75494</u> /ISA-(UK)-530.pdf. <u>Back to section</u>
- 6 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 <u>Amey LG Ltd v Cumbria County Council [2016] EWHC 2856</u> (TCC) (11 November 2016) (<u>bailii.org</u>). From paragraph 25.99. <u>ABack to section</u>
- 7 In circumstances where advanced sampling methods are used (eg, cluster sampling), one might adjust the sampling frame to first identify clusters and then sample from within each cluster. <u>A Back to section</u>
- 8 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 was to be drawn using the same design, then a different set of units would be randomly selected, and therefore the estimate drawn from that second sample may differ. Statisticians measure such uncertainty by reference to margins of error and confidence intervals. If the confidence level is 95 per cent and the margin of error is 9 per cent, then this indicates a 95 per cent confidence interval of 25 per cent ± 9 per cent, or 16 to 34 per cent. This confidence interval means that there is a 95 per cent chance that the true proportion of essential patents in the broader portfolio is between 16 and 34 per cent. A <u>Back to section</u>
- 9 'Amey LG Limited v Cumbria County Council' (2016) England and Wales High Court (Technology and Construction Court), case 3MA500110. Available at <u>https://www.bailii.org/ew/cases/EWHC/TCC/2016/2856.html</u>. From here on, referred to as 'Amey v Cumbria' for brevity. <u>> Back to section</u>
- **10** Amey v Cumbria, 25.110. <u>Amey v Cumbria</u>, 25.110.
- 11 See "selection bias" in B. S. Everitt and A. Skrondal, The Cambridge Dictionary of Statistics (2010). <u>A Back to section</u>
- 12 See: Peverill Squire, *Why the 1936 Literary Digest Poll Failed* in The Public Opinion Quarterly (Spring, 1988), <u>https://www.jstor.org/stable/2749114?seq=1#page\_scan\_tab\_contents</u>. <u>A Back to</u> <u>section</u>

- **13** 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*, 25.143 and 24.145. <u>Amey v Cumbria</u>, 25.143 and 24.145.
- 14 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, 25.153 and 25.167. <a href="https://www.backtosection">Back to section</a>
- 15 '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*, 3.76 and 3.77. <u>ABack to section</u>
- 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 per cent), and therefore whether the defendant is in breach of warranty, or not. Statistical samples can also be used to test such hypotheses, explicitly and quantitatively. <u>ABACK to section</u>



200 Aldersgate, Aldersgate Street, London EC1A 4HD, United Kingdom

Tel:+44 20 3727 1000

https://www.fticonsulting.com/

Read more from this firm on GAR