

A Numbers Game? Statistics in Public Law Cases

ALBA Annual Lecture, 5 July 2021

Lady Rose, Justice of the Supreme Court

I. Introduction

1. We are living in an increasingly statistical world.¹ Complex statistical modelling is now commonplace, and, as a consequence, assessing the validity and reliability of statistical evidence is becoming ever more difficult. In daily Coronavirus updates, we are shown facts and figures displaying both absolute numbers and also changes in the rate of vaccinations, infections, hospitalisations, and deaths. Should we wish them to do so, our phones or other electronic devices can monitor and provide us with statistics showing our daily steps, screen time, and even our sleep quality. Statistics are also increasingly visible and influential in the world of sport. In football it is now easy to track the distance that a player runs throughout a game, the top speed that they reach, and even how many goals they scored as compared to their “expected goals”, i.e., the number of goals that they ought to have scored given the quality of the chances that they had.
2. There are some areas where humans are particularly fallible at making use of statistics to take rational decisions. An important one is in assessing risk and probability. I remember in the mid-1990s there was an unfortunate coincidence of two major developments in the news. The one was the transmission of BSE (mad cow disease) to humans who contracted variant Creutzfeldt-Jakob disease from eating contaminated beef. 177 people died from contracting variant CJD and the Government tried to overcome the reluctance of the public to buy and eat British beef by stressing what a vanishingly small risk this involved. At the same time, however, they were promoting the national lottery which was launched in 1994 with the odds of winning standing at 1 in 45 million. The ubiquitous slogan “It could be you” was regarded as much as a sinister warning about eating beef as it was an encouragement to buy a lottery ticket.
3. On a more tragic note the case of *R v Clark (Sally)*, provides an example of how dangerous statistics can be when they are misused.² In *Clark*, the defendant mother was charged with murder after the sudden death, on separate occasions,

of two of her infant children. A pathology expert appearing for the Crown testified that the probability of two Sudden Infant Death Syndrome deaths in one family matching the defendant's profile was 1 in 73 million. Mrs Clark was convicted of murdering her two sons by a majority of 10 to 2.

4. Unfortunately, as the Court of Appeal subsequently identified, the statistical evidence was dangerous and should never have been before the jury. Even if the 1 in 73 million figure correctly represented the chance of two separate SIDS deaths in a family matching the defendant's profile, the pathologist was wrong to suggest that this meant that, in circumstances where there had been two infant deaths in a family, the chance that both of the children had died from natural causes was 1 in 73 million. By conflating the two, the pathologist had fallen victim to what is known as the "*prosecutor's fallacy*".³ The Court of Appeal rightly noted that the pathologist's statistical evidence "*was tantamount to saying that without consideration of the rest of the evidence one could be just about sure that this was a case of murder*".

II. The Use of Statistics in Public Law

5. Inevitably, the world of public law litigation has not escaped the statistical revolution. In competition and intellectual property law, with which I am familiar through my time in practice and sitting in the Chancery Division, statistical evidence is now so commonly used in certain contexts that comprehensive guidelines have been developed by the courts to control its use and case management routinely involves an early discussion between the parties and the court as to what evidence will be presented at trial and how it will be prepared. The evidence is usually presented by experts subject to the usual duties and responsibilities that expert witnesses owe to the court. In public law, by contrast, it strikes me that comparatively little thought has been given to the use of statistics, or how and when that evidence ought to be subjected to judicial scrutiny.
6. As a consequence, I hope in this talk to elucidate some thoughts about the use of statistics in the law, and how other areas of law might prove instructive for courts and practitioners grappling with the difficulties of analysis in the context of public law.

UNISON

7. Both European and domestic courts are generally content to rely upon statistical evidence in public law cases. In fact, in some cases statistics can carry the day, being key to establishing or disproving the various elements of a cause of action.
8. One relatively recent example is the well-known and much loved *UNISON* case, in which a trade union challenged an Order in Council that imposed fees for bringing claims in the employment tribunals (the “Fees Order”).⁴ UNISON claimed that the fees were unaffordable for certain claimants and that, as a consequence, the Fees Order was ultra vires the fee setting power because it interfered unduly with the constitutional right of access to the courts. When the case ultimately came before the Supreme Court in 2017, statistical evidence —some of which had not been available to the courts below—proved vital.
9. Beyond simply showing that there had been a very sharp fall in the volume of claims after the introduction of the Fees Order, statistics also enabled the court to conclude that fees order did not achieve the Government’s stated objectives. They showed that the Lord Chancellor had failed to set the fees at the optimal price so that the reduction in numbers of claims meant that the fees had not achieved the stated objective of transferring more of the costs of running the tribunals from taxpayers to tribunal users. The figures also failed to support the Lord Chancellor’s claims that the fees were helpful either in deterring unmeritorious claims or in incentivising earlier settlements. In fact, the union could show that more unsuccessful claims (and fewer successful claims) were brought after the introduction of the Fees Order, and the proportion of cases settled through mandatory conciliation decreased after fees were introduced.⁵ Finally, statistics helped to make good UNISON’s case on indirect discrimination. Statistical evidence showed that a higher proportion of women (and others with protected characteristics) brought the more complicated Type B kinds of claims, which were subject to higher fees. This was sufficient to establish a case of *prima facie* discrimination.
10. The Supreme Court’s decision in *UNISON* therefore demonstrates the great potential of statistical evidence in public law claims.

What Type of Statistical Evidence is Allowed?

11. Given the dual importance and risks of statistical evidence, one might expect strict controls on its use and a close and early analysis by the courts of the sources and methodologies used behind the figures relied on. And one obvious candidate for such a control mechanism would be a restriction on the type of evidence that can be used.
12. For example, it might be thought that courts should require parties only to rely upon officially produced data, since those are likely to provide the most reliable material. Thus, in *Hoogendijk v the Netherlands*,⁶ an early authority of the European Court of Human Rights in 2005 on the use of statistics in indirect discrimination, the Court held that “*where an applicant is able to show, on the basis of undisputed official statistics, the existence of a prima facie indication that a specific rule - although formulated in a neutral manner - in fact affects a clearly higher percentage of women than men, it is for the respondent Government to show that this is the result of objective factors unrelated to any discrimination on grounds of sex*” (emphasis added).
13. Unfortunately, the reality is that this would probably be too restrictive: in many public law challenges, official statistics will not be available and so restrictions on the type of evidence that can be relied upon might prevent good claims being brought. The European Courts have recognised this problem, again in the context of indirect discrimination claims. In *DH v Czech Republic*,⁷ in 2008 the Court observed that “*applicants may have difficulty in proving discriminatory treatment*”, and said that, “*[i]n order to guarantee those concerned the effective protection of their rights, less strict evidential rules should apply in cases of alleged indirect discrimination.*”⁸ This was said to follow from *Nachova and Others v. Bulgaria*⁹ in which the Court said in 2006 that there are no procedural barriers to the admissibility of evidence or pre-determined formulae for its assessment.
14. Domestic courts appear to have adopted the same approach. And as a consequence, recent discrimination cases have shown us a host of innovative ways in which parties have employed statistical evidence in support of their arguments. Leaving aside statistics derived from survey evidence (which I want to discuss in more detail later in this talk), two examples warrant particular

mention for their creativity:

- i. First, in the Joint Council for the Welfare of Immigrants' recent challenge to the Government's "*Right to Rent*" scheme,¹⁰ "*mystery shopping*" exercises were used by both parties in order to test whether the scheme resulted in discrimination in housing access. Both these exercises involved inquiries of landlords from six fictitious tenants who differed in various ways (for example, as to whether they had British citizenship, an ethnically British name, or a British passport). Landlords' responses (as well as failures to respond) were then tracked and analysed to see if evidence showing the potential for discrimination emerged.
- ii. Second, in *Doreen McNeil and others v Commissioners for Her Majesty's Revenue and Customs*,¹¹ a differential distribution analysis was used in support of the Appellants' argument that HMRC's pay practices disadvantaged women. The analysis considered the distribution of men and women within each of HMRC's pay grades, and showed, statistically, that women were disproportionately over-represented at the lower end of the pay scale for each of HMRC's grades, and disproportionately under-represented at the upper end.¹²

15. Whilst neither challenge was ultimately successful, both pieces of evidence were held to be admissible and the JCWI's mystery shopper exercise was highlighted as a key piece of evidence that indicated that, as a result of the Right to Rent Scheme, some landlords did discriminate against potential tenants without British passports, and particularly those without a British Passport or an ethnically British name.¹³

How do Courts Control the Quality of the Statistical Data Used?

16. Since prescribed restrictions on the type of statistical evidence that can be used might prove problematic, we must focus our attention on how courts control the quality of statistical data. This leads us back to *DH*, in which the European Court said that, although statistics are not necessary, nor do they have to be "*official*", they must at least be "*valid and significant*" or "*reliable and significant*".¹⁴

17. In examining what these terms mean, we can start with the meaning of statistical

significance. This was explored by the European Court of Justice in *R v Secretary of State for Employment, ex parte Seymour-Smith and Perez* in 1999.¹⁵ The case concerned an order that raised the qualifying period for protection against unfair dismissal from one to two years of continuous employment. The applicants, who were women, challenged the rule on the basis that it constituted unlawful indirect discrimination. The House of Lords referred various questions to the Court in Luxembourg, including the legal test for establishing whether a measure adopted by a Member State has the necessary degree of disparate effect so as to give rise to a *prima facie* case of indirect discrimination.

18. The Court set out the rival positions adopted before it. Thus, (a) the applicants contended that where there are significant statistics covering the entire workforce which cannot be regarded as fortuitous, "anything more than a minimal difference" in the impact of the measure could be a breach; (b) the United Kingdom contended for a requirement that there be "a markedly different impact"; (c) the Commission proposed a test by which "the existence of statistically significant" evidence is enough to establish disproportionate impact. The Court held that the question that had to be answered was whether the statistics available indicate that "*a considerably smaller percentage of women than men is able to satisfy the [two-year] condition*". The condition could be satisfied if "*the statistical evidence revealed a lesser but persistent and relatively constant disparity over a long period between men and women who satisfy the requirement...*" However, it was ultimately "*for the national court to assess whether the statistics concerning the situation of the workforce are valid and can be taken into account, that is to say, whether they cover enough individuals, whether they illustrate purely fortuitous or short-term phenomena, and whether, in general, they appear to be significant.*"¹⁶

19. Courts have interpreted this to mean that the requirement of significance runs broader than "*statistical significance*" in the technical sense.¹⁷ But the question of significance still causes problems. When the *Seymour-Smith* case came back before the House of Lords¹⁸ their Lordships split 3-2 over whether the ratio of women qualifying for protection as compared to men (which had slightly increased from 89:100 to 90.5:100 in the six years since the order amending the qualifying period was introduced) was sufficiently significant to give rise to a

prima facie case of indirect discrimination. The majority decided that the test was met so that the Government did have to justify the reduction of the qualifying period.

20. What about reliability and validity? The answer is that courts seem to take a case-by-case approach. In some cases, the courts scrutinise competing statistics quite carefully. For example, in *JCWI*, the Court of Appeal looked closely at the competing mystery shopper exercises (and in particular the *JCWI*'s exercise, which had been especially influential at first instance). In assessing the weight to be given to that exercise, the Court took into account (amongst other things): (i) the level of consistency of the evidence as a whole; (ii) possible biases inherent in the design of the exercise; and (iii) the replicability of the data produced by the exercise.¹⁹
21. In other cases, courts have been unwilling to get into a debate over whether statistics are reliable and valid at all. For example, in *R (on the application of Roberts) v Commissioner of Police of the Metropolis and another*,²⁰ Lord Justice Moses sitting in the Administrative Court, declined to rule on rival statistical evidence in relation to the difference in likelihood, based on one's ethnic group, of being stopped and searched by police through their powers under section 60 of the Criminal Justice and Public Order Act 1994. He held that, "[i]f a court had to rule on the use of the power of stop and search under s.60, it would require properly prepared expert reports identifying those issues upon which the experts were agreed and those on which they disagreed. The areas for resolution by the court where the experts disagreed would have to be identified. Issues simply cannot and should not be resolved merely by reading undigested material in the hope that the court's own expertise in statistical analysis will enable it to reach a conclusion."²¹ The Court of Appeal agreed. Lord Justice Maurice Kay said: "I do not think we should become embroiled in tendentious statistical material" in circumstances where it was "readily apparent that the statistics are controversial and give rise to difficult issues of interpretation which it would be difficult to resolve without expert assistance of a kind with which [the court had] not been provided."²²
22. In many cases, it seems to me, the courts take a rather impressionistic approach

to what strikes the judge as helpful with little express analysis of what the data can or do show. One aspect which has struck me is reliance on a comparison between absolute numbers and reliance on comparative proportions. In some cases, courts appear to be satisfied that a *prima facie* case of discrimination will be established if it is shown that a seemingly neutral measure negatively affects a far greater number of individuals who share a protected characteristic. As an example, in a number of recent welfare benefits cases,²³ it has proven uncontroversial that measures that limit access to benefits for single parents bringing up children inevitably give rise to a *prima facie* case of indirect discrimination against women since the number of single women bringing up children is much greater than the number of single men doing the same.²⁴ Neither the court nor the parties seems to have thought that the correct question is to ask what proportion of single women bringing up children qualify for or claim benefits as compared with the proportion of single men bringing up children and qualifying or claiming benefits.

23. In other cases, it is the proportions that are considered relevant. For example, the *Essop* decision in the Supreme Court in 2017 gave an important boost to statistical analysis.²⁵ In that case, immigration officers employed by the Home Office complained that the requirement to pass a Core Skills Assessment test in order to be promoted to certain civil service grades was indirectly discriminatory. A report from independent consultants revealed that candidates from Black, Asian and minority ethnic communities and older candidates had a much lower pass rate than White and younger candidates. No one knew why. The question for the Supreme Court was whether the claimants had to show the reason for the lower pass rate in order to succeed in their claim for indirect discrimination. The Supreme Court held that they did not. But the important thing for my purposes is that it was assumed that it was not enough to show that more White people passed than people from Black, Asian and minority ethnic communities since that might just reflect that there were many more White candidates taking the test. The statistics presented to the Court therefore compared the proportions of those passing.

24. One might say that that was obvious but it is actually difficult to express why one parameter is relevant in one indirect discrimination challenge and the other in

another. One case in which this was controversial was the Court of Appeal's recent decision in *R (on the Application of QSA and Ors) v Secretary of State for the Home Department & Anor.*²⁶ QSA concerned a challenge to, amongst other things, the criminalisation of loitering or soliciting in a street or public place for the purpose of prostitution. It was argued that the criminalisation of soliciting was in breach of Article 8 read with Article 14 of the ECHR because it was gender discriminatory. The Applicants presented evidence to show that although the statutory provision rendering the conduct criminal had been revised in 2003 so that the offence could be committed by men as well as women, the statistics showed that over 98% of those convicted of the relevant offences were women.

25. The Court of Appeal held this was not sufficient because all it showed was that many more women commit the crime than men. As Lord Justice Bean said, most criminal offences, certainly most offences of violence, are committed overwhelmingly by men, but that has never so far given rise to a successful argument that the existence of the particular offence is a breach of the defendant's rights under Article 14 read with Article 8, or that it has a disparate adverse impact on men. What needed to be shown to get a case of discrimination off the ground was that women who solicit were in some way disproportionately disadvantaged compared to men who solicit.²⁷

26. Perhaps the answer is that figures about single women bringing up children are necessarily reflective of proportions since, presumably for every woman with a child there must be a corresponding man with that same child though he may long since have disappeared from the scene. But I have not seen cases analysed in this way.

III. Survey Evidence: Indirect Discrimination vs Trade-Mark Cases

27. We can therefore see that courts in public law cases take a highly individualised and flexible approach to statistics. In particular I am not aware of it being common practice for there to be any analysis by the court in advance of the substantive hearing of what statistics are going to be presented and how they should be produced. It does not seem to be considered in the case management stages of an application before the substantive hearing. As I have foreshadowed, this is in contrast to the relatively consistent and rigorous approach taken in

competition and intellectual property law cases. In order to provide a point of direct comparison, I want to now look at the treatment of survey evidence in each area of law.

Surveys in Public Law

28. In line with the generally flexible approach to statistics in public law, both domestic and European courts appear to adopt varying approaches to scrutinising survey data in public law cases. For example:

- i. In *UNISON*, the Supreme Court adopted a relatively light touch to scrutinising survey data. In support of his conclusion that the Fees Order impeded the right of access to the courts, Lord Reed relied upon a 2015 survey of individuals whose employment disputes had failed to settle through mandatory conciliation, but who had not proceeded to file a claim in the employment tribunals. The most frequently mentioned reason for not bringing a claim following a failed conciliation was that the fees were off putting, and more than two thirds of claimants who gave that reason said that they could not afford the fees. Without exploring the issue in detail, Lord Reed said it was “*not obvious why the explanation given by the claimants should not be accepted.*”²⁸
- ii. In *JWCI*, however, the Court of Appeal adopted a stricter approach. It looked closely at the results of a considerable number of surveys carried out by JCWI and other organisations asking questions to landlords, letting agents and organisations working in fields related to migration, housing and discrimination. (this was separate from the mystery shopper exercise). It was aimed at establishing whether the introduction of sanctions on landlords for letting to people who had no entitlement to be in this country led them to discriminate against prospective tenants on grounds of ethnicity. The survey evidence was specifically criticised because, amongst other things: (i) the JCWI’s survey had no control group; (ii) the sample size was small, limiting the Court’s ability to extrapolate the results; and (iii) it did not identify the extent to which the new Scheme had exacerbated discrimination in the housing market which already existed before the scheme was introduced. As a consequence, although the Court

did place some weight on the survey (in light of the evidence as a whole), Lord Justice Hickinbottom noted that the Secretary of State's submission that the survey needed to be approached with "some caution" had "some force".²⁹

Surveys in Trade-Mark Infringement and Passing Off Cases

29. With that said, even the more detailed analysis of survey evidence in *JCWI* is relatively light touch in comparison to the strict two-stage approach taken in trade-mark infringement and passing off cases within the field of intellectual property law.
30. Survey evidence has historically been commonplace in these cases.³⁰ In particular, surveys in which members of the public are shown different products and asked questions about what the products mean to them have often been relied on to show either that a particular trade-mark was distinctive or had acquired goodwill, or to establish that the defendant's actions create confusion on the part of the public over the seller of particular goods and/or their association with the claimant.
31. Because trade-mark cases are concerned with people's unconscious assumptions, there is a particularly acute risk that the wording or design of surveys risks distorting what they test (e.g., because respondents search for the answers they "should" give to the questions).³¹ Perhaps for this reason, courts have laid down rigorous guidelines which need to be followed when preparing survey evidence. These guidelines help the courts to decide whether the evidence is admissible and, if so, whether it ought to be given significant weight.
32. The first hurdle for claimants is getting permission to conduct a survey. The debate about whether and how a survey should be conducted is therefore a key element in the early case management of the trial. Whilst previously the Court would admit technically admissible evidence (including survey evidence) unless the judge could be satisfied that it would be valueless, this procedure was abolished in the first *Interflora v Marks & Spencer* case.³² Now: (i) a party may conduct a true pilot survey without permission, but at its own risk as to costs; (ii) no further survey may be conducted or adduced in evidence without the court's

permission; and (iii) no party may adduce evidence from people replying to survey questions without the court's permission. The court will only grant permission for a survey to be conducted if it is satisfied that the evidence is likely to be of real value. If permission is sought to carry out a full survey, the applicant will also have to provide the court with: (i) the results of any pilot survey; (ii) evidence that any further survey will comply with what are known as the Whitford guidelines (which I will consider further in a second); and (iii) the cost of carrying out the pilot survey and the estimated cost of carrying out the further survey.³³

33. Once parties have been given permission to carry out a full survey, that survey is assessed³⁴ by reference to the "Whitford Guidelines". These guidelines were formulated by Mr Justice Whitford as long ago as 1984 in *Imperial Group plc v Philip Morris Ltd*,³⁵ and were recently summarised in the first *Interflora v Marks & Spencer* case as follows:

- "i) if a survey is to have any validity at all, the way in which the interviewees are selected must be established as being done by a method such that a relevant cross-section of the public is interviewed;
- ii) any survey must be of a size which is sufficient to produce some relevant result viewed on a statistical basis;
- iii) the party relying on the survey must give the fullest possible disclosure of exactly how many surveys they have carried out, exactly how those surveys were conducted and the totality of the number of persons involved, because otherwise it is impossible to draw any reliable inference from answers given by a few respondents;
- iv) the questions asked must not be leading; and must not direct the person answering the question into a field of speculation upon which that person would never have embarked had the question not been put;
- v) exact answers and not some sort of abbreviation or digest of the exact answer must be recorded;
- vi) the totality of all answers given to all surveys should be disclosed to the opposing party to the litigation; and

vii) the instructions given to interviews must also be disclosed”³⁶

34. The guidelines therefore set out strict criteria against which surveys can be assessed. Further the results of the survey are usually presented to the court by the expert witness who devised and conducted the survey and who can be cross examined about the methodology used and the conclusions he or she has drawn from the data. There have grown up a number of expert litigation consultants who have become familiar faces giving expert evidence in the IP Courts because they are trusted by the parties and by the judges to act in a fair and impartial way compliant with the guidelines and with their duties to the Court.
35. An example of how strictly and rigorously they are applied can be found in *Glaxo Wellcome UK Ltd and others v Sandoz Ltd and others*.³⁷ The case involved a dispute about the colour purple. The claimants (“Glaxo”) sold inhalers under the trade-mark “*Seretide*”. The inhalers were coloured in two shades of purple and were also sold in packaging featuring a shade of purple. The defendants launched a generic inhaler which contained the same active ingredients as the Seretide inhalers. The competing generic inhaler was also largely coloured in a shade of purple and was sold in packaging featuring a shade of purple. Glaxo argued that the defendants passed off the generic inhaler as being: (i) connected in the course of trade with Glaxo; or (ii) equivalent to one of their Seretide inhalers.
36. Glaxo sought to support their claim by reference to four different surveys. Two were carried out in March 2015, one of 200 GPs and the other of 200 pharmacists. In response to the defendants’ various criticisms of the first two sets of surveys, two further surveys were conducted from March-April 2016, one of 251 GPs and the other of 252 pharmacists. In 2017, Glaxo received permission from the Court to rely upon the surveys.³⁸
37. Glaxo was not as successful at trial. In a detailed judgment, Mr Justice Arnold (as he then was) devoted 46 paragraphs to considering the surveys and the application of each of the Guidelines.³⁹ This exercise led him to conclude that the 2015 surveys did not comply with the Guidelines in a number of respects and consequently were of no value. The 2016 surveys fared slightly better but was still found to be defective in a number of important respects. For example, one

question – “How do patients typically refer to the inhaler that you mentioned?”⁴⁰ was held to be overly vague; it was likely to result in responses biased toward mentioning purple because interviewees had been given a purple colour card at the start of the interview which they kept sight of throughout. In a final nail in the coffin of Glaxo’s surveys, Mr Justice Arnold noted that, even taken at face value, they did not demonstrate that the colour purple was distinctive of Glaxo’s Seretide inhalers. They did not prove that GPs or pharmacists would assume that another inhaler bearing that shade of purple (let alone a different shade) emanated from the same trade origin even if they assumed that the inhaler would serve the same function as the Glaxo product. In other words, the survey did not show that people regarded the colour as being distinctive of inhalers produced by Glaxo even if they did associate them with inhalers containing that active ingredient or for treating a particular illness.⁴¹

38. The *Glaxo* case is an interesting example of a dissection of the methodology used but it is unusual in rejecting survey evidence. More commonly potential problems are resolved in the course of the case management hearing at which permission for the use of the survey evidence is sought and obtained. An example of successful survey evidence being deployed to help the court is a delightful case I presided over in the Chancery Division. The Champagne House Louise Roederer which produces the very prestigious and extremely expensive CRISTAL champagne were suing some Spanish wine produces who were selling bottles of cava for £5 at Sainsburys and labelling them “Cristalino”. Two surveys were devised and carried out for Louis Roederer by a consultant who is one of the doyens of this kind of work. He wanted to establish whether the word “Cristal” was associated with premier cuvee champagne by a significant number of people questioned. About 570 people were stopped randomly in the street and simply shown a piece of card with the word CRISTAL written on it and asked “What can you tell me about what you are looking at?”. 14 per cent mentioned champagne seeing the word CRISTAL printed on a card without any context. I held that this showed that the brand had a strong reputation, particularly given that very few of the interviewees were likely actually ever to have bought a bottle. But my point is not just the inherent interest of the exercise but in the close control that the court maintains over the conduct and analysis of these data.

Surveys in Market Investigations

39. A final point of comparison can be found in the scrutiny of survey evidence in market investigations within the context of competition law.
40. An important element in many cases is how to define the relevant product market. That enables you to tell which products available to consumers compete with each other and that in turn tells you whether the seller of any one of those products is free to act in an anti-competitive or monopolistic way without fearing that customers will simply switch their business to one of its competitors. So if an increase in the price of a product, say bananas, would lead consumers to switch their purchases towards other kind of fruit, then the banana producers are not able to exercise monopoly power over the pricing of bananas because there is demand-side substitution. But if there is a significant proportion of consumers who cannot switch to other fruit, we would say that bananas occupy a separate product market because there is little demand side substitution and the producers of bananas may then be in a position to exercise monopoly power.
41. Modelling demand-side substitution provides an ideal opportunity for the use of survey evidence. This is explicitly permitted by both the European Commission⁴² and the Competition and Markets Authority, and the CMA has produced detailed guidance on the design and presentation of consumer survey evidence in merger cases.⁴³ The CMA's guidelines have proven important in practice. A recent example can be found in the failed merger between Tobii and Smartbox, both of whom operate in the field of assistive technology for communication. Although the CMA was not initially notified of the merger, it investigated the merger's competitive effects and ultimately required Tobii to divest itself of the Smartbox business on the basis that, the CMA concluded, the merger could be expected to result in a substantial lessening of competition.
42. In making its case before the CMA, Tobii sought to rely on a survey that it conducted of end users of assistive technology as evidence to support its arguments on demand-side substitution.⁴⁴ The CMA refused to give any weight whatsoever to Tobii's survey and relied instead upon its own customer questionnaires in support of its assessment that, taking account of the likelihood of demand-side substitution, competitors would not provide a sufficient constraint

to mitigate the effects of the merger on competition.⁴⁵

43. When the CMA's decision was challenged in the Competition Appeal Tribunal,⁴⁶ a key point of argument concerned whether it was unreasonable for the CMA to place no weight on Tobii's end user survey; and whether the CMA's own customer questionnaires were so flawed that the CMA could not reasonably rely on them.
44. Dealing first with Tobii's survey, the CAT upheld the CMA's refusal to give it any weight, noting that the CMA had "*provided a number of intelligible and adequate*" reasons why it refused to do so.⁴⁷ These included that: (i) the survey was based on an online panel (in circumstances where the CMA's survey guidance explained that samples for online panels are not random and so are deserving of less evidential weight); (ii) the relevance and size of the sample (which was 101 end users) fell short of the CMA's usual requirement for survey evidence in merger cases; (iii) the response rates on a number of important questions were very low (again, falling short of the CMA's usual requirement); (iv) it was not clear that all of the survey's respondents were in fact customers of Tobii or Smartbox, or even end users of assistive technology for communication; and (v) the questions used in the survey did not provide the necessary data to accurately model the likelihood and scale of demand-side substitution.⁴⁸
45. Turning to the CMA's questionnaires, the CAT looked closely at each of the criticisms and concluded that despite there being some flaws in the CMA's questionnaires, when placed in context, that did not render it unreasonable for the CMA to have relied upon them.⁴⁹
46. Finally, in considering survey evidence it is ironic that the Strasbourg case that is frequently cited as authority for the proposition that statistical evidence is valuable, the *DH* case about the Roma children, was itself one in which the survey evidence relied on was intensely controversial as between the different members of the Court. That was the case about whether Roma children in the Czech Republic were victims of discrimination in being disproportionately placed in schools for children with special educational needs rather than in main stream schools. The applicants relied upon unofficial survey data obtained from questionnaires sent to head teachers at schools in Ostrava (a region of the

Czech Republic). The figures obtained showed that 56% of the total number of pupils placed in special schools in Ostrava were from the Roma community although Roma children represented only 2.26% of the total number of primary-school pupils in Ostrava. A Roma child in Ostrava was 27 times more likely to be placed in a special school than a non-Roma child. The Government argued that the statistical evidence was not sufficiently conclusive as the data had been furnished by the head teachers at the schools and therefore only reflected their subjective opinions. There was no official information on the ethnic origin of the pupils. The Government further considered that the statistics had no informative value without an evaluation of the socio-cultural background of the Roma children, their family situation and their attitude towards education. They pointed out in that connection that the Ostrava region had one of the largest Roma populations in the Czech Republic.

47. The majority in the Court recognised that the statistics may not be entirely reliable but said that they revealed a dominant trend that had been confirmed by the State and other bodies who had looked into the question. The statistics gave rise to a strong presumption of indirect discrimination, shifting the burden of proof to the Government to show that the difference was the result of objective factors unrelated to ethnic origin.

48. The strongly worded dissenting judgment of Judge Jungwiert contained a barrage of other statistics comparing the efforts made by the Czech Republic to bring Roma children into the education system compared with the position in other Contracting States. He noted that the Roma people living in the Czech Republic had almost all arrived there since the Second World War. They were not only largely illiterate and completely uprooted but did not speak the Czech language. Nonetheless whereas very significant numbers of Roma children in other states failed to attend any school at all, almost 100 per cent of Roma children in the Czech Republic were attending school by 1990. The implication of the majority judgment was, he said, that it is probably preferable and less risky for a state to do nothing and to leave things as they are elsewhere, in other words to make no effort to confront the problems with which a large section of the Roma community is faced. In his view, he said, such abstract theoretical reasoning rendered the majority's conclusions wholly unacceptable.

IV. Conclusion

49. In conclusion, the statistical revolution is taking hold inside and outside the courts. The judicial role in scrutinising statistics is becoming increasingly important. In public law cases, courts still adopt a case-by-case approach. However, if practitioners in the field would welcome more consistency and a more structured framework for appraising statistical evidence, their colleagues in the fields of intellectual property and competition law cases might be able to provide a useful starting point.

¹ I am grateful to my Judicial Assistant, Ian Simester, for his invaluable help in preparing this lecture.

² [2003] EWCA Crim 1020.

³ H. Malek, M. Howard, *Phipson on Evidence* (19th Ed., Sweet & Maxwell, 2017) at 34-18.

⁴ *R (on the application of UNISON) v Lord Chancellor* [2017] UKSC 51.

⁵ *Ibid.* at [57]-[59]; [101].

⁶ (2005) 40 EHRR SE22.

⁷ (2008) 47 EHRR 3

⁸ *Ibid.* at [186] et seq.

⁹ (2006) 42 EHRR 43 at [147].

¹⁰ *The Secretary of State for the Home Department v R (on the application of The Joint Council for the Welfare of Immigrants)* [2020] EWCA Civ 542.

¹¹ [2019] EWCA Civ 1112.

¹² *Ibid.* See paragraph [38]; [45]. This was said to have resulted from HMRC using length of service to determine one's pay within a particular grade, with disproportionately more women recruited in recent years.

¹³ Note 10 at [75].

¹⁴ Note 10 at [187]-[188].

¹⁵ (C-167/97) EU:C:1999:60.

¹⁶ *Ibid.* at [62].

¹⁷ Note 11 at [19].

¹⁸ *R v Secretary of State for Employment, ex p. Seymour-Smith and Perez (No.2)* [2000] UKHL 12.

¹⁹ Note 10 at [55]; [71]; [75].

²⁰ [2012] EWHC 1977 (Admin).

²¹ *Ibid.* at [50].

²² [2014] EWCA Civ 69 at [32]. Note no appeal was made against the Court of Appeal's Article 14 analysis, so these issues were not considered by the Supreme Court in its judgment ([2015] UKSC 79).

²³ E.g., *R (on the application of SG & Ors) v Secretary of State for Work and Pensions* [2015] UKSC 16; *R (on the application of SC & Ors) v The Secretary of State for Work and Pensions & Others* [2019] EWCA Civ 615. *SC* is currently on appeal to the Supreme Court.

²⁴ *Ibid.*, *SG* at [61] (per Lord Reed); [180]-[182] (per Lady Hale). See *SC* at [126]. The logic behind this is simple: roughly 90% of single parents are women, whilst roughly 50% of joint parents are women.

²⁵ *Essop and others v Home Office (UK Border Agency), Naeem v Secretary of State for Justice* [2017] UKSC 27.

²⁶ [2020] EWCA Civ 130; [2020] 1 WLR 2062.

²⁷ *Ibid.* at [27]; [64].

²⁸ *Ibid.* at [93].

²⁹ Note 10 at [73].

³⁰ In fact, the use of surveys was so commonplace that, in 2015, a publication by Bird & Bird observed that, for a period between the early 1990s through to 2013, it would have been considered unusual for there not to have been some sort of survey evidence in cases where these arguments were being made: see U. Lüken, P. Brownlow, N. Aries (2005), *How courts view surveys in trade mark cases*, accessible [here](#).

³¹ *Glaxo Wellcome UK Ltd v Sandoz Ltd* [2019] EWHC 2545 (Ch) at [218].

³² *Interflora Inc v Marks & Spencer plc* [2012] EWCA Civ 1501.

³³ *Ibid.* at [149]-[151].

³⁴ It is generally understood that, at this stage, the Whitford Guidelines go to the weight to be given to the survey, though they may also go to admissibility: see Note 31.

³⁵ [1984] RPC 293 at 302-303.

³⁶ Note 32 at [61].

³⁷ Note 31.

³⁸ Permission was granted by Birss J: [2017] EWHC 3196 (Ch).

³⁹ Note 31 at [200]-[246].

⁴⁰ Ibid. at [214].

⁴¹ Ibid. at [245]-[246].

⁴² Commission Notice on the definition of the relevant market for the purposes of Community competition law (OJ 1997 C372/3)

⁴³ Accessible at: <https://www.gov.uk/government/publications/mergers-consumer-survey-evidence-design-and-presentation/good-practice-in-the-design-and-presentation-of-customer-survey-evidence-in-merger-cases>

⁴⁴ [2020] CAT 1 at [85(7)(i)]

⁴⁵ See, e.g., the CMA's Final Report (dated 15 August 2019) at [24]. The report is accessible [here](#).

⁴⁶ Note 44.

⁴⁷ Ibid. at [203]-[204].

⁴⁸ Ibid. at [202]

⁴⁹ Ibid. at [230].