UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper A11/01

Applications and Policy

Statistics In Society:

Three Case Studies In The UK

Tom King

Abstract

ABSTRACT

Public statistics are increasingly produced to facilitate public accountability and engagement at various levels of society i.e. to empower citizens. However, the impact of these statistics and their comprehension has not been systematically studied, indeed the arena in which statisticians work is generally more technical and less oriented to public scrutiny. We explore the role of statistical literacy in citizenship and provide explanations as to how this is necessary to live up to normative ideals. Case studies show how statistics can create social reality, both in terms of categories and numerical summaries, although there may be more than one 'data story' derived from the same information, as seen in the case of Mid Staffordshire Foundation Health Trust. Furthermore, public interpretations of these data stories need not be accurate and are rarely appropriate in their criticism of the objectivity represented by statistical analysis. Statistical description for a public audience requires much more than reporting figures relating to the phenomenon being described: such a presentation can be misleading or exclusive as well as restrict political discourse.

KEY WORDS: descriptive statistics, citizenship, public understanding accountability, statistical literacy, data stories

Statistics in Society:
Three case studies in the UK

Tom KING

Division of Social Statistics and Southampton Statistical Sciences Research Institute

University of Southampton

tom.king@soton.ac.uk

ABSTRACT

Public statistics are increasingly produced to facilitate public accountability and engagement at various levels of society i.e. to empower citizens. However, the impact of these statistics and their comprehension has not been systematically studied, indeed the arena in which statisticians work is generally more technical and less oriented to public scrutiny. We explore the role of statistical literacy in citizenship and provide explanations as to how this is necessary to live up to normative ideals. Case studies show how statistics can create social reality, both in terms of categories and numerical summaries, although there may be more than one 'data story' derived from the same information, as seen in the case of Mid Staffordshire Foundation Health Trust. Furthermore, public interpretations of these data stories need not be accurate and are rarely appropriate in their criticism of the objectivity represented by statistical analysis. Statistical description for a public audience requires much more than reporting figures relating to the phenomenon being described: such a presentation can be misleading or exclusive as well as restrict political discourse.

KEY WORDS: *descriptive statistics, citizenship, public understanding accountability, statistical literacy, data stories*

Introduction

Statistics have been an issue of public concern for many decades, not least because they can be misunderstood or misinterpreted (Percival 1919). Many are produced and reported but little is known about what use they are put to by the public (as opposed to specific user or interest groups), how well they are understood, or whether they are considered to be valuable. Statistics can attain extraordinary status, coming to symbolise an event in a way that obscures other detail, such as the number of deaths in the wars in the Congo (Coghlan, Brennan et al. 2006). De Santos (2009) describes this status as totemic and shows how it is present the lives of people subject to this description but this does not probe into the accuracy or appropriateness of the statistics used. More precisely, we know that statistics are always trying to measure or to represent something (Hand 1996) and so when they measure something of civic importance this can attain the status of an objective fact even though they are nothing of the sort.

In order to look further into the publics' understandings of statistics and how this may be evolving, as there is precious little prior knowledge, we adopt a case study approach to consider certain public statistics and how these are used. Specifically we consider some examples, and then study three cases in depth. These allow us to develop definition of the pivotal issues around public statistics and then focus in on how these issues are evinced in our cases. We are then able to draw out some conclusions on the nature of statistics which might be useful in the purposes to which they are intended to be employed. Our cases are of different types, so this is not a multiple case study approach so much as three separate case studies put together.

Our paper reviews the context in which statistics exist along with some historical review of statistics in public life. This gives us particular focus from which to draw the cases and the detail of their situation. We then consider how public understanding has been constructed in education and in empirical studies to consolidate the existing conceptions. This allows us to describe the role statistics play in public life and draw out some examples of how they are used, particularly in governance and accountability. This will elucidate the issue of objectivity, thought so important in science, which is generally only an aspiration to avoid certain undesirable practices in society (Porter 1995). We follow this with a description of our case studies, and our

discussion of what makes a useful statistic for public audiences. Finally we advance some recommendations about how statistics could be more valuable to society and how statistical literacy would empower citizens who desire accountability and understanding. Thus this is not about the best production or statistical practice so much as the citizen user viewpoint on the utility of statistics.

History of Public Statistics

The idea that statistics can be used in different ways to develop and evaluate policy, to facilitate public accountability and to inform publics about performance will be familiar to many in industrialised countries (see Goldstein and Spiegelhalter 1996; Evans and Plows 2007). How this situation has arisen, and the variability of this process in different fields and cultures, is reviewed by Derosières (1998), revealing some specific disparities. In particular he notes three separate aims of original developments: problem-focused analyses of self-contained issues; equality-focused centralisation of data for distribution of resources; and description-focused collation of information for understanding of societal variation. Thus statistics has developed with different meanings in different contexts, deriving from several epistemic stand points. However, there are commonalities in the desire for an objective public presentation (Porter 1995).

Derosières (1998) identifies the post-war period as being the beginning of the common use of statistics in public life, such that they are now commonly used in decision-making. Standardisation of numbers in professional contexts met with much resistance about judgment in many fields, delaying their use considerably, but also developing a crucial awareness of the issues involved (Porter 1995). In many countries, including the UK, performance indicators have been developed in relation to public bodies and delivery of services (Bird, Cox et al. 2005). Such indicators have been identified as useable in two distinct manners: facilitating comparison of institutions and thence choice between them; and understanding differences in performance as a tool for those responsible, thus facilitating accountability to some extent (Goldstein and Spiegelhalter 1996). However, it is clear that the publication of such indicators in ‘league tables’ plays a slightly different role, indeed the avoidance of stigma is now a concern (Lilford, Mohammed et al. 2004), with good reason as shown by the example of country risk in Argentina (de Santos 2009). Hence in little

more than a generation, statistics have appeared to take a strong role in the public sector and civic society by providing a more objective accountability, without pausing to ensure that citizens are keeping up or even gaining their assent.

The problem that publics may face understanding statistical outputs and processes has been raised several times in the past two decades, perhaps first defined in Wallman's presidential address to the American Statistical Association (Wallman 1993). Indeed, these concerns are often raised in the reflections of senior statisticians in this way, and also in some texts by experienced teachers (e.g. Chatfield 1995; Barnett 1999) who emphasise problem solving and understanding over the mechanistic calculations which can become the focus of statistics education through the prejudices of both teachers and learners (Petocz and Reid 2010). This situation would make the lack of research by statisticians into public understanding surprising, were it not for several issues: the statistical establishment belongs to the problem-focused tradition, which is engaged with industry and other fields of application rather than the wider world, and this has meant little need to engage with the public directly. Statistics used by government have been similar to the descriptive political arithmetic of past centuries which present large tables of data without estimating models or making inferences with official statistics being quite separate from academic branches; the exception to this, economics, has developed as a separate field in its own right (Desrosières 1998). However, recent developments in performance indicators, risk analysis and climate sciences have moved statistical issues into the public domain, and perhaps none more so than in the UK.

Statisticians often do not address the issue of communicating with the public, but other groups must do this as a matter of course and so have researched some concerns about communicating statistics in their respective fields. Crettaz von Roten (2006) identifies the public understanding of statistics as a major barrier to the understanding of scientific concepts, sometimes hampered by the statistical skills of scientists themselves. Often there has been a problem of quality control (Percival 1919; Andersen 1990) so critical reading is an essential component of using statistics (Wallman 1993; Gal 2002). Informed consent is a major concern for individual doctors and patients, so the communication of risks and likely effects of treatments in a way which empowers patients is very important and one study has found graphical

presentation to be the most useful (Goodyear-Smith, Arroll et al. 2008). The media are often criticised by statisticians for misleading presentation of risk but one study shows that the presentation of information does not necessarily improve understanding of risk (Miller and Barnett 2010). There may well be more studies like these but the lack of a coherent literature on the topic has meant even this has been difficult to track down. However even these few build up a clear picture of the importance of statistics in communicating issues of public interest and associated concern from professionals of getting their message across. This present work goes a bit further in building up a clearer picture in the specific context of the UK.

The UK Context

Statistics can be relevant to any person in any society but we are going to focus on the politically constructed uses that they have to play in society and the consequences of these constructions. For these purposes, it is useful to focus just on one country where the statistical infrastructure is well developed: that of the UK. Our immediate interest here is in what is expected of citizens in various roles, and what their capabilities are. The remainder of this section reviews the current work on the former but firstly we note the issue of statistics curricula which would underpin public understanding.

Holmes (2003) identifies an important lack in teaching of statistics in England (not so different from the rest of the UK in this respect) in the 20th century, with curriculum being available but being utilised by individual teachers rather than systematic provision. The aims of those projects which were implemented were about the needs of workers, as well as trying to develop teaching so it gave an understanding of inference in a broad sense, rather than solely facility with techniques. Thus they did identify that project work was necessary for understanding of the purpose of statistics but even into the 80s and 90s this was rarely done, so good school education in statistics will be increasing in coverage in the general population but far from uniform. This wide variety in ability by age and education we will look for in its effects on citizen engagement.

The concept of statistical literacy is fairly novel; indeed the use of statistics in public decision-making in the UK evolved only in the 1960s (Desrosières 1998). Public data and reporting are standardised in many organisation, and government statistics are

monitored by an independent body (the UK Statistics Authority (UKSA)). Now we have an emphasis on evidence-based policy (Sowden and Raine 2008), citizen choice and accountability; all underpinned by quantitative data about performance (Bird, Cox et al. 2005). Sadly, none of the implementations of such public schemes has involved assessments of citizen ability to engage with such information, although some specific effects are now being studied (e.g. Boyne, James et al. 2009). However, from the preceding, it is apparent that there will be a bias by age (due to these changes taking effect over time, and so applying to different generations of school children) as well as the level of education of citizens as to how well they utilise public data. Thus we turn to the obvious question of how this may create inequality and failings in public life where expectations of citizens reach beyond their complicity and their capabilities which is a problem of both policy and engagement.

Citizens of the UK may take on various roles in public life: making choices for themselves and their dependents; challenging prevailing systems of governance; and accepting positions of public responsibility. These roles may be predicated by their locations, proximity, impartiality, expertise or popularity but there is a movement to make relevant information accessible to any such group and strengthen the voice of such citizen users in the determination of data collection (UKSA 2010). The intention of making data publicly available is to empower ordinary citizens, and may have the benefit of making debate more representative early on, a problem seen as important in the public understanding of science (Evans and Plows 2007). Public data allows individuals to judge from their own expertise rather than requiring professional accreditation or networks. However, most statistical work focuses on the validity specific uses of specific data, as well as issues of design, so that understanding of public use is often desired but never addressed (e.g. Bird, Cox et al. 2005). Furthermore, citizens may not agree with the validity or utility of political quantification (Porter 1995), partly because of the complexity of its constructions. Evaluation of to what extent citizens find data useful or empowering is very limited but it is clear that some performance indicators do generate thresholds of acceptable performance.

Statistics are entering further into the public domain for several reasons: science is more dependent on statistical models of large systems e.g. global climate; more

complex economic policy (which increasingly relates to forecasts and international indices of credit worthiness rather than manufacturing output domestically); and scientific research (which has a negotiation of public consent role in crisis management such as swine 'flu rather than advising or researching for government). Public contact with, and expectation to see, figures is increasing but levels of confidence in official statistics continues to disappoint the government. This stands against a backdrop of misrepresentation of data by government, media and other agencies. Indeed further investigation into many peer-reviewed publications leads to doubts of the veracity of calculations and methods (Andersen 1990). Thus statistics are becoming increasingly important to democratic functions but do not seem to be reaching the potential they are thought to have and they can serve to exclude some groups from participation (Demeritt 2001; Evans and Plows 2007) and create arbitrary objectives of normalcy (Llewellyn and Northcott 2005). We need to establish some of the reasons for this, to clarify the expectations and evaluate whether they are realistic, and the quality of the implementation.

Government has introduced monitoring of performance of public bodies by various indicators in recent years (Bird, Cox et al. 2005). The most prominent of these relate to schools and hospitals, considering exam results and death rates respectively (Goldstein and Spiegelhalter 1996). Statistical models put both of these in context by allowing for some aspects of local circumstance to be taken into account so that these measures can be regarded as relating to institutional performance. Such adjustments are considered to represent an improvement over the use of crude rates of productivity or output, like mortality and exam passes. The values are then made publicly available so that they can assist citizens in making choices between institutions and services. However, they also play an important part in local accountability as media present 'league tables' of institutions, and local and national averages, to compare the performances against these indicators for other groups.

How the external use of data affects the position of those formally accountable, i.e. the governing bodies and executives of institutions, has not really been considered but is potentially damaging (Ghobadian, Viney et al. 2009). The rapidity of implementation and lack of the technical barriers to working with such performance data have presented an enormous challenge for those involved in governance,

particularly as governors are lay citizens. This is not withstanding the ongoing debates in statistical circles about the appropriate nature of comparisons made and the use of the same information as performance indicator and audit tool and judgement (Lilford, Mohammed et al. 2004). The role of governing bodies in relation to such statistics is multi-faceted: representing them to the media; challenging inaccuracies of external agencies; and challenging failures of internal performance.

Defining Public Understanding and Literacy in the context of Statistics

In this section we make reference to science as it offers a useful precedent as well as having a strong reliance on statistics in matters of public interest. Statistical literacy is often constructed as a problem of education (Gal 2002; Utts 2003), similar to the deficit model applied in the public understanding of science. This model constructs scientific understanding as a problem of scientific instruction and knowledge more than understanding and critical thinking (although it must be noted that such a simple model has been criticised (Sturgis and Allum 2004)). However, science is taught in schools and often thought to contain domain specific knowledge as well as understanding of scientific methods (Evans and Plows 2007). Statistical literacy is dependent on general literacy and basic numeracy, combining these with critical thinking about problems in context to consider their statistical validity (Gal 2002). Because an important action of university education is to develop critical facility, albeit in a particular domain, the emphasis on numerical critical thinking makes it unsurprising that most efforts at statistical literacy are focused on education in universities, even setting aside the teaching of probability theory or statistical inference. Moreover, as in scientific understanding (Sturgis and Allum 2004), statistical understanding is about context and problems, not just learned concepts and techniques (Petocz and Reid 2010; Pfannkuch, Regan et al. 2010). However, there are currently plans to introduce more advanced statistics and to deliver this curriculum to all students in the UK and New Zealand.

Historical reading suggests that the use of problem-focused solutions negotiated solely between chosen experts and those commissioning was typical of the UK engagement with statistics in public life (Desrosières 1998). Indeed such systems of statistical accountability are more advanced in the UK than elsewhere in the world, meaning opportunities for research on public understanding abound in the UK context.

There are several prominent opportunities to examine these issues in detail: public and political choices in relation to scientific issues of particular importance such as climate change and epidemic management; citizen roles in relation to the issues of comparison generated by government statistics on issues such as schools and hospitals; and the skills expected of individuals taking on public responsibilities and the extent to which external monitoring will facilitate or obstruct their role. These issues are all in a reflexive political system which is developing at many levels which research may itself influence so that a snapshot is not sufficient evidence to base any change on. This is concordant with the historical reading which shows how accountability emerged at the same time as policy: measurement comparing institutions does not occur until equality of provision is a political aim (Desrosières 1998).

Concerns about public understanding and literacy with regard to science began in the 1980s but still leave some inconclusive points when we go beyond systems for education (Evans and Plows 2007). The roles of citizens in making decisions about science policy (Feyerabend 1982) and how their understanding can be facilitated is challenging (Evans and Plows 2007), although with respect to this, it is interesting that statistical understanding has been suggested as a particular barrier (Crettaz von Roten 2006). Furthermore, boosting public understanding of science by remedying a knowledge deficit has been criticised without leading to conclusions. This has facilitated the growth of mature scepticism of scientific endeavour in the public domain (Sturgis and Allum 2004). Statistics remains a topic similar to science in that it is sometimes featured in public affairs as a justification for a decision but this feature is inaccessible to those citizens and journalists without statistical literacy as it is not part of the subject domain. That said, ideas of literacy are rather more nascent in statistical circles (Gal 2002), although here there is the more public-focused drive of National Statistical Offices (NSOs) which need to give public access and understanding in their work. From a government perspective, the challenge of reconciling the judgment of professionals with open public disclosure has been a project of piecemeal progress for decades (Porter 1995).

Just as science is not a single homogeneous domain, applied (empirical, as opposed to theoretical) statistics contains different specialisms, with experimental design being very different to public health epidemiology and sociological statistics. Complex

statistical modelling and simulation, particularly in a Bayesian framework, used in fields like climate science might be considered different again. In the English tradition, most statistical work consists of expert consultation on specific problems (Desrosières 1998), so the interpretation remains within a specialised field and most statisticians work in specialised industrial or academic areas. More complex problems which include many of a more social nature and wider-ranging public interest are becoming accessible to modern statistical techniques but their public representation is unfamiliar to both statisticians *and* citizens.

The difficulties of understanding complex problems are further complicated by a shortage of those qualifying at a high level in statistics (Smith and Staetsky 2007) and the use of statistical analyses by many scientists who may not have sufficient training in matters of statistical inference. Thus, despite there being much attention on improving public understanding of statistics, there are still clear shortfalls in not only engagement with this project but also a lack of those with the ability to engage. At this point, we venture a speculation that there is a problem in conceiving the breadth of statistical work and the range of its understanding, the latter being well expressed by Petocz and Reid (2010), and that there is indeed a need for a 'public understanding of statistics' (Crettaz von Roten 2006) to help address this range with a public perspective. This public understanding is about group constructions of the epistemic which may not be well-informed or hampered by a lack of data.

Public Statistics

The notions of public understanding and literacy in statistics need not be separate but it is useful to present a distinction between formal hierarchical training and critical thinking. There is also a separation of issues for those presenting, in terms of structures (Crettaz von Roten 2006), and those seeking to read for themselves, the agency of individuals and groups (Evans and Plows 2007). Gal (2002) conceives statistical literacy as joining up ideas of literacy and numeracy in context with an understanding of types of uncertainty and probability. These last two are generally taught to students in the hierarchical mode of knowledge development common in mathematics, allowing them to be applied to real problems as appreciation advances and a cyclic model of planning, implementation and reflection is encouraged. Gelman

and Nolan (2002) discuss the idea of teaching statistical literacy to students of statistics only in the context of reading and criticising published work.

Statistics will be taught and can be applied in many other fields by non-statisticians who do not develop their facility in such a hierarchical fashion but by problem-focused or incidental study, and service courses aimed at enabling certain activities. There are others still who will meet statistical work only as it is presented to them, rather than analysing data themselves (Evans and Rappaport 1999) and these people may carry knowledge and prejudices which develop over time through their experience. Moreover, the models produced by statisticians are constructed not begotten and so we must retain a social perspective on statistics (Hennig 2010). Indeed the experience of everyone affects their understanding and develops them in domains not quite the same as those they start from and it is the stories that emerge from the context and the data which are important.

Petocz and Reid (2010) have developed a typology of statistical perspectives which conceptualise its role, and its objectivity (in the sense of how impersonal it is (Porter 1995)), rather than the complexity of its application. They apply this to teaching and learning situations and also develop a hierarchical structure but one showing the difference between broader and narrower views. Thus they range from a focus on techniques, through a focus on working with data, to a focus on understanding; giving a range of freedom in approaches to problem solving. As well as presenting a challenge to the conventional idea of upskilling, perhaps broadening perspectives may be more fruitful in many cases. Indeed several statisticians have written, late in their career (e.g. Chatfield 1995; Barnett 1999) or in presidential addresses, about the importance of statistics being about problem-solving and data rather than simply following a process.

Thus we can develop several different roles for statistical literacy: being able to follow through one's own needs as far as to make inferences; being able to criticise the statistical work of others; being able to understand the limitations and implications of statistical analyses. Gal's (2002) conception of a need to understand probability and uncertainty is not essential to these in all cases, indeed public understanding (Crettaz von Roten 2006) is certainly relevant to the last two. The ideas of Petocz and

Reid (2010) are certainly important to the last but this hinges on a conception of why statistics are, which may not be a consensus or considered. Thus we will take a more personal definition to use, the utility of which will become apparent later, that statistical literacy is 'familiarity with reading and writing "data stories"' (Pfannkuch, Regan et al. 2010).

Consideration of perspectives may give an insight into public perceptions and prejudices about statistical work.

> "The aim of statistical work is to make a priori separate things hold together, thus lending reality and consistency to larger, more complex objects." (Desrosières 1998)

Many people have misconceptions about quantitative methods of summarising things and many mistakes are made by people using statistical methods. The epistemological foundations are actually consistent and defensible but have emerged from several different historical traditions (Desrosières 1998). There is genuine confusion about the intentions of statistical work and teaching and other communication can actually limit understanding by design (Petocz and Reid 2010).

> "a richly nuanced or profound analysis of a large question is never logically excluded by the attempt to quantify parts of it." (Porter 1995)

Thus we can see that the objectivity of statistics is the major issue in the public understanding of statistics, where we mean that the method is replicable rather than saying the result is truth.

Citizenship and Statistics

Exactly what constitutes citizenship is difficult to fix, partly because it is negotiated and therefore evolving, and partly because a definition tends to hinge on exclusions and activities rather than something holistic. The intention here is to explore this empirically rather than theoretically but we will consider Bellamy's (2008) specification of three strands, namely inclusion, rights and involvement. These

function under the definition of citizenship as a state of "civic equality", once included i.e. equal rights and equal involvement. The interest is therefore in how statistics affect the involvement and rights of citizens in the UK.

One of the main empirical interests of political theorists with regard to citizenship is in voting behaviour (Bellamy 2008). Performance indicators may support the judgment of citizens about the suitability of others to govern them, but this requires the individuals to understand the indicators. Empirical work has shown that very poor performance is punished under such schemes but good performance may not be rewarded (Boyne, James et al. 2009). It is possible that this effect is caused by publicity but the study was unable to show this thus more research is required. As regards statistical literacy, simple, descriptive rating systems have been produced to facilitate maximal understanding from the public but are contentious in how they may be gamed or misrepresent institutions, as well as unfairly damaging their reputations (Lilford, Mohammed et al. 2004).

The issue of equality of involvement is of empirical interest because of declining levels of participation, particularly among certain groups. However, equality is about fair representation of your views, not necessarily a compulsion to represent them oneself. Thus surveys and polling might be considered a modern form of participation which statistical offices attempt to make representative and, in the UK at least, relevant to the public (UKSA 2010). The reporting of statistics is certainly a part of popular understanding of issues (Crettaz von Roten 2006; Miller and Barnett 2010), but access to these statistics through statistical literacy requires critical understanding of the individual readers and may also require them to collect their own information in order to engage in the debate (Evans and Plows 2007). Furthermore, consultation with users and other stakeholders is expected, even of the smallest groups accepting public funds, which may impose substantial demands on groups if they are to consider each individual equally.

The matter of rights of equality of access to the benefits of citizenship is part of the cause of the statistical performance indicators we see today: in order for citizens to make choices, they must be statistically literate to the extent that they can compare different institutions and do so objectively. The premise of locally accountable boards

representing performance objectively is in conflict with nationally determined policies and indicators for improvement. However, this requires an understanding of local context which may not be possible in a national indicator. We return to the issue of local use of performance indicators later but the concept of governance in relation to statistics will be developed further in the next section.

Hence the level of statistical literacy required to participate directly as a citizen (and the alternatives available to those who are not) is of importance in various ways. In particular the lack of understanding of uncertainty is limiting for many people: e.g. there are considerable difficulties appreciating risks, in terms of traffic (Evans 2003), climate (Demeritt 2001) and nuclear power but we have professionals to mediate this uncertainty. Statistical literacy is unlikely to be uniformly distributed in the population. Thus we are led to observe that it seems inadequate levels of statistical literacy cause disenfranchisement due to these statistical means for participation in democratic activities. This is complicated the fact that it is a state concern to have responsibility for improving the quality of life of all citizens, in part by educating them, so that the abilities of and demands on citizens are constantly developing.

There is a further problem engendered by the structures to assist citizens in their rights, namely that these structures exclude other forms of action. Indeed, in Britain,

> "opponents of airports and powerplants may be denied the opportunity to challenge government studies unless they can match the official experts and produce comparably full and detailed analyses of their own." (Porter 1995)

However, there is potential for interest groups to organise themselves (Evans and Plows 2007), for there to be ordinary people 'barefoot statisticians' who are capable enough the engage with these dialogues (Evans and Rappaport 1999), and official engagement is concerned to involve users in the whole of the data cycle (UKSA 2010). Nevertheless, statistical structures have the potential to regulate society to exclusion of desirable freedoms (Porter 1995; Desrosières 1998).

Governance

The concept of lay oversight of public work is well-established in the UK, to the extent of having a place in determining the direction for scientific research funding (Feyerabend 1982; Evans and Plows 2007). Lay representation will be familiar as the trustees of entities such as pension funds, as well as the governing boards of police forces and health trusts. Much lower level organisations, such as schools, local charities and community groups will also have a governance structure with outside or community representation. The governance responsibility of such bodies includes financial oversight and senior appointments but is becoming more focused on setting targets, monitoring performance, adjudicating value for money, and managing risk. Each of these developing tasks requires a level of confidence with statistical concepts and an ability to ask questions about data. Governing bodies, as we shall refer to all of these boards, are empowered to seek expert advice, or appoint persons to their number with skills in certain areas but the actions they take are expected to be based on their own consensus decisions. They should also expect to represent themselves to the community they serve and the sources of their funding. Thus they need to engage a number of different audiences with the performance of their organisation.

The variety of tasks and their importance to the organisation mean that public appointments require a wide range of specific abilities and high level experience. Financial expertise is often required as many bodies require a chartered accountant to act as the independent chair of their audit committee. High level experience in HR, media or PR, IT, and legal work is also often required as well as board level experience due to the strategic oversight nature of the role. Facility in statistical issues is not something which is mentioned yet, as the case of ambulance trusts in England shows (Bevan and Hamblin 2009), this is something which is important.

Since the year 2000, ambulance trusts in England have been required to meet targets for response times to critical (category A) calls within a clinically indicated time of seven minutes. Responses to this have included fabrication of data, post-arrival re-categorisation and delayed start times, as well as improved systems for responding, all of which have an effect on the number of responses meeting the target. Some methods of gaming the system (as such responses have been called) are fairly obvious from simple plots of time frequencies, and an investigation by the Healthcare Commission

was able to take action where this behaviour was particularly egregious. Sadly, although it is not clear if the governing bodies responsible seen such data, but they certainly had not identified concerns and most likely did not see it. This demonstrates the possible effect of public performance management to distort priorities away from improving performance towards meeting targets (Lilford, Mohammed et al. 2004; Ghobadian, Viney et al. 2009). This case also demonstrates the need for governing bodies not only to ask questions of data presented to them but to have appropriate skills to be able to agree appropriate data to use to monitor organisational performance by reading and writing their own data stories.

It has already been noted that performance indicators could distort accountability structures (Ghobadian, Viney et al. 2009) but they can also provide perverse incentives. These may include overproducing unnecessary products or organisations selecting only those cases likely to succeed such as healthy patients or advantaged students. Thus the RSS (Royal Statistical Society) responded with a working party on the appropriate use of performance indicators which reported many recommendations, including allowing for case-mix in an appropriate manner, and piloting the system before it becomes a statutory accountability, allowing flaws and perversity to be identified (Bird, Cox et al. 2005). The outcome of this has been much more complex methods of measuring performance, although government has been attached to the idea of league tables which are clearly not identifiable from the data available (Bird, Cox et al. 2005). It is presumed that such systems will have a beneficial effect (Hodgson, Farrell et al. 2007) but this requires a better understanding of the nature of decisions that would be based on such indicators. Possible distortion of priorities comes down to whether we are measuring the whole performance and if the data are appropriate for this so for schools new models and concerns are evident (see later).

When statistical work has been done with performance indicators, it has limited their uses and placed constraints on their construction but specifically it has specified the terms of comparisons which can be made. Goldstein and Spiegelhalter (1996) identify two types of decisions to be made using such indicators, namely identification of factors affecting performance in the whole population, and comparison between a small number of organisations, e.g. for the purpose of making a choice between them. Thus they identify the need to correctly specify confidence intervals depending on the

comparisons being made. Bird et al. (2005) extend this to three aims: research, managerial and democratic. These have the advantage of focusing on the people roles, rather than the specific comparisons being made. Thus one can take the perspective of: trying to understand what the indicator tells us about the whole picture; identifying variations in performance which are unusually high and making a response to this at an appropriate level; or evaluating the performance of, as examples, a person, an administration or a policy. How these three roles interact and may conflict when individuals have unclear responsibilities or artificially constrained responsibilities is something we shall develop further in this paper. The issue of whether comparison to an average which may not mean anything in any intuitive sense but may be a rather misleading goal is a moot point (Llewellyn and Northcott 2005).

Mid-Staffordshire Foundation Health Trust

An independent inquiry was launched into the reasons for the continuing poor performance at Mid Staffs hospitals trust in 2009 due to findings of the Healthcare Commission that standards of care were low and action had not been taken and that which had had not been successful. Many of the problems encountered there had nothing to do with statistics and so the discussion here is not causal but it is of interest that of eight sections of the report (Francis 2010), two were ‘mortality statistics’, ‘governance’ and ‘the board’. Moreover, one of the issues which stimulated concern about the standards of care in the trust was the report that “between 400 and 1,200 needless and unnecessary deaths” were associated with the poor management of the trust. Thus the issue of statistics and responses to them were of significant interest in identifying the underlying issues which contributed to continuing poor experiences for some patients treated by the trust as these were raised at several points prior to the inquiry.

Specific concerns raised in the inquiry were about the treatment of the statistics by those in charge, and the more general issue for the public of engaging with and having confidence in the mortality statistics. The report of the independent inquiry makes this conclusion:

> “Too much comfort was taken from the coding as an explanation for concerning figures and insufficient consideration was given to other explanations.”

And the following recommendation:

> "In view of the uncertainties surrounding the use of comparative mortality statistics in assessing hospital performance and the understanding of the term 'excess' deaths, an independent working group should be set up by the Department of Health to examine and report on the methodologies in use. It should make recommendations as to how such mortality statistics should be collected, analysed and published, both to promote public confidence and understanding of the process, and to assist hospitals in using such statistics as a prompt to examine particular areas of patient care."

The inquiry went so far as to take advice from outside the UK about the mortality statistics in order to maintain an appearance of independence in this endeavour. Thus our analysis may well not be impartial and so instead it will look at the conflict between the different options for those responsible and uses to which the data were put.

Warnings were received about high levels of comparative mortality in the trust and the action taken was to assess whether the case-mix was being appropriately judged due to the different recording of primary and secondary diagnoses. Thus, of the three options governing bodies have in this instance, they took only one, to challenge the data, rather than interrogating it or challenging the management. Challenging the management to improve the comparative mortality might have been a sensible response alongside this but the issue of asking for further analyses at a high level, or asking questions of what was presented was not really taken and is not really covered in the inquiry. To a statistician this seems strange, but that is precisely because this action would require some statistical competence on the part of those involved.

The warnings received by the trust were not simply about high overall mortality, which was reported in the papers, but about high mortality in relation to specific conditions. Thus there are natural questions to ask about the variation in comparative mortality which should be plausibly explained by the issue of coding secondary diagnoses before this is accepted as the only response necessary to the warnings. While it might appear that high mortality couldn't be hidden by moving the diagnoses around, it should be remembered that the trust did not treat all possible maladies, and also that Simpson's paradox could

occur in such a situation. Thus the answers to these questions could be conceptually rather complex but the need to ask them remains, indeed it is the function of lay accountability that once asked, the questioners should be satisfied without recourse to acceptance of management opinion (c.f. Nolan principles). We can conclude that such warnings require certain levels of statistical understanding of individuals which, currently, do not form part of the criteria for their appointment although individuals with high level experience may well have used such indicators.

Part of the government strategy to make public services accountable is to publish performance indicators, although its preferred method of the league table has been challenged (Bird, Cox et al. 2005). Concern has been raised over the way such publication can distort existing relationships of accountability (Ghobadian, Viney et al. 2009) and evidence is that poor performance is picked up by the public whereas good performance goes unnoticed (Boyne, James et al. 2009). Further to this, it has been asserted that a system of accountability must inflict reputation damage for poor performance in order to be effective, and that if nobody is trying to game the system then it is not challenging enough (Bevan and Hamblin 2009). This makes things much more complicated for those trying to evaluate the performance of public officials and for those in formal accountability roles who may have to deal with internal gaming, as well as managing the effect on external relations. However, it should be clear that those attempting to game the system are aware their numbers reflect poorly on them, they need appropriate oversight to improve performance rather than manipulating the indicator.

The Inquiry found deep anxiety amongst the board members that deaths were being directly ascribed to mismanagement, a conclusion which indicators are not able to make as the average is not a baseline expectation. There was a further concern amongst members of the public about whether particular deaths could be amongst this number. The Inquiry did find that the numbers didn't seem to have a basis obvious in the data, and that they were often ascribed to the wrong (and much shorter) time periods. The issue of comparisons, which are fundamental to such performance indicators, are not developed in the report. In particular, comparative mortality compares to the prevailing average nationally in that year, adjusted for case-mix etc. This does not assert that any deaths are necessary or that any number is the right number, indeed it is possible to identify over-performing hospitals where mortality is surprisingly low. The point, which was not made, was that this allows comparison, and that confidence intervals are provided so that

random variation and the imperfect fit of the model can be considered for the variation between trusts rather than ascribing performance as the only factor, the view that was generally represented by the media, albeit occasionally as a deterministic range. Representation of this complex issue requires statistical skills of people in the organisation and of those in the media as well as amongst the public. The Inquiry's recommendation illustrates that there is no realistic mechanism for this to have occurred.

We can see several data stories arising from the standardised mortality ratios provided by Monitor and the Healthcare Commission: Their inferences were that mortality was significantly above what might be expected nationally, both for the whole hospital and for certain specific conditions. Managers presented another story, challenging the objectivity of the methodology employed, reasoning that this was about the coding of the deaths, rather than poor outcomes. The media came to the story about 'excess' deaths, interpreted as caused by the hospital, either through negligence or other errors (the original report of 400 to 1,200 can be traced to a leaked draft which was never actually published http://www.guardian.co.uk/theguardian/2010/nov/16/corrections-clarifications). However, what is clear from the report is that none of these reports were well understood by the public: they reported concerns about identifying if specific deaths were amongst the 'excess' (not something which could be possible from simple monitoring systems) and; they were not worried by the concerns so much as that they wanted the hospital to be good in the future (perhaps this relates to needing a good hospital, but not needing it until one is ill). Thus we can describe a failure of statistical literacy of citizens who were unable to read and write data stories for themselves, becoming victims of the totemic status of the SMR, without being able to engage with it (c.f. de Santos 2009).

School Performance in England

In contemporary England, most stories about schools are stories about data in one way or another: the attendance, behaviour, drug-taking of students; the progression or exam performance of students; the shortage or qualifications of teachers; success or failure of a new policy; the finances and building programmes and; the standards and challenge of the curriculum. These are often compared between schools, between countries, or as a time series over a number of years; the large number of pupils and institutions involved making data essential for large scale validity and interpretation of the whole system. Thus we can try to understand the whole public dialogue about school performance by considering the data stories that arise and the attempts to make inferences with them (c.f.

Pfannkuch, Regan et al. 2010), and we shall do so by taking a lay citizen perspective rather than a statistical expert one (as followed by the Royal Society (2008) in considering these matters). Thus we are analysisng representations of belief, and can hardly claim to be impartial, let alone divining objective truth.

The citizen interest in education has several aspects: as it is publicly provided through public taxation, there is an expectation of quality and accountability. There is universal provision and so all citizens responsible for children have to make choices about schooling, both about enrolment and challenging available provision. Schools in England each have a governing body responsible for formulating the local strategy of the school and holding the leadership of the school to account for delivery of this strategy. There is also a wider role in determining the development of education for the society and evaluating its progress as well as the representatives who purport to deliver this. All of these roles hinge on making comparisons, between the actual and other years, other areas, an expectation, or a promise. Thus some considerable time has been invested in producing information that facilitates these sorts of assessments, while remaining comprehensible enough to engage citizens equally.

The aims of education are many, but one of some significance to citizens is the renewal and upskilling of the labour force. This is deemed important for economic prospects at a national level, and employment prospects at an individual level. Thus the standards of education deriving from the education system and whether they meet the needs and expectations of the nation are a key accountability for education policy. One of the outputs of education is public examinations which accredit the standard of education achieved by different individuals within the same cohort, and so it is natural to compile some kind of list of achievement against this measure. At a national level this would allow for considering progress and comparing the uptake of, and achievement in, different subjects of study. At a local level, this would allow again for assessment of progress but also for the comparison of different institutions. To these ends national results are published annually and schools have been placed within local league tables showing their performance. However, it is well understood that such tables are only showing one output of schools (Royal Society 2008) and we return to further data later in this section.

The measures used have included the idea of a school leaving standard of 5 GCSEs at grade ‘C’ or above but these have developed beyond this basic measure for several

reasons. By creating this public measure, the accountability structure becomes distorted towards improving this value, in the way described by Ghobadian et al. (2009), this has led to gaming of system, rather than an increase in educational outputs, in some cases. This gaming includes targeting tuition at borderline students and switching to easier courses or qualifications equivalent to a number of passes and teaching to the exam, rather for the education. Part of the problem here is that this is breaking students into two groups and not looking at the variation within those above and below the threshold, so the story is about meeting a standard rather than the quality of provision. Another problem is that this doesn't take into account how far students have to travel in order to achieve this standard: schools are there to make progress, rather than achieve an absolute level. Pupils do not enter the system being equal, so we shouldn't use an output measure which assumes they were. The allowance of equivalence to GCSE of many other qualifications has made a distortion away from an aim on leaving school to give rise to a concern that 5 passes can be achieved without learning functional skills in certain key subjects. Further, statisticians have raised concerns about the use of league tables to make comparisons (Goldstein and Spiegelhalter 1996), both on the simple level of presentation and then to look at progress over many years or expectation of performance in the future (Leckie and Goldstein 2009). However, many of the concerns of statisticians relate to national and multiple comparisons whereas local descriptive comparisons which show large differences are both indicative and useful to the public.

The indicators available and the ways they are used have evolved: value-added measures which take into account the local context have been developed for assuring the effectiveness of schools; other outcomes like %NEET (not in education, employment or training) have been used on a community level; output is expected include functional skills, such as 5 GCSEs at grade 'C' or above, including English and maths amongst the five. These other approaches all have new weaknesses but none has been able to engage with the data story of falling standards, as opposed to increasing achievement. It is useful to recall that "a richly nuanced or profound analysis of a large question is never logically excluded by the attempt to quantify parts of it" (Porter 1995). Thus inspection of schools includes much more than simply the assessment of statistical measures, and generally so does public dialogue, it is nevertheless instructive to consider the data stories which arise at various levels in relation to these new indicators.

The value-added is a contentious approach to education as the name promises rather more than the method tries to deliver. The model (for it is only a model) determines the value-added by a school as the variation between the pupil progress there compared to other schools, controlling for various individual level factors, and a few at school level. This makes it entirely possible that there are other local level factors which are not included in the model which affect outcomes. Local factors include a binary measure of poverty (there is some defence of this i.e. that below a certain threshold there is no space for independent study in the home but wealth should not make so much difference above that (Jesson 2008)) which is financial in nature and does not capture social and cultural poverty which are known to affect educational attainment (Royal Society 2008). Although generally published, it is not particularly embraced by the public and is most used within schools (although even there few people understand it well). Information on how this indicator is determined is not very accessible and is very recently becoming available, principally for the information of school managers. Training for governors, who are expected to hold their school management accountable for the performance of these indicators, is not available in a suitable form except on an ad hoc form on a school by school basis. As governors, and to some extent head teachers, have little training in statistics (SGOSS doesn't even list this as a desirable competency) the accountability of these indicators is confused, resulting in feelings of stigma and political manipulation which not a fair reflection on the process of development.

Some of the pressures are created by artificial use of the same data for several purposes as happened in the previous case study: value-added used for public accountability, formal inspection and governor monitoring could cause a conflict over the presentation of the data. However, this is not necessary as Ofsted make national inferential comparisons based on the statistical models whereas the public would be most interested in what this indicates for their own children. Governors should be considering variation within the school and progress over years rather than an annual snapshot. Indeed this system is still possible to game:

"identifying more pupils with special educational needs resulted in a positive influence on the school's contextual value-added score." (Ofsted 2010 p.22)

Further research is establishing a more reasonable model of school attainment by considering family and neighbourhood effects (Rasbash, Leckie et al. 2010), as well as developing models which allow for more personalised information than one number per school such as quantile models. This still leaves the problem of facilitating the understanding of users both inside and outside of schools, as well as the obvious problems of different primary school effects and the fact that there is more to education than what is measured by national tests.

The development of the indicators for NEET and the inclusion of English and Maths are more sensible in that they relate to something more comprehensible. They still fail to engage with the nature of education and the desire for the level of each child to be assessed. Public dialogue still includes the basic measure of 5 passes even though this is widely discredited. League tables have been ordered alphabetically and include information on the context of the school so that they allow for a more complex comparison rather than just a simple ranking. However, this ignores the nature of community expectation and understanding of the context: comparisons are not just made on the basis of the data presented but also on expectations. Thus certain schools locally will be expected to perform better than each other, based on knowledge of the community, and perhaps the circumstances of the school. Such comparisons can only be facilitated by the openness of the measures used to indicate performance. However, these are all still limited in their applicability to individual parents making a choice for their child: they have much more utility in community accountability than parental choice due to the six year period between enrolling and completing school (Leckie and Goldstein 2009). Thus indicators do not show outcomes of interest to parents but are also part of a complex data decision process likely to extend beyond individuals.

Handling Uncertainty in Science

The Royal Society held a discussion meeting entitled ‘Handling Uncertainty in Science’ on 22-23 March 2010 (royalsociety.org/Handling-uncertainty-in-science/). In part this was a more general meeting as a public engagement during the 350th anniversary year of the Society; it also addressed recent concerns about scientific treatments of uncertainty and risk and how to make policy decisions under uncertainty. Discussion meetings are limited in attendance but open to any participant without

charge, the presentations are recorded and available on the web, as well as being the subject of publication. Thus although the author attended, there is no reason to accept this analysis as complete so it focuses on the role of the data stories within the scientific talks and policy implications. Proceedings are published but are not available for this meeting at the time of writing but it is actually the dialogue and anecdotes which are of interest, rather than the formal academic presentation.

The structure of the meeting was four talks on theory of uncertainty, eight on substantive issues, and three more on communication, concluding with a plenary discussion each session being chaired by another scientist. Although the philosophy of science and classification of uncertainty are of interest and will be relevant to parts of this analysis, the main focus will be on the substantive talks as these provide data stories which are now the focus of our analysis. Interpretation of these which followed in the communications sessions will also be covered, from a perspective critical of the objectivity perceived in the sense that sciences aspire to objective methodology (Porter 1995).

Substantive issues covered were weather, economics, risk and physics, reflecting the range of applications of science as it impacts on the public consciousness. Thus the issues of uncertainty include measurement, modelling, prediction and risk, with Donald Rumsfeld becoming an unlikely point of reference for his infamous classification of knowns and unknowns. Another familiar theme was the classification of substantive fields on two axes of ‘ease of measurement’ and ‘process complexity’ such that the most difficult problems, notably most of social science, having relatively high values on both. Thus, fitting for an interdisciplinary meeting, there was an appreciation of the challenges faced by problems in different areas meaning that solutions had more or less certainty depending on the complexity. In this way we can see the outreach nature of the meeting and the determination to engage in dialogue rather than criticism although it is exactly this, where the dialogues are misunderstood and criticisms are made, that is of interest to us.

The data stories which come out of the meeting are more varied than in the other cases as they refer to totally different data, but they are expressed with much more conviction as they are described by leading substantivist. Thus there is much more to

learn about objectivity from these representations than from previous cases but actually rather less about statistical literacy from the citizen perspective. Indeed we should remember that scientists have their own role as ordinary citizens on top of their scientific input. Thus the stories that were expressed during the meeting were about experiences of communication relating to the substantive topic and topical issues of the time, rather than the actual content of the talks themselves, scientists being aware of many of the issues of uncertainty within their own work.

In keeping with the title of the meeting, most of the data stories centre on uncertainty and how it can be understood by members of the public through the means of models and other representations. A typical example is the risk of rare events: this is familiar in the field of medical testing where the sensitivity and specificity of a test need to be considered in conjunction with the population prevalence and any prior selection probability for testing. The issues here are that tests often have different probabilities for false positives and false negatives, yet these can only occur to people who do not and do have the condition respectively. Thus if the condition is very rare, a positive result is likely to be a false positive, even if the test is very specific. A case like this was discussed during the meeting in the understanding of a villager of the reliability of flood warnings which would only be correct seven times out of ten. The villager replied that their guesses would only be right half of the time so this was still better than they could do on their own. This response conflates all of the issues above leaving it unclear whether the true point had been understood. Indeed it was admitted that no erroneous warnings had been made to date and it would be interesting to see their impact.

The problem of treating (95%) confidence intervals as representing upper and lower bounds on the range of possibilities was also raised during the meeting. In the case of economic forecasts, these are usually made with some kind of probability fan indicating central forecasts and variants. In 2009, of course, economic growth deviated rather sharply from the forecasts but this was just outside the 95% region, not a deviation which had been thought impossible. The Bank of England has also been trying to develop multidimensional representations of the uncertainty but it is not clear how easy these would be to engage with.

National monitoring of swine 'flu involved regular public reporting of the situation and forecasting of future expectations of prevalence. Media generally reported these forecasts and later compared expectations to the realisations through reported cases. This presented some issues in the shift in testing to measure the actual number of cases, to relying on other means of estimating when cases were better understood and increasing. However, a further problem was apparent in that forecasts were made of median expectation and a worst case scenario, for which the government was duty bound to plan, even if it was unlikely. Media presentations and assessment tended to use this worst case value for reporting and comparison without making it clear that this is what it was, and that other forecasts were being used. This gave a public impression of overestimation, scaremongering, and wasteful stockpiling of medication.

Other topics raised made it clear that the ethical behaviour of scientists and the scientific quality of their work conflated in the public consciousness to their detriment on disclosures of some poor practice. Moreover, there was scientific consensus on the validity of the work, even when popular opinion may not have been in agreement, such as the issues raised in the plenary where the assembled asserted public trust in epidemic warnings but a lack of current confidence in economic forecasts, in contrast to the chair of that session. The data stories from the meeting have a frustrated air which does not particularly admit of public interpretation so much as dictation of instructions from scientists: that there is one interpretation and scientists have worked it out. However, there was agreement that social science input was important to these issues and that further meetings of this kind would be worthwhile in the near future. Perhaps this is a surprising admission of significant scientists at the Royal Society but actually quite promising also that public engagement is seen as so important, although this is not a new idea (Feyerabend 1982).

The issue of objectivity is very important to many of these data stories, but it often gets hidden in the interpretation of the results, given the assumptions made. Climate models make many assumptions, some of which are theoretically derived, some from empirical and historical data, and some drawn from a range of plausible possibilities. Other aspects represent the bounds of scientific understanding of the climate and the extent of feasible computation e.g. in the resolution of global climate models. Making a comprehensive understanding of these possible variations given the information

provided by scientists is difficult, especially when there are concerns raised about some issues of judgment in these matters. However, it is probably in issues like climate change in which citizen engagement in the scientific understanding is most important as consequences have such far reaching potential.

Descriptive Statistics

The nature of statistics which are not inferential, the purposes they might be put to, and their effectiveness at communicating information is not well defined. However, these are certainly not negligible and they rely on long history of monitoring political processes (Porter 1995; Desrosières 1998). Moreover, aspects of their use have been the subject of some study, and our present analysis shows their importance in contemporary data stories: where the public have met inferential statistics, they have failed to engage with them. Thus we will in this section attempt to clarify the nature, purposes and effectiveness of non-inferential statistics.

The definition of a statistic derives from its use referring to some item of data about a state, within some larger region of governance (Desrosières 1998). This has been generalised to a sample estimate of a population parameter, although the nature of the estimation and the extent of the population are sometimes hard to specify. Parameters are typically quantiles (e.g. the median) or averages (e.g. the arithmetic mean) although they may relate to any aspect of a model (this should draw to attention that the population may refer to a model, rather than the process of data generation). The population may also be rather nebulous in terms of time and space, as well as more selection based aspects of membership. Estimation is often a literal process of approximation, where an analytic solution is infeasible, and may itself involve an arbitrary choice of loss function. Thus any statistic has a complex relation to the data and a population which can easily be misunderstood.

By saying a statistic is an empirical estimate of some parameter, it should be clear that it represents an attempt to reduce information, to distil it into a form which encapsulates much data in one statement. Thus we define every statistic to be a summary, with the understanding that to be inferential it requires reference to some probability distribution (Cox 1981). However, we contend that not all statistics are descriptive, although inferential statistics may also be descriptive but need not

necessarily be so. Moreover, we will show that the descriptive nature of a statistic will be dependent on the audience it is presented to as description is an inherently comparative process.

Inference has contrast at its heart: considering the difference between the data and some model of how it might have been generated and evaluating whether it is sufficient to reject a comparison and force a contrast. Description aims to represent a novel or unknown fact (in our case some aspect of the data at hand) by comparison to known things. Thus it depends on the knowledge of the audience through this description by terms external to the data.

In general, one can become familiar with a certain kind of data by repeated experience with similar modes of generation. For technical reports and trade audiences, it can be assumed that standard reports on similar problems will be understood, even though individual readers may have to seek expert support. For a public audience, much more needs to be understood of their familiarity with the process, taking account of possible misconceptions. This knowledge may not even be available but is certainly worth pursuing if the intent is genuine communication. As even graphics require domain specific knowledge (Roth 2004), it should not be assumed that conventional representations are sufficient for description.

Thus we must establish what embodies a suitable referent to facilitate a description. The cases of this paper are useful in establishing weaknesses in some descriptions: simply reporting the number of responses recorded within the target time for ambulances (Bevan and Hamblin 2009) was insufficient to describe the progress in improving response times. The confidence interval for hospital standardised mortality, or the equivalent ‘excess’ deaths, for a given time period did not describe the nature or extent of problems at the health trust. Annual entry data and pass rates for public examinations do not describe the education received by the children in schools. Predictions of climate models tend not to describe the experience that any given person would have of climate change, nor do their parameters describe the physical factors of the system. Thus there are two problems here: describing what one wishes to, and utilising the public descriptive power of a statistic. The weaknesses can be

summed up as the data story was not the same as what the statistic was able to describe, leading to misrepresentation and confusion.

The simple line graphs of Bevan and Hamblin (2009) are sufficient to show the distribution of times of response and therefore develop the descriptive power of the target statistic. Further statistics, quantiles of the distribution are also conduits of this information, in the absence of prior knowledge of the distribution. For comparative mortality which is intrinsic to the output of a hospital but less controlled than specific response times, prior information of previous years or similar hospitals (in the sense that one expects to be similar to them) may facilitate description. The breakdown of mortality by diagnosis should be rather more descriptive to those who know the strengths and weaknesses of the institution as they can compare the observed range to the overall deviation. However, none of these would necessarily give an absolute understanding of performance as people die but a reasonable rate is difficult to specify. Thus in simple cases, grasp of the distribution and comparators, and understanding of the range and variation in other subgroups, helps but so do well defined expectations.

Statistics are published about school performance frequently but there are problems of the extent to which they truly describe anything (we do not consider the effect of innovations in policy, such as new types of schools or specific methods teaching as these have more technical critiques). There is an essential challenge that the nature of education evolves so that new ideas percolate down into the curriculum, both directly and through their influence on thought and the nature of society. However, this should not prevent many of the intended comparisons: between schools to see variation in performance; and at the same school in recent years to track improvement. The CVA is intended to describe performance variation which it may do from a national perspective (ideal for government and Ofsted) but I contend that it is not descriptive, not least because it is not understood. Six points correspond to one grade over eight GCSE subjects, thus variation is not as great as it might be perceived – it does not relate to expectations. There is the further weakness of the modelling that it does not facilitate comparisons over time, or take into account family effects (i.e. all local variation is ascribed to the school). Specifically it is a national model where many comparisons made by citizens would be much more local, thus the lack of statistical significance attached to the data is counterintuitive and national comparisons made

are frustratingly irrelevant (e.g. wealth varies nationally, so a national indicator of poverty will mean different things in different regions).

Many measures are applied to schools such as indicators of attendance and modes of travel, as well as the proportions coming from within a certain area around the school. It is very difficult to make these concepts descriptive as not much prior information is available and therefore one would fall back on graphing the distribution, mapping the locality and comparisons to previous distributions and other schools. Public examination results are rather more familiar. Intended as certification for employers, they also form a record of the output of a school over time and comparison over time within an area is more feasible. There are problems of gaming and the validity of the indicators but they are readily understood and their common currency makes them natural by-products of education, even if they do not represent the education themselves. Thus the essence of the task is to make these more descriptive (perhaps several markers of attainment from 5A-G and 5A-C to include a language as well as standard requirements) as well as realising that they represent the likely outcomes of education as context is not something for which allowances are often made in employment. Indeed it would be more useful to validate these descriptive statistics against outcomes than to expect public understanding of CVA models. Descriptive statistics should facilitate questions whereas CVA models inferences and also confuses, leaving citizens baffled or sidelined but certainly not empowered.

Statistics for communicating science are more various and often rather complex constructions, often of scientists' design rather than statisticians'. Thus public presentation may cause misinterpretation or challenge for the results, methods or conclusions, meaning that construction of models for uncertainty must acknowledge the nature of their objectivity. Some citizens are unfamiliar with the construction of empirical models as theoretically plausible fits to the data (Petocz and Reid 2010) but most would be surprised by the contention that some concordance with other models is required to retain the confidence of funders. Thus though predictions and estimates derived from models may well be descriptive, the concern is that they describe the output from the models, not an objective representation of scientific truth. More is also required to put these descriptive statistics in context by providing a commentary

(c.f. UKSA 2010), particularly when the statistic initially contains some information about uncertainty or variation.

With all statistics, they are intended to be useful to the audience, indeed for citizens we go so far as to say they should be convivial tools (Illich 1973), helping people to live their lives without putting arbitrary constraints on them. Thus we can see a weakness in general inferential statistics that the accompanying probability distribution gives a misleading sense of objectivity which disempowers. Thus descriptive statistics are convivial but inferential statistics may become so when their uncertainty is mediated by a professional. There are of course some weaknesses that mean multivariate statistics can mislead (Chatterjee and Firat 2007), and unfamiliar constructs will require further information such as a plot of the distribution. Therefore descriptive statistics are both socially constructed and convivial only to certain audiences, so that good commentary and a level of public understanding are essential to their use.

Statistics in Society

Statistics keep advancing into new roles in society, partly to fulfil expectations of objectivity (Porter 1995), and also to make sense of complex bodies of data. This means that in order for citizens to take part in society that the level of basic education needed will begin to include some understanding of how statistics can be used and how they can be collected. The UK is more advanced in this practice than most countries, and the demands on citizens are greater because of the role that lay people take in governance and accountability, as well as the prevailing agenda of public disclosure and public choice. The conception of citizenship as giving equality of participation means that a conflict between abilities and requirements demands adjustments to education, expectations, or both.

The case studies considered here show that although there may be problems of the ability of citizens to engage with statistics, there are also further problems of the appropriateness of the statistics provided for certain purposes. In making choices, citizens should be left with some freedom whereas the confidence intervals provided with inferential statistics certainly appear to preclude individual decision making. Indeed there is no reason why citizens should be bound by a particular size of test,

like the 5% error rate generally applied by statisticians. Furthermore, the statistics provided often neglect that the range over which choices are made are often much smaller than the national scale for which the statistics were originally designed. This leads to several problems of comparability, but also a frustrating failure to include local knowledge which is important to local citizens. Thus statistics provided cannot grant equal representation without some flexibility to local needs. This might include the use of local definitions, or comparisons over local areas.

Statistics can have positive roles in society by focusing on particular aspects of issues and standardising some processes. They have certainly been instrumental in improving the clarity and accountability of certain government actions (Porter 1995), although these processes continue to evolve. Many current developments in governance require representative consultation activities and others make use of official survey data on the presence of certain groups. Thus for the needs of groups to be properly taken into account, they must be properly represented by such surveys. The many issues which affect response to and inclusion in such government data become therefore very important to citizens. Indeed this is represented by the term statistical citizenship in political literature. The extent to which dealing with inadequate representation is a problem of engagement or compensating methodology will be of great important in resolving this.

Much more could be written about the weaknesses of particular statistics for specific purposes, but the main concern of this article is that such statistics can limit political activity by artificially constraining the arguments considered. This may not be a deliberate act and in fact may be caused by expedience or feasibility but, when such factors are not explained, this will cause exclusion rather than the intended advance of understanding. Porter (1995) makes many arguments for uses standardised processes to increase the objectivity of political decisions, but every advance has its limits which expose other problems. Indeed these practices set a high bar for participation in a political dialogue, although this is not insurmountable it often requires significant vested interest for a group to make this engagement (Evans and Plows 2007). The essence of this is that citizens should be able to criticise the statistics they are presented with (Gal 2002), and to write their own data stories. This does not require everyone to be a statistician but all to be aware of the power of statistics to stigmatise

and create social truths. Thus developments should take into account the abilities of those who will come into contact with new statistics and engage with developing data stories from them. Fundamentally, those being held accountable by statistics should be enabled to understand the mechanisms by which these judgments will be made, and so the accountable can hold the system of accountability to account.

Initiatives continue to appear to support user engagement; the question is to what extent they conceive the engagement users wish to have, as opposed to that which is politically desirable. The issue of how to engage the wider public is present in the efforts of the UKSA (2010) and the present campaign of the Royal Statistical Society (www.getstats.org.uk) but in both cases they represent the weakest aspects of their plans. What is actually understood by individuals and how this informs their behaviour is not understood. What further information or education would actually enhance their understanding, confidence and trust in their representatives is not understood either. How the present situation falls short of its aims to empower and enable citizens is hinted at in the present analysis but these three cases only scratch the surface of what may become a substantial problem. Misconceptions and misuses abound but the fundamental problem is that statistics are used when there is uncertainty, not only limits on precision but also more qualitative aspects of definition and measurement. Much more research is needed on the public understanding of statistics which may be informed.

Acknowledgments

This paper is a development of joint work with Clare Woodford (QMUL) presented at the Royal Statistical Society Annual Conference in Brighton 2010 under the title ‘Statistical Literacy for Citizenship: the role of descriptive statistics’. I am very grateful to Clare for challenging my prose and my political naïveté. Various ideas have been developed through conversations with colleagues in Southampton and those involved with the ‘getstats’ campaign. I am also grateful to the insights gained whilst a governors at a secondary school in Hampshire. However, views expressed in this paper are those of the author alone.

References

Andersen, B. (1990). Methodological errors in medical research: an incomplete catalogue. Oxford, Blackwell.

Barnett, V. (1999). Comparative Statistical Inference. London, Chapman & Hall.

Bellamy, R. (2008). Citizenship: a very short introduction. Oxford, Oxford University Press.

Bevan, G. and R. Hamblin (2009). "Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK." Journal of the Royal Statistical Society Series A - Statistics in Society **172**(1): 161-190.

Bird, S., D. Cox, et al. (2005). "Performance Indicators: good, bad, and ugly." Journal of the Royal Statistical Society Series A - Statistics in Society **168**(1): 1-27.

Boyne, G., O. James, et al. (2009). "Democracy and Government Performance: holding incumbents accountable in English local governments." Journal of Politics **71**(4): 1272-1284.

Chatfield, C. (1995). Problem Solving: a statistician's guide. London, Chapman & Hall.

Chatterjee, S. and A. Firat (2007). "Generating Data with Identical Statistics but Dissimilar Graphics: a follow up to the Anscombe dataset." The American Statistician **61**(3): 248-254.

Coghlan, B., R. Brennan, et al. (2006). "Mortality in the Democratic Republic of Congo: a nationwide survey." The Lancet **367**(1): 44-51.

Cox, D. (1981). Applied Statistics: principles and examples. New York, Chapman and Hall.

Crettaz von Roten, F. (2006). "Do we need a public understanding of statistics?" Public Understanding of Science **15**(2): 243-249.

de Santos, M. (2009). "Fact-Totems and the Statistical Imagination: the public life of a statistic in Argentina 2001." Sociological Theory **27**(4): 466-489.

Demeritt, D. (2001). "The Construction of Global Warming and the Politics of Science." Annals of the American Association of Geography **91**(2): 307-337.

Desrosières, A. (1998). The Politics of Large Numbers: a history of statistical reasoning. Cambridge, MA, Harvard University Press.

Evans, A. (2003). "Accidental Fatalities in Transport." Jounal of the Royal Statistical Society A **166**(2): 253-260.

Evans, J. and I. Rappaport (1999). Using Statistics in Everyday Life: from barefoot statisticians to critical citizens. Statistics in Society. D. S. Dorling, S. London, Arnold**:** 71-77.

Evans, R. and A. Plows (2007). "Listening Without Prejudice? rediscovering the value of the disinterested citizen." Social Studies of Science **37**(6): 827-853.

Feyerabend, P. (1982). Science in a Free Society, Verso.

Francis, R. (2010). Independent Inquiry into care provided by Mid Staffordshire NHS Foundation Trust January 2005 – March 2009. London, The Stationery Office.

Gal, I. (2002). "Adults' Statistical Literacy: Meanings, Components, Responsibilities." International Statistical Review **70**(1): 1-51.

Gelman, A. and D. Nolan (2002). Teaching Statistics: a bag of tricks. Oxford, OUP.

Ghobadian, A., H. Viney, et al. (2009). "Explaining the Unintended Consequences of Public Sector Reform." Management Decision **47**(10): 1514-1535.

Goldstein, H. and D. J. Spiegelhalter (1996). "League tables and their limitations: Statistical issues in comparisons of institutional performance." Journal of the Royal Statistical Society Series a-Statistics in Society **159**: 385-409.

Goodyear-Smith, F., B. Arroll, et al. (2008). "Patients Prefer Pictures to Numbers to Express Cardiovascular Benefit from Treatment." Annals of Family Medicine **6**(3): 213-217.

Hand, D. (1996). "Statistics and the Theory of Measurement." Journal of the Royal Statistical Society A **159**(3): 445-492.

Hennig, C. (2010). "Mathematical Models and Reality: a constructivist perspective." Foundations of Science **15**(1): 29-48.

Hodgson, L., C. Farrell, et al. (2007). "Improving UK Public Services: a review of the evidence." Public Administration **85**(2): 355-382.

Holmes, P. (2003). "50 Years of Statistics Teaching in English Schools: some milestones." The Statistician **52**(4): 439.

Illich, I. (1973). Tools for Conviviality. London.

Jesson, D. (2008). Using Data, Improving Schools. London, Ofsted.

Leckie, G. and H. Goldstein (2009). "The limitations of using school league tables to inform school choice." Journal of the Royal Statistical Society Series A - Statistics in Society **172**(4): 835-851.

Lilford, R., M. Mohammed, et al. (2004). "Use and Misuse of Process and Outcome Data in Managing Performance of Acute Medical Care: avoiding institutional stigma." The Lancet **363**: 1147-1154.

Llewellyn, S. and D. Northcott (2005). "The Average Hospital." Accounting, Organizations and Society **30**: 555-583.

Miller, B. and B. Barnett (2010). "Understanding of Health Risks Aided by Graphics with Text." Newspaper Research Journal **31**(1): 52-68.

Ofsted (2010). The Special Educational Needs and Disability Review. London, Ofsted.

Percival, A. (1919). "How to Read Statistics." British Medical Journal **1**(3044): 540-542.

Petocz, P. and A. Reid (2010). "On Becoming a Statistician - a qualitative view." International Statistical Review **78**(2).

Pfannkuch, M., M. Regan, et al. (2010). "Telling Data Stories: essential dialogues for comparative inference." Journal of Statistics Education **18**(1).

Porter, T. (1995). Trust in Numbers: the pursuit of objectivity in science and public life. Princeton, NJ, Princeton University Press.

Rasbash, J., G. Leckie, et al. (2010). "Children's Educational Progress: partitioning family school and area effects." Jounal of the Royal Statistical Society A **173**(3): 657-682.

Roth, W.-M. (2004). "Emergence of Graphing Practices in Scientific Research." Journal of Cognition and Culture **4**(3-4): 595-627.

Royal Society (2008). Exploring the relationship between socioeconomic status and participation and attainment in science education. London, Royal Society.

Royal Society (2008). Science and Mathematics Education, 14 –19. State of the Nation Reports. London, Royal Society.

Smith, T. M. F. and L. Staetsky (2007). "The teaching of statistics in UK universities." Journal of the Royal Statistical Society: Series A (Statistics in Society) **170**(3): 581-622.

Sowden, S. and R. Raine (2008). "Running Along Parallel Lines: how political reality impedes the evaluation of public health interventions." Journal of Epidemiology and Community Health **62**(5): 835-841.

Sturgis, P. and N. Allum (2004). "Science in Society: re-evaluating the deficit model of public attitudes." Public Understanding of Science **13**(1): 55-74.

UKSA (2010). Strengthening User Engagement – interim report. London, UKSA.

Utts, J. (2003). "What Educated Citizens Should Know About Statistics and Probability." The American Statistician **57**(2): 74-79.

Wallman, K. K. (1993). "Enhancing Statistical Literacy - Enriching Our Society." Journal of the American Statistical Association **88**(421): 1-8.