The Use of Algorithms in Regulation and Public Service Delivery

Summary

1. This submission summarises the results of research on the use of algorithmic decision-making tools in regulation and public service delivery conducted by academics from King’s College London and London School of Economics and Political Science.

2. In two separate studies we showed that algorithmic decision-making tools do not help regulators decide which hospitals or universities should be prioritised for inspection because of problems with the quality, timeliness and granularity of indicator data; the small 'n' and base rate fallacy; and the ambiguity, instability and contestedness of regulatory goals.

3. The wider applicability of these challenges has been confirmed in further ongoing research with some 20 other public bodies, which has identified further challenges in using algorithms to improve public service delivery and regulation including: data protection duties; anti-discrimination and due process concerns; leadership; staffing and skill shortages.

The Use of Algorithms in Regulation

4. Over the last two decades, UK regulators have increasingly adopted so-called ‘risk-based’ approaches to organising regulatory enforcement and targeting inspections according to the probability and consequence of non-compliance amongst their regulatees\(^1\)\(^{(Rothstein, 2012 #104)}\). Indeed, risk-based approaches to inspection have been legally mandated for all UK regulators since 2008\(^2\).

5. Traditionally, regulators used a combination of subjective expert judgment and information from past inspections to decide which regulatees to prioritise for inspection. Pressed to follow the ‘Big Data’ analytics example of Amazon and other private sector organisations, however, a number of UK regulators have started to develop and use algorithmic decision-making tools to predict which regulatees are at greatest risk of non-compliance.

6. Algorithmic tools offer the hope of targeting regulatory inspections more accurately, objectively and efficiently by applying a variety of statistical and machine learning techniques to large sets of indicator and regulatee performance data. For example, Ofsted\(^3\), the Care Quality Commission\(^4\) (CQC), and the Rural Payments Agency\(^5\) automatically aggregate large volumes of performance data to decide which organisations pose the greatest risk and prioritise their resource accordingly.

7. To date, however, there has been little peer-reviewed research on whether these new algorithmic approaches to regulation actually work. Our research evaluated the use of algorithmic analysis of existing datasets to predict inspection outcomes in two regulatory domains; healthcare and higher education quality. In both cases we found that algorithmic decision-making tools failed to improve regulatory targeting.
Written evidence submitted by Dr Alex Griffiths, Dr Henry Rothstein and Professor David Demeritt (ALG0065)

**Case study 1: Use of algorithms to regulate healthcare quality**

8. The Care Quality Commission is responsible for regulating the quality of care delivered by over 30,000 health and social care providers in England, including 150 NHS Hospital Trusts. In order to optimise its use of inspection resources, the CQC has invested heavily in developing a series of statistical surveillance systems to crunch the wealth of administrative data generated in the NHS to identify potential risks to quality and thus help ‘make better decisions about when, where, and what to inspect’.

9. The CQC developed its first statistical surveillance system when it was established in 2008; adapting an earlier risk-rating system developed by its predecessor, the Healthcare Commission, to calculate what were known as *Quality and Risk Profiles* (QRPs) for all health and social care providers. QRPs, which were developed in conjunction with leading statisticians, drew on more than a thousand healthcare indicators to estimate the probability of a provider failing to meet 16 regulatory standards. However, despite the sophistication of the QRP’s underlying algorithm, the CQC was unable to use the tool to prevent a number of high-profile failures across the health and social care sectors.

10. Following criticism by the Francis Inquiry into the failings at Mid-Staffordshire NHS Foundation Trust, the CQC altered its regulatory approach and replaced QRPs with a simpler statistical surveillance system to help decide what and where to inspect. The new “Intelligent Monitoring” system, which was introduced in 2013, was developed at a cost of £1.2m by *McKinsey & Co* and was claimed to be “world leading”. The system uses just 150 unweighted indicators that were selected following a broad consultation process to identify those indicators that were the ‘most important for monitoring risks to the quality of care’. Those indicators were then combined in a simple algorithm to produce a single risk estimate to help the CQC decide which providers to prioritise for inspection.

11. In order to test the reliability this new system, our ESRC-funded research compared the Intelligent Monitoring risk scores for 110 NHS Hospital Trusts against their subsequent Ofsted-style quality ratings awarded following detailed on-site inspections by large expert teams from the CQC. Our analysis showed that the Intelligent Monitoring risk scores could not effectively prioritise Trusts for inspection. Indeed, according to the measure used by NAO and by the CQC itself, Intelligent Monitoring predictions of hospital quality were wrong more often than they were right.

**Case study 2: Use of algorithms to regulate teaching quality in higher education**

12. The 2011 Higher Education White Paper *Students at the Heart of the System* called for the Quality Assurance Agency (QAA), the body charged with monitoring and advising on standards and quality in UK Higher Education, to adopt “a genuinely risk-based approach” and to “explore options in which the frequency – and perhaps need – for a full, scheduled institutional review will depend on an objective assessment of a basket of data, monitored continually but at arm’s length.”

13. In contrast to the CQC’s consultative approach to indicator selection, the QAA co-sponsored an ESRC PhD studentship at King’s College London by Alex Griffiths to use machine learning...
Written evidence submitted by Dr Alex Griffiths, Dr Henry Rothstein and Professor David Demeritt (ALG0065)

techniques to identify whether there was any available data that could be used to reliably identify which higher education providers should be prioritised for inspection. To that end, Griffiths gathered data on over 1,000 metrics in diverse areas such as student satisfaction and employability, drop-out rates, finance, staff and student characteristics, and complaints.

14. Griffiths then used a range of machine learning techniques to see if that data could be used to predict the results of hundreds of institutional reviews conducted by the QAA between 2007 and 2014. Even accounting for changes in performance over time and standardising for the type of institution, his analysis demonstrated that no automatically-aggregated combination of metrics could be found that could help the QAA reliably prioritise its inspections.

Why algorithmic techniques failed in these two cases

15. In both these cases algorithmic decision-making tools were used to help regulators decide which regulatees should be prioritised for inspection. In our published research we identify three major reasons why algorithmic techniques failed to improve the regulation of healthcare and higher education quality:

i. The quality, timeliness, and granularity of indicator data. Healthcare and teaching quality are difficult to measure, and key performance indicators, like patient waiting times or National Student Survey scores, are liable to gaming by regulatees. Indicator sets were often based on annual data collections and compiled at NHS Trust or university-level. As such indicator data lagged behind current performance and failed to reflect significant variation in quality within the large, complex organisations.

ii. Small ‘n’ and the base rate fallacy. Algorithmic techniques work best when there is a large number of data-rich cases so that statistical associations can be found between the outcomes of regulatory interest (i.e. non-compliance) and various independent indicators on which the probability of those outcomes statistically depend. In our cases, predictive power was limited by the small number of regulatees and inspection outcomes available for use in training algorithms and correctly identifying rare events with low base rates.

iii. The ambiguous, unstable and contested nature of regulatory goals. In the NHS successive governments have struggled to agree on the relative priority that should be given to cost efficiency, safety, waiting times, and patient satisfaction or to acknowledge the trade-offs between those different measures of healthcare quality. Likewise in higher education should university quality be assessed in terms of student employability, satisfaction, retention, widening participation or A-level tariffs? Algorithms can do little to improve regulatory decision-making when the underlying goals and priorities are poorly defined or inconsistent.
Wider Use of Algorithmic Decision-Making in the Public Sector

16. The specific challenges that we identified in the two studies of algorithmic regulation discussed above also apply more generally across the public sector, as we are discovering in a follow-up study on the use of algorithmic decision-making tools by government.

17. As part of that research, which is funded by King’s College London and is being conducted in collaboration with our KCL colleagues Prof Karen Yeung, Prof Ben Bowling, Dr Elizabeth Sklar and Prof Peter McBurney, we have talked to representatives from some 20 government departments, non-ministerial government departmental, local authorities, and other public bodies about their experience of developing and using various algorithmic decision-making techniques.

18. That research study has identified a number of further challenges for public bodies in using algorithmic techniques to improve public service delivery and regulation. These include:

i. **Data protection** rules that limit the ability of public sector bodies to link and share data sets or use them for purposes beyond those for which they were originally collected.

ii. **Anti-discrimination and due process concerns** about acting on cases identified as high risk through purely algorithmic techniques.

iii. **Leadership** willing both to support potentially risky innovation and also cognizant of the limitations and caveats about the use of algorithmic decision-making tools.

iv. **Hiring and retaining staff** with the necessary skills to effectively develop, operate, and maintain algorithmic decision-making tools and the data feeds necessary to keep them accurate and up-to-date.

Dr Alex Griffiths, London School of Economics and Political Science, Centre for Analysis of Risk and Regulation

Dr Henry Rothstein, King’s College London, Department of Geography

Professor David Demeritt, King’s College London, Department of Geography

April 2017

References


Written evidence submitted by Dr Alex Griffiths, Dr Henry Rothstein and Professor David Demeritt (ALG0065)


