Regulators' Forum

Predicting Quality Failure

What strategies exist to predict failure? What does the experience with existing methods of predicting quality failure tell us? This session of the Regulators' Forum focused on the experiences in higher education in particular, allowing for comparison with other regulated sectors. Existing findings place the future of a 'data-driven' world of quality failure prediction into a critical light.

Different strategies regarding the prediction of quality failures exist, all three of which can be seen as risk-based. One approach is to rely on a small set of contextual indicators that are used to categorise different regulatees. A second approach is to rely on a 'data-informed' approach in which a (often) large set of indicators is taken so as to generate a risk rating. This rating is then used to inform expert judgement. A third approach is to rely on a 'data-driven' or algorithmic' approach in which machine-learning techniques are utilised to develop an optimal model, based on large sets of performance data in order to develop a risk rating.

Such a reliance on algorithmic approaches is highly promising, but studies in higher education have offered a somewhat mixed picture. For higher education institutions, the relationship between data and review findings was so poor that no effective model could be developed. For further education institutions, the model's predictive powers were no better than chance. However, in the area of 'alternative providers', a model with predictive power did emerge.

Such findings raise considerable challenges for the future of 'data-driven' models in predicting quality failure. One fundamental issue affects the definition of what quality is. As there might be no agreed definition as to what 'quality' in particular sectors is supposed to be, there is therefore little hope of developing effective ways of measuring quality. Quality was inherently contested,

given that regulators changed their standards, that there were different understandings as to whether quality was an absolute concept, or should be understood in 'fit for purpose' terms. Metrics could not therefore be assumed to offer a straightforward solution for directing regulatory activity. There were also questions about the quality of the data, in terms of level of granularity, timeliness, quality and variability.

Other potential explanations are that none of the existing data manage to successfully measure quality, or that the existing inspection-based assessments of quality equally fail to measure quality. Such accounts should at least raise questions about the existing ways of measuring quality. It raised questions about how a regulator sought to ensure that there was consistency across its inspections. Inspections might also look at different aspects of an organisation than by means of data. Data might be arguably advantageous where the organisation had a clear function – in cases of complex organisations, there was considerable variation. Furthermore, there were question as to whether outcome definitions were too broad, so that individual indicators were not able to provide varied insights.

In higher education, it was not possible for a data-driven approach to predict an incident before it happened using either current data points or trends over time. The regulator, however, has asserted that expert interpretation of trends data as part of a data-informed approach will benefit the prioritization of interventions. Other regulators also used a data-informed approach in that they looked at low-level concerns so as to develop an indicator of trends. By looking at patterns, one could start initiating secondary controls.

There was a general concern about questions of quality when quality could not be defined in binary terms. Higher education is a case in point; there were numerous different aspects of student experience that could be identified as 'quality'. In other words, there were issues about measurability and aggregation. Furthermore, there were regular changes to the regulatory standard itself. There were also problems in joining up the different parts of organisations that were generating data – quality potentially varied enormously across organisations.

Regulators varied in terms of their target population – some regulating a relatively small set of firms with clearly defined activities, others look at a vast universe of highly diverse business activities. This alone had an impact on the time a data-informed approach might take. There were

also questions about how informed such a data-informed approach could be, raising questions about the feasible extent of peer-review.

A further question related to the extent to which data-informed approaches could be proactive, or whether they could also be reactive: they could use an incident to drive change. But there was also the question about identifying those features that were driving 'excellence' so as to enhance the possibility of cross-organisational learning. In other areas, the key question related to 'fitness to practise' and here patterns were used to identify weaknesses, such as the type of common hazards that were encountered or how individuals worked with others.

If, therefore, there were questions about a world in which algorithmic regulation could predict quality failure, there were also questions about a data-informed approach. Such an approach had advantages – if it included multiple sources of data, then it was very difficult to game such a system. It allowed the application of tacit knowledge, peer-review and nuanced opinion. It also could be used to target individual regulatees. At the same time, in order for such a data-informed approach to work, it was important to be aware of certain biases. One was the kind of biases produced by bounded rationality – confirmation biases and pattern- seeking behaviours risked making decision-making very problematic. There were questions about the selection of the set of indicators that could be used *a priori*: this in itself would be a source of bias. And finally there were also questions about the size of the set of indicators – the larger the quantity, the more likely it was that a single signal could be lost.

ML
10/16