# Insurance & Big Data

*Written submission to the Commons Select Committee on Science & Technology: 'The Big Data Dilemma' from the Insurance Law Research Group, University of Southampton.*[*]

03/09/2015

***Executive Summary:***

- The use of 'Big Data' by insurers (and those who advise them, such as actuaries) provides both opportunities and threats to the relationship between the industry, its customers and the State.
- We consider these to be emerging issues, with the current evidence suggesting that these are not yet in play in the United Kingdom.
- Our proposal is that these issues be investigated and monitored, rather than met with any immediate State action.
- Opportunities:
    - The rapid, automated use of large data sets may reduce the cost of risk assessment, permitting 'personalised' design and pricing of insurance products.
    - It will enhance the ability of insurers to detect fraudulent claims by identifying patterns of customer behaviour not honestly communicated to the insurer (such as smoking status or the extent of ongoing physical disability attributable to an accident).
    - It will enable insurers to better monitor customer's compliance with coverage conditions (such as the routine use of a security system).
- Threats:
    - It may reduce the current redistributive effects of insurance, by which the fortunate subsidise the unfortunate. 'Risk based pricing' brings not only the promise of more efficient contracting, but risks of extending the class of 'the uninsured and uninsurable'.
    - It may provide insurers with access to data currently considered private or personal. Should an insurer have routine access to customers' internet browsing history or supermarket loyalty card records? These factors may signal the need for ongoing revision of the extent to which insurance law and regulation controls the use of sensitive information.

[*] Co-authored by Professor James Davey and Mateusz Bek, School of Law, University of Southampton. Correspondence to J.A.Davey@soton.ac.uk.

## A. Insurance and Information

[1] Insurance is a product that is based fundamentally on the assessment of risk through information. Historically, the assumption was that the insured held all relevant knowledge pertaining to the risk, and was duty bound to communicate that information to its underwriter prior to the formation of the contract, so that the information asymmetry could be minimised and risk better priced. This position- adopted almost exactly 250 years ago in *Carter* v *Boehm*[1]- remained the presumed basis of the insurance relationship long after it stopped bearing any resemblance to reality.

[2] In truth, insurers have independently collected and utilised data to inform their actuarial processes for many decades.[2] This shift has been recognised in recent legislation[3] with a rebalancing of the informational duties in insurance, but does not go so far as to regulate insurers' use of data that they obtain for themselves.

[3] In this brief note, we highlight two areas in which the expanded use of 'Big Data' by insurers might further alter the relationship between the industry and its customers, to the extent that additional regulation might be desirable. These are, first, the use of 'Big Data' in the risk classification process that is the basis for pricing of insurance contracts; and second, the use of data to monitor the insured's behaviour during the term of the policy. Our interest in this area should not be mistaken for a call for immediate State intervention, but proposes instead a watching brief.

[4] Much of our analysis builds upon recent work in the United States, where the regulatory framework for insurance often comes under closer academic scrutiny than in the United Kingdom. The threat and opportunity that 'Big Data' presents within the insurance sector was the focus of a symposium in the leading US insurance law journal, the Connecticut Insurance Law Journal. There is an immediate need for a similar review of the UK position, as the differences in the regulatory landscape mean that UK and US insurers are likely to use 'Big Data' in significantly and subtly different ways.[4]

[5] The opportunities and threats within the US insurance markets are described by Professor Rick Swedloff:

---

[1] (1766) 3 Burr 1905, 1910 per Lord Mansfield.

[2] M Clarke, *Policies & Perceptions of Insurance in the Twenty-First Century* (OUP, 2007), 114-116.

[3] The Insurance Act 2015 received royal assent earlier this year and will enter into force on 12 August 2016.

[4] One obvious difference is the existence of a strong 'Federal' branch of anti-discrimination law within the EU, that manifests itself as the United Kingdom's Equality Act 2010. Within the US there is a much greater diversity of legal protections on race, gender, sexuality, genetic status and the like as the controls are largely set at the State level. For a recent review of each model, compare D Schwarcz *et al*, 'Towards A Universal Framework For Insurance Anti-Discrimination Laws' (2014) 21 Connecticut Insurance Law Journal 1 and J Davey, 'Genetic discrimination in insurance: lessons from *Test Achats*' in A De Paor *et al* (eds.), *Genetic Discrimination - Transatlantic Perspectives on the Case for a European Level Legal Response* (2014, Taylor & Francis).

'There are… few industries with as voracious an appetite for data, in any form, as the insurance industry. Carriers likely can no longer ignore the possibility that the algorithms driving big data will offer greater predictive accuracy than traditional statistical analysis alone. And, if realized, this additional accuracy could potentially benefit insurers in at least three ways. First, by analysing purchasing patterns, carriers could better target those individuals most likely to buy new coverage and retain those insureds most likely to switch to a different carrier. **Second, insurers may be able to use claims and settlement patterns to better distinguish between real and fraudulent claims. Third, again, to the extent greater predictive power is realized, carriers could use big data analytics to price their products more accurately**'.[5]

[6] Our focus in this paper is on the second and third areas: pricing by risk classification and identifying fraud. To this we add a further opportunity: changes in the nature of monitoring compliance with coverage conditions (such as the routine use of a burglar alarm, or fire prevention protocol).

[7] We would also echo the potential threats that Swedloff identifies:

'… the potential benefits of big data (to the extent carriers can recognize them) will not be free. Like any improvement in risk classification, additional expenditures on big data analytics could be socially wasteful and privately inefficient. Further, like all risk classification refinements, to the extent that the promised gains in predictive accuracy materialize, classifying risks with big data analytics may undermine important risk spreading goals of insurance. Lastly, mining individual data to build the data sets or to identify whether a potential insured falls into a particular risk category could invade policyholder privacy'.[6]

[8] Insurance, as currently operated, provides relatively little differentiation between high and low risk groups. The costs of obtaining, verifying and analysing risk data often outweigh the competitive advantage to be gained by more accurate pricing of the product. This means that many current risk groups operated by insurers are relatively heterogeneous, with invisible subsidies flowing from those who are below the average level of risk in the pool to those above that line. If 'Big Data' delivers on its promise of rapid, automated analysis of huge volumes of data then risk groups will be made more discrete as the price of risk assessment falls. There will be more risk groups with a greater degree of homogeneity in the actual risk profiles of the members of each group.[7] There is a danger that this will exacerbate the tendency of insurance to assist the fortunate,

---

[5] R. Swedloff, "Risk Classification's Big Data (R)Evolution" (2014) 21 Conn. Ins. L.J. 339 (emphasis added). Hereinafter, "Swedloff".

[6] Swedloff, p. 343.

[7] As Swedloff put it (at p. 346): 'All insurance pools are somewhat heterogeneous with low risks subsidizing higher risk policyholders. Risk classification can remove some of the heterogeneity by putting like risks

wealthy and influential more than it does the unlucky, impoverished and marginalised.[8] Moreover, we have chosen as a society to prevent insurers from pricing risk on a range of personal factors, and that list may need to be revisited when 'Big Data' is in full force. We consider this issue of 'personalised pricing' below in Part B: Big Data and Insurance Risk Classification.

## B. Big Data and Insurance Risk Classification

[9] Ever since the decision of Lord Mansfield in *Carter v Boehm*, the prospective insured has been under a duty to disclose to the underwriter all facts material to the assessment of the relevant risk. Such an obligation is thought to be the most effective way of minimising the impact of information asymmetry upon the classification of risk. The recent reform of insurance law has generally preserved the nature of that duty.

[10] It is not suggested here that the use of Big Data will lead to predictive certainty, eliminating risk, and undermining the viability of insurance as a product. Neither is it suggested that it will result in the reversal of the information asymmetry which has characterised the insurance relationship at the risk classification stage ever since the inception of insurance business, albeit with gradually diminishing effect. Rather, as Professor Peter Siegelman puts it, "the economic theory of insurance suggests that market equilibria are highly sensitive to small changes in underlying assumptions or parameters, so things might look very different if insurers were able to use Big Data techniques to discover more about policyholders' riskiness than the policyholders themselves knew".[9] Insurers may thus come close to knowing as much as assureds do, but the actual benefit of using 'Big Data' lies in fine tuning risk classification by finding new correlations and therefore making better-informed predictions about consumer behavior.[10]

[11] Nonetheless, such fine-tuning might impinge on the protection currently afforded to policyholders. It is not merely privacy that is at stake when it comes to the use of 'Big Data' in insurance, but also confidentiality, transparency, identity and free choice.[11] Following the judgment of the Court of Justice of the European Union in *Test-Achats,*[12] Part 5 of Schedule 3 to the Equality Act 2010 imposes greater controls on the use of protected characteristics (chiefly gender) in

---

together. The more refined the classification scheme, the more homogenous the resulting pools will be, which will then require less subsidization from low risks to high risks'.

[8] The social effects of insurance is a growing area of academic debate, on this point see eg T. Baker and K. McElrath, 'Whose Safety Net? Home Insurance and Inequality' (1996) 21 Law & Social Inquiry 229, examining the insurance claims outcomes for the broad range of employees of the University of Miami following Hurricane Andrew.

[9] P. Siegelman, 'Information & Equilibrium in Insurance Markets with Big Data' (2014) 21 Conn. Ins. L.J. 317, p. 330.

[10] Swedloff, p. 342.

[11] See N Richards and J King, 'Big Data Ethics' (2014) 9 Wake Forest L Rev 393.

[12] C-236/09 *Association Belge des Consommateurs Test-Achats ASBL* v. *Conseil des ministres.*

insurance pricing. Even though such characteristics may correlate with certain types of loss, there are good public policy reasons not to charge higher premiums purely on that basis.[13] Not least, as AG Kokott identified, because that correlation may simply reflect prior social attitudes to protected groups (a norm where husbands drive, with wives as passengers) rather than permanent or desirable differences.[14] Similar (but more contested) claims can be made in favour of regulating characteristics that identify individuals as members of vulnerable groups,[15] characteristics which are undeserved[16] and those that are consequent on bad brute luck.[17]

[12] The harvesting of 'Big Data' might allow insurers to amass a plethora of information from which to discern relevant characteristics of prospective or current assureds which insurers should not otherwise use for the purpose of risk classification. Equally, insurers may collect from a variety of public sources an abundance of private information about their customers which they are prohibited from asking about directly.[18] The danger that comes with the use Big Data is that the algorithms used to process it might simply indicate higher prices for certain assureds without insurers explaining why this is so. Swedloff argues that "insurers may treat the information as propriety and thus have an incentive to conceal the reason for pricing from the policyholder".[19]

[13] While it is not the aim of this paper to make a call for immediate legislative action, it is suggested that the handling by insurers of 'Big Data' might require regulatory oversight. Such monitoring could include inspection of the data sets derived from 'Big Data' as well as the underlying algorithm used. We recognise that this might require expertise not currently held by regulators. An argument could also be made for imposition of a corresponding duty of disclosure on insurers, requiring them to reveal to the prospective assureds what information is being taken into account when pricing their risk.

[14] It may well be that the decrease in the cost of coverage that would result from greater accuracy in risk classification owing to 'Big Data' would be exceeded by an increase reflecting the extra cost of obtaining that accuracy. If so, the issue of potential use of 'Big Data' in the context of insurance pricing might be nothing more than a storm in a teacup.[20] However, it would only be prudent to observe the storm and, if necessary, ensure that it does not get outside the tearoom.

---

[13] Swedloff, p. 360.
[14] See Davey, above n [5].
[15] Ibid.
[16] See T. Baker, "Containing the Promise of Insurance: Adverse Selection and Risk Classification" (2001) 9 Conn Ins LJ 371.
[17] R Dworkin, *Sovereign Virtue* (Harvard University Press, Cambridge, MA, 2000), pp. 73-83.
[18] Swedloff, pp. 368-369.
[19] Swedloff, p. 363.
[20] Inaccurate classification of an assured is not in itself unfair. See K. Abraham, *Distributing Risk: Insurance, Legal Theory, and Public Policy* (Yale University Press, New Heaven and London, 1986), pp. 84-85

### C. Big Data and Insurance Contract Compliance Monitoring

[15] Insurers do not only use information to price insurance products, but to influence the behaviour of those who already have insurance. There are good reasons for doing so: notably, the 'moral hazard' that arises where the insured purchases cover.[21] If uncorrected, the insured would be indifferent to the occurrence of insured losses, and so insurers control for this by imposing excesses, policy limits, and conditions that limit cover unless the insured acts prudently. The difficulty is that the underwriter is not able to actually observe how the insured behaves and is reliant on the use of proxy devices (such as excesses that force the insured to bear some of the cost of losses) or on discovering non-compliance with contractual requirements. 'Big Data' provides the possibility of increased monitoring of behaviour during the currency of policy by significantly reducing the costs of identifying risk-sensitive behaviour.

[16] The **first** of the two possible areas for the use of data analytics is already under consideration within the insurance industry: the detection of fraudulent claims. This has been an area of activity for government, industry and the judiciary in recent years.[22] Insurance frauds are extremely costly to the insurance industry in the UK. The ABI estimated that in 2013 the size of detected insurance fraud was £1.3 billion. At the same time, it is believed that annual undetected insurance fraud is in the region of £2.1 billion.[23] Investigation of potentially fraudulent claims is both complex and expensive. More accurate detection of objectively suspicious patterns might also improve industry-customer relations as aggressive reviews of legitimate claims may lead to unmerited litigation and eventually to insurers losing customers.[24]

[17] The current Insurance Fraud Taskforce, co-hosted by the Treasury and the Ministry of Justice, is due to report on the end of 2015 and is expected to make recommendations in favour of a substantial increase in the sharing of data, across insurance providers, across financial services providers and across the public/private divide.[25] However, 'Big Data' is unlikely to be perfectly accurate in identifying correlations between suspicious behaviour and actual fraud and there is a need to ensure that human systems can properly respond to data sourced across many systems. So, an apparent failure to declare a change of welfare status should not perhaps, of itself, trigger assumptions of dishonesty in an unconnected insurance claim. Underwriters have- at least in their

---

[21] See T Baker, above n 16.

[22] For a critical review of the current position, see J Davey & K Richards 'Deterrence, human rights & illegality: the forfeiture rule in insurance contract law' [2015] Lloyd's Maritime & Commercial Law Quarterly 314.

[23] Such was the estimate of the National Fraud Authority in the two years prior to its closure. See NFA, Annual Fraud Indicator 2012 and 2013.

[24] C Nyce, "Predictive Analytics", White Paper 4, American Institute For Chartered Property Casualty Underwriters and Insurance Institute of America, 2007.

[25] See Ministry of Justice / Treasury Insurance Fraud Taskforce: https://www.gov.uk/government/groups/insurance-fraud-taskforce. The final report is due in late 2015. Professor Davey participated in these discussions as an invited stakeholder.

evidence to courts- shown themselves to be extraordinarily unforgiving of dishonesty in any form. In *Roselodge* v *Castle*, a case concerning the non-disclosure by a company director of a conviction more than 20 years old, McNair J noted that the underwriters '… were driven in cross-examination to state such extreme views that I am unable to accept their evidence on this point. [One] stated that in his view a man who stole apples at the age of 17 and had lived a blameless life for 50 years is so much more likely to steal diamonds at the age of 67 that if he had told him this when putting forward a proposal at the age of 67, he would not have insured him'.[26]

[18] Our **second** application of data as a control mechanism for behaviour during the life of the contract considers the need to minimise moral hazard by monitoring compliance with policy terms. Many insurance policies will require that an approved security system be fitted, and window locks. These 'features' are used to rate the policy as lower risk. However, insureds may then fail to renew security contracts or routinely secure windows when out of the property. The 'internet of things' promises exponential growth in the information available to insurers to detect how we behave whilst insured, and use this information to deny liability for claims that eventuate. This will not mean that people are taking greater risks, merely that the cost of detecting failures is reduced, making them more visible. Law is often not what you know, but what you can prove.

[19] As with fraud, increased monitoring may change the nature of the insurance relationship. Where insurers only had limited information as to our conduct, its ability to establish non-compliance with policy terms was limited to obvious cases (or bad luck). If- as is likely- our everyday failures to act with absolute car and attention is evident from datasets, then insurers will need to develop more human models of 'ordinarily careful' behaviour and write their contracts accordingly.

## D. Conclusion

[20] The role of information in the sensitive equilibrium in insurance markets, and the influence it has on the role of insurers, the State and customers makes 'big data' a clear insurance issue. It is likely to arise slowly- perhaps with the growth of 'black boxes' in cars to monitor driving patterns and improved detection of suspicious claims. However, the extended use of data might diminish the real strength of insurance to provide cover across communities, whereby the lucky subsidise the unlucky, often without even realising they are doing so. Whilst the redistributive effects of insurance might only be of tangential interest to insurers and those interested in markets, it has a social significance that ought not to be overlooked.

[26] [1966] 2 Lloyd's Rep 113, 132. This (and similar cases) is discussed in J Davey, 'Materiality, Non-Disclosure And False Allegations: Following *The North Star*?' [2006] LMCLQ 517.