

Big data, pooling risks and exclusion in the insurance industry

Rémi Steiner,

Engineer from Corps des Mines, Conseil Général de l'Économie

Abstract:

In a market economy, the theory of insurance leads insurers to set the premiums for policies as a function of the risks specific to policyholders. A hard question thus crops up: how to segment risks (an action reducing the effects of pooling risks) so as to keep clients from taking their business to competitors? Answering this question involves big data, a phenomenon referring to: the exponential growth of the data collected (in particular via connected devices), the opening of public data, the nearly unlimited storage facilities for data, and the powerful statistical tools for processing them. The “right” use of big data is a key to the insurance industry’s digital transformation. It might lead to “behavioral insurance” policies. Sensitive questions arise about how personal data are to be used. Big data can help limit the “moral risk” and “asymmetry of information” in insurance, but it might also lead to unacceptable forms of discrimination, which would eventually necessitate laws and regulations. Will big data increase the number of persons excluded from coverage under insurance? The opposite might be possible (and would be desirable).

The insurance business essentially relies on processing statistical data. There is no doubt that the quantitative explosion of electronic data and the improvement of the techniques that make sense out of these data are major catalysts of change. Insurers are facing the digital transition: big data are a part of this trend. The disruption of distribution channels, the dematerialization of contracts and the automation of operations are major technological and commercial challenges (VILLANI 2018).¹

Improved data processing during all phases in the life of an insurance policy is likely to deeply affect the insurance business and open new areas of competition. Big data could lead to making more relevant selections and setting fairer rates for the risks covered by a policy. The analysis of data from vehicles, health sensors and others connected devices could boost, during the duration of an insurance policy, exchanges of mutual benefit to both the insurer and policyholder. It could also result in better risk prevention and additional, appreciable services. Big data are also likely to improve the fight against fraud as well as the identification of the recipients of unclaimed benefits.

¹ This article has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor’s approval, completed a few bibliographical references.

Pooling or selecting risks?

Pooling risks, the foundation of insurance

Pooling risks underlies the insurance business: a group of individuals, subject to the same risk, might have an interest in pooling coverage for the risk. They prefer paying a fixed premium instead of having to pay more for the eventual damages corresponding to the risk. Like their insurance company, these clients make the following unspoken assumptions: that the probability of a disaster or accident is stable over time, that they all members of the pool have more or less the same degree of exposure to the risk, and that their individual risks are relatively independent of each other. Each insured party willingly accepts to pay a premium without receiving anything in return for coverage against a risk that does not occur one year because, the next year, his/her turn might come to be hit by an occurrence corresponding to the risk — in which case, he/she will be protected thanks to this pooling of risks. The insured parties belonging to this pool bundle, we might say, their interests. If some of them, contrary to expectations, turn out to be excessively vulnerable, imprudent or negligent with regard to the covered risk, the premiums of all persons in the pool might be jointly raised.

Competition is forcing insurance companies to select risks

The solidarity between the insured parties could be absolute in a world where the insurance business constitutes a monopoly (as in compulsory health insurance in France). The insurer benefitting from this monopoly would be able to accept all clients, regardless of their *ex-ante* probability of risk; and impose the same premium rates on all of them, not withstanding differences in their risk probabilities. Joint and several liability for the risk would apply fully.

In a situation with competition however, when a subpopulation is identified as having a lesser risk in relation to the event to be covered, a new insurer will tend to offer its members an advantage of mutual benefit (*i.e.*, more favorable conditions). If all insured parties act in full compliance with their interests, the clients with a lesser risk exposure will switch to the new company. As a consequence, their former company will end up covering a group with higher risks than initially expected. Unless it accepts to operate at a loss, it will be forced to increase its premiums.

Pooling and selecting risks are not antagonistic however. On the contrary, they are necessarily mixed. In a market economy, an insurance company segments its customer base as a function of its assessment of various levels of risk, and applies different premiums to different clients depending on coverage provided by the policy. Furthermore, insurers are allowed to purely and simply refuse a request for coverage.

The difficulties of segmentation

For a slue of reasons, segmentation is a difficult art (CHARPENTIER & BARRY 2015, CHARPENTIER *et al.* 2015). Even supposing there are relatively stable traits or behaviors with a high predictive power, the insurer would have a hard time obtaining information about them. For practical reasons, the decision to classify a person asking to subscribe to a policy in a given segment is normally based on simplistic data (age, place of residence, etc.). As a result, the segments cannot be homogeneous. Even when a certain parameter strongly affects the level of claims in relation to the covered risk, it might simply be a correlation instead of a risk factor as such. For example, age is a far from perfect indicator of the probability that a driver will be careless or inexperienced.

Moreover, there is usually an asymmetry of information: insured parties have no reason to give the insurer information that would identify them as having higher risks, since this would justify a decision to make them pay higher premiums. This asymmetry might have an “antiselection” effect.

If an insurer distinguishes clients poorly from each other and applies the same premium to all, the persons with a lower exposure to the risk, aware that the conditions imposed on them are rather onerous, might prefer either forgoing insurance or subscribing a policy under better conditions with a company that has more foresight. At the rate of premiums set, the only customers who would be left would be those with the most exposure to the risk. The insurer’s assessment failed to take this into account when setting rates for premiums. In a pioneering article, George Akerlof (1970) has shown how such asymmetries of information can keep supply and demand from matching and thus forestall the emergence of a market.

How big data might change this situation

New prospects

The “big data” phenomenon takes various forms. The phrase itself mainly brings to mind the volume of data along with the capacity for computer processing with new possibilities for statistical analyses. Of course, proliferating data and increasing computational power cannot magically lift the theoretical difficulties of segmenting a customer base or validating a schedule of individualized rates for premiums. Nonetheless, there is considerable room for progress as the access to public data is broadened. For instance, data on automobile accidents or on health are underused because they are currently hard to obtain. If made anonymous whenever necessary, they could be made more available for analysis and thus deliver information of major interest not only to insurers but also for the purpose of risk prevention.

Meanwhile, the EU’s General Data Protection Regulation (GDPR) is setting up a new regulatory framework for personal data.² Under the previous system in France, insurance companies (like other firms) were subject to an *ex-ante* declaration and authorization of the data-processing actions that they wanted to perform. The National Commission on Informatics and Liberty (CNIL) closely oversaw the choice and use of the personal data necessary for managing insurance policies and subscribing to them, whether the data were to be processed for studies of clients’ specific needs, for the assessment, control and monitoring of the risks covered, or for the purpose of managing policies from the precontractual phase till termination or cancellation.³ As of 25 May 2018, the new European regulation has put an end to these *ex-ante* arrangements and replaced them with a more flexible system of *a posteriori* self-regulation and control. This change is likely to increase the freedom for insurance companies to collect and process personal data under their own responsibility and with important safeguards.

² The GDPR (General Data Protection Regulation): “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data”. Available via: <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1478961410763&uri=CELEX:32016R0679>.

³ CNIL: Commission Nationale de l’Informatique et des Libertés, created by the 1978 law “Informatics and Liberty”. On insurance companies, see CNIL (2014), *Pack de conformité assurance* and the CNIL’s decision n° 2013-212 of 11 July 2013 on the automated processing of personal data available at: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027838440&categorieLien=id>.

The deployment of big data could be spurred by the rollout of sensors and connected devices that objectify the person's behavior patterns before he/she takes out insurance and then during the duration of the policy. Everyday devices (vehicles, smartphones, connected watches and household appliances) are increasingly recording data about behaviors behind the steering wheel, during sports, etc., not to mention physiological or geolocation data. Insofar as such personal data objectively affect the insurer's assessment of the exposure to the risk to be covered by a policy, and on condition that the insured party consents, this data could enter into the processes whereby insurers assess risks and set premiums.

From segmentation to behavioral analysis

Initial experiments have been conducted. For group coverage of their employees under complementary health insurance, Generali France is offering firms Generali Vitality, a policy with incentives for wage-earners to improve prevention and hygiene: checkups, screening, vaccinations, personalized physical activity goals (measured by electronic bracelets), participation in sporting events, pledges to not smoke, daily monitoring of diabetes, etc. Allianz Conduite Connectée offers an automobile insurance policy based on a "driving score" that uses data, collected via a smartphone during a test period, about: geolocation, number and intensity of instances of acceleration and of braking, cornering, dates and times of departure and arrival via GPS-plotted routes. If a driver takes out a subscription after this test period, a smart device for collecting this information will be permanently installed in the vehicle.

"Behavioral insurance" policies will spread since they create the conditions for the interests of the insurer and insured parties to converge. Such policies ward off the "moral hazard" of insurance — the deviant effect of classical policies that arises when the insured alter their behaviors and take fewer precautions to protect themselves because they are insured. Carefully calibrated, behavioral insurance policies hold promise as a means for averting moral hazard. In a reversal of the usual situation, the signature of an insurance policy would induce an individual to do more to prevent the risk being covered. Were this promise to prove true, the general interest would instruct us to foster such policies. In France however, shocked comments have been aimed at behavioral insurance policies; they underscore the risk of insurers excessively meddling in the policyholder's private affairs.

Reducing the asymmetry of information

Should there be restrictions barring insurers from knowing too much about their clients? The question is not new, but the upsurge of big data has honed its pointedness. Is it not legitimate, and in the common interest, for both parties to enter into an engagement with full knowledge — with complete, objective information about the risks taken by each? In comparison with yesterday's world (where information was scarce, the insurer was located far from the insured, risk policies were opaque, and dissimulation was easy), the current abundance of information enables both parties to conclude in good faith a contract under the best conditions for building trust.

All the data and statistics of the world will never make uncertainty about future events disappear, but big data will lift at least a corner of the "veil of ignorance" that obfuscates the perceptions of insurers and the insured. Both parties would be better able to assess risks, and the asymmetry of information between them would be lessened. In tomorrow's world, the need for insurance will not vanish, but the insurance market could become more competitive and transparent (BERBAIN & SALAMANCA 2015, DUMORA 2018).

Big data and exclusion from insurance coverage

Concerns about a more selective marketplace

For reasons already explained, this more competitive, transparent marketplace will probably also be selective. Some applicants for policies will be refused, or will be extended partial coverage, or will have to pay higher premiums (FROMENTEAU *et al.* 2011). This might happen in two distinct situations, each of them raising a completely different set of questions about basic principles.

The first situation has to do with the moral hazard of insurance or the very low risk aversion of some policyholders, *e.g.* those who accumulate traffic violations, who take no precautions to avoid burglary or who pay no attention to their health (for reasons of personality and comportment). The insurer cannot always detect this sort of situation, which carries a risk much higher than average. In extreme cases, some policyholders can be said to be free-riders, who draw unwarranted benefits from their coverage to the detriment of both the company and the other insured parties. Questions arise about whether this type of risk should be pooled. Identifying situations of this sort is in the general interest.

A completely different situation arises when the higher risk is an intangible fact, a specific characteristic over which the applicant has no hold (*e.g.* a pre-existing health condition or exposure to a natural catastrophe). Refusing insurance to such applicants or restricting coverage can be deemed outrageous. *Ad hoc* regulations might be adopted to cope with this situation, or insurers might willingly forgo using certain risk assessment criteria. In effect, the adoption of regulations and concerns by insurance companies about their reputation being sullied are two factors that tend toward pooling risks in this sort of situation. The insured more or less tacitly accept to pay higher premiums as a counterpart to the absence of discrimination in relation to this sort of risk.

Laws and regulations

In many cases, regulations compel insurance companies — and the insured — to pool risks. An example is home insurance coverage for natural catastrophes, a legal requirement under Article L. 125-1 of the French Insurance Code. Another example arises under the Evin Act,⁴ which stipulates that, when employees are covered as a group under a complementary health insurance policy, the insurance organization cannot refuse to reimburse or cover the expenses resulting from an illness to the persons covered by the contract who want to maintain coverage. Nor can it reduce the coverage corresponding to the premium set for the category to which the insured person belongs. Nor can it later increase the premium for reasons related to the insured person's state of health. If the insurance organization wants to raise the premium related to a type of guarantee or contract, the increase must be the same for all the insured who have subscribed to that type of guarantee or contract.

Another case has aroused much more discussion: the pooling of risks by sex when statistics clearly indicate a difference. EU directive 2004/113/CE of 13 December 2004, which establishes the principle of equality in the access of men and women to goods and services, granted member states the possibility to allow exonerations from a strict application of this principle in insurance. However the Court of Justice of the European Union (CJEU) has decided that setting different premiums for men and for women is a form of discrimination based on sex and is prohibited by the Charter of Fundamental Rights of the European Union.⁵

⁴ Specifically Article 2 of Act n° 89-1009 of 31 December 1989 on the guarantees extended to persons insured for certain risks. Texts of French law are available at: <https://www.legifrance.gouv.fr/Droit-francais>.

⁵ Judgment of the Court (Grand Chamber) of 1 March 2011. Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministres. Available via: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A62009CJ0236>.

Big data and high-risk health conditions

A major social issue is life insurance coverage for persons with increased health risks. Since banks require that borrowers subscribe to life insurance, these persons have fewer possibilities for obtaining a loan to buy a house and become a homeowner. Will big data ostracize part of the population by shamelessly exposing life expectancies?

Despite our legitimate worries about future trends, we must admit that this sort of situation already occurs. According to French statistics for 2016, collected under the AERAS convention (about increased health risks when taking out insurance or a loan),⁶ 514,449 out of 3,447,038 requests for life insurance — 15% — concerned persons with higher health risks. In more than two out of three of these cases and following additional examinations, the proposal was made to insure the person under normal conditions. Nonetheless, 133,025 applicants — 4% of the total — received a proposal with a higher premium. In half these cases, this premium was 50% higher; but it amounted to more than 300% in 2,117 cases. In addition, life insurance was simply refused to 18,046 applicants. In 2,569 cases, a proposal was made without a higher premium but with limitations on the coverage provided. For undetermined reasons, applicants left 89,664 applications in abeyance. We assume that self-censorship led an indeterminate number of persons not to file an application. Restricting or eliminating access to insurance is far from unordinary.

Under pressure from associations of patients and from public authorities, Article 190 of the Act of 26 January 2016 on the modernization of the health system has enshrined the principle of a “right to be forgotten”, in particular for persons who have had cancer. This right, under articles L. 1141-5 and L. 1141-6 of the Public Health Code, is to be gradually extended to other pathologies (in particular chronic illnesses) whenever advances in therapy and scientific data attest that a treatment can significantly and lastingly limit the pathology’s effects.⁷

Much progress is yet to be made to improve the access to insurance coverage. It is clearly possible that the effects of big data — owing to the combination of: an improved statistical processing of public and private data, a more transparent insurance market and greater competition — will lead insurers to be less timorous and thereby decrease the number of exclusions from coverage. In any case, this is the objective that should continue guiding public authorities’ interventions.

References

- AKERLOF G. (1970) “The market for ‘lemons’: Quality uncertainty and the market mechanism”, *Quarterly Journal of Economics*, 84(3), pp. 488-500.
- BERBAIN C. & SALAMANCA E. (2015) *L’assurance de demain. Reconnaitrons-nous notre assureur en 2030?*, thesis submitted to the Corps des Mines, July 2015.
- CHARPENTIER A. & BARRY A. (2015) “Passer d’une analyse de corrélation à une interprétation causale”, *Risques*, 99.
- CHARPENTIER A., DENUIT M. & ELIE R. (2015) “Segmentation et mutualisation, les deux faces d’une même pièce?”, *Risques*, 103.
- DUMORA R. (2018) “Les nouveaux chemins de l’assurance” in J.B. Mateu (ed.), *Les banques face à leur avenir proche* (Paris: Eyrolles), pp. 265-288
- FROMNTEAU M., RUOL V. & ESLOUS L. (2011), “Sélection des risques: où en est-on?”, *Les Tribunes de la Santé*, 31.
- VILLANI C. (2018) *Donner un sens à l’intelligence artificielle. Pour une stratégie nationale et européenne*, mission parlementaire, March, 242p. Available via: https://fichiers.acteurspublics.com/redac/pdf/2018/2018-03-28_Rapport-Villani.pdf.

⁶ Federation Francaise de l’Assurance, *Convention AERAS: statistiques 2016*.

⁷ AERAS, “S’assurer et emprunter avec un risque aggravé de santé”, convention of 6 July 2006, with amendments, annual statistics and a reference grid in the appendix. The grid was last updated on 30 March 2017.