Organisation for Economic Co-operation and Development

DAF/AS/PEN/WD(2018)6

For Official Use **English - Or. English**

18 April 2018

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
INSURANCE AND PRIVATE PENSIONS COMMITTEE

Working Party on Private Pensions

Big data and annuities

Working Party on Private Pensions, June 4-5, 2018

This document is circulated for discussion under the agenda of the WPPP meetings.

Jessica MOSHER [Tel: +33 1 45 24 91 28; Email: Jessica.Mosher@oecd.org]
Pablo ANTOLIN [Tel: +33 1 45 24 90 86; Email: Pablo.Antolin@oecd.org]

JT03430420

Big Data and Annuities

Background and motivation

1. Recent OECD work on has highlighted the potential benefit of annuities that are more tailored to the needs of specific segments of the population, and in particular different socioeconomic groups (OECD, 2016[1]), (DAF/AS/PEN/WD(2017)14). However, often these types of targeted products lack supply or come at a relatively high cost.

2. Enhanced annuities, which offer higher payments for individuals with lower life expectancies due to health or lifestyle factors, could provide a more affordable solution for low income groups to hedge their longevity risk in retirement. Yet these products are not widely available. One potential explanation for the lack of availability of enhanced annuities is the difficulty in determining the underlying mortality assumptions for a particular subset of the population due to the lower amount of data available at a granular level. This results in much higher uncertainty around the true expected mortality, making the expected time that payments will be made more uncertain and exposing the annuity provider to increased levels of longevity risk. This increased risk comes at a cost in terms of higher required capital for the annuity provider and a lower income for the pensioner (OECD, 2016[2]).

3. Products offering more flexibility around investment options and access to the underlying capital, such as the variable annuity products common in the United States, could be attractive to those with higher levels of financial literacy. However, this flexibility requires the annuity provider to make assumptions regarding the behaviour of the individual. Behaviour is very difficult to model and predict, and getting these assumptions wrong can be quite costly to the annuity provider. As a result these types of products tend to be rather expensive (OECD, 2016[2]).

4. Assumption setting relies on data analysis, and as such the increasing prevalence of big data offers a rich source of information for annuity providers on which to base these assumptions. Big data refers to large and complex sets of data typically characterised by their *volume* (size of data set), *variety* (type of data, e.g. text, image) and *velocity* (speed of data generation). The advances in technology that allow for increased ability to access, store and analyse big data - both internal and external data sets - represent a huge opportunity to increase the analytical capacity of annuity providers to assess the underlying assumptions required for annuity products and the resulting risks to which they are exposed. Increased precision in mortality estimates and behavioural assumptions could allow for products which more accurately target specific sub-groups of the population. It could also improve product design and distribution to increase the likelihood that the segments of the population who could benefit most from certain products would access them.

5. Nevertheless, the increased segmentation of the population that big data allows for also raises questions as to how much segmentation is desirable for an insurance product, and what we, as a society, would consider to be ‘fair’. While the annuity provider would likely be better able to reduce its overall risk exposure, the use of an increased number of variables in determining segments would also reduce the amount of risk pooling and cross-subsidization across different groups. The use of complicated algorithms could also lead to less clarity and transparency as to the relevance of the variables used and how they are influencing pricing.

6. The objective of this note is to discuss both the opportunities that big data presents for improving the provision of targeted annuity products as well as the challenges that advanced analytical techniques exploiting big data could introduce. The first section explores the potential use of big data for annuity products. The second section discusses the implications big data could have on the assumptions used for pricing annuity products. The third section considers the implications that big data could have for the distribution of annuity products. The final section summarises with some concluding remarks.

The potential of big data for annuities

7. Insurers have already begun using big data to improve their business processes. While property and casualty insurers have been more advanced in its use compared to life and annuity providers, a majority of the latter have plans to use big data namely for areas relating to risk (including product development, underwriting, and risk management) and product marketing/distribution. The sources of data that are being considered are largely related to health and financial situations, as expected, though insurers are also considering using data from social media and purchasing behaviour (Breading & Garth, 2014).

8. Life and annuity providers are beginning to use big data to better inform their products and risk exposures. Wearables, such as Fitbits or activity tracking apps, are able to provide real-time feedback regarding an individual’s health and activity. These have been used to incentivise individuals having life or health insurance policies to improve their health or be physically active in order to receive discounts on their insurance premium (Edwards, 2017[3]). As the risk insured with annuity products is that of living longer than expected, big data cannot be used to provide incentives for individuals to reduce their risk, as this would imply that they would increase their chance of dying. Nevertheless, there are opportunities for big data to improve the assessment of the risk profile through improving mortality estimation, and thereby offer higher annuity payments to those who are disadvantaged with respect to their shorter life expectancy.

9. Some progress is being made on using big data to improve estimates of mortality, though this seems yet to be used in practice for annuity products. The European Commission is investing in the CompBioMed initiative, which aims to improve predictive models of health and disease – and by extension, mortality – based on human biology in order to improve healthcare. Another research programme, funded by the Institute and Faculty of Actuaries (UK), is basing its research on the The Health Improvement Network (THIN) primary care database along with mortality data from the Continuous Mortality Investigation (CMI) (Kulinskaya and Gitsels, 2016[4]). Its objective is to improve the modelling of mortality and trends in mortality for the UK population and different subgroups of the population. It will also model trends in morbidity and uptake of different medical interventions, which can differ across socioeconomic groups and therefore has implications for how quickly medical advances will impact the

mortality of lower socioeconomic groups in particular. The increased availability of genetic data could also improve mortality estimates from common risk factors already being used in practice. Not all smokers, for example, are equally susceptible to developing lung cancer.

10. As for other types of assumptions, improved predictive analytical methods are already being used to refine behavioural assumptions for annuity products. Nevertheless, much of this analysis currently seems to rely on in-house data or aggregated industry data. As insurers develop their expertise and capabilities, however, external data sets could be used to refine these models even further, potentially in combination with neural nets.[1] (Campbell et al., 2014[5]).

11. One of the leading potential applications of big data relating to marketing and distribution activities of life and annuity providers is the improved segmentation of distribution channels (Breading and Garth, 2014[6]). Big data could allow providers to better target products to specific demographics or geographical areas and rely on channels which are the most effective for a given segment, which could improve consumer uptake of annuity products across all segments.

12. Nevertheless, just because the data is now available for use does not mean that it should necessarily be used for all purposes. The refinement of product pricing and distribution using more variables to segment the population presents both significant benefits as well as potential challenges relating to providers' risk exposure and access to insurance. The following section discusses these issues in detail.

The implications of big data for pricing annuities

Segmentation and assumption-setting

13. The use of large amounts of data to predict risk and segment the population according to different risk profiles is certainly not new for insurers and annuity providers. They are in the business of pricing and managing the risk for their clients, and they need data in order to assess the risk as accurately as is possible. Big data does, however, increase the possibilities regarding the variables which can be used for segmenting risks and the consumer base.

14. When it comes to setting assumptions used for pricing, annuity providers have two main motivations for refining the segmentation of the consumer base. First to improve the accuracy of the assumptions, thereby reducing the risk exposure of the annuity provider. Second, to reduce the level of adverse selection by consumers through more accurate pricing of the annuity to reflect the specific risk profile of different segments.

15. With respect to risk reduction, more refined segmentation of risk implies that there will be a lower volatility of experience around the true mean. A population consisting of both individuals with chronic illness and perfectly healthy individuals will experience a much wider range of mortality experience than a group of healthy individuals who all consistently maintain regular exercise and good diets. Splitting these groups into two will result in segments with less variability around the true mean within each group.

[1] Neural nets refer to models that rely on machine learning and are able to incorporate both linear and non-linear variables for the analysis.

16. Big data may also allow for a more accurate estimation of the true mean for different segments of the population. One challenge for insurers in segmenting the population is that the more their in-house data is divided up into segments, the less data is available on which to establish assumptions. This reduces the certainty around those assumptions because of the smaller number of data points on which they are based. The use of external big data could improve the certainty by increasing the amount of data available by drawing from external sources as well as increasing the number of variables used for estimation. For example, estimating the mortality for someone with diabetes could also take into account the type of treatment followed.

17. Big data could also improve the accuracy of more dynamic assumptions, such as those relating to how consumers will behave. Consumers' decisions to terminate the contract, withdraw funds, change investments, or take annuity payments can depend on a multitude of external variables relating to the economy and their financial circumstances. Such assumptions are rather difficult to determine using traditional actuarial techniques. Here again, the analysis of big data could help to establish more accurate assumptions which take many other variables into account and react to changes in the environment in a dynamic way.

18. The ability to dynamically assess changes in the variables that are used to set assumptions could also lead to better real-time risk assessment. This would allow the annuity provider to better control its risks from existing books of business and potentially could lead to new types of products. For example, if an annuity provider decided to begin offering enhanced annuities, they might be better able to determine the impact this would have on the mortality of their existing annuity business going forward. The ability to dynamically underwrite annuity holders could also lead to new products which could change payment levels, for example products which increase payments for individuals whose health deteriorates (IFoA, 2017[7]).

19. More accurate assumptions and improved risk assessment could lead to a lower cost for annuity products. The margins that annuity providers include in assumptions to account for their inherent uncertainty could be reduced, thereby lowering the price of the product and the calculated reserve requirements. In addition, lower risk exposure means that risk-based capital requirements could be reduced, resulting in a lower cost of capital for the annuity provider. The use of big data to verify the status of policies, i.e. whether the individual is still alive or has been reported dead, could reduce the impact of unreported claims, further lowering costs to annuity providers. If these savings are passed on to consumers, this should result in higher annuity incomes being offered from annuity products.

20. Having more accurate pricing for specific segments of the population not only allows annuity providers to better control risks for existing business, but also helps to control their risk exposure from those who will purchase the annuity product in the future. Indeed, preventing adverse selection is one of the main reasons that insurers propose different prices to different groups of people.

21. Adverse selection refers to the tendency for high-risk individuals to purchase insurance while low-risk individuals will opt-out if the price is higher than the expected value of the payments they will receive from the insurance. In the case of an annuity, the high risk individual is the one that expects to live a long life. If given the choice, a person with a chronic illness would not be willing to pay the same price for an annuity as a healthy person. If the annuity is priced based on the average mortality of these two groups, only the healthy person would purchase the annuity and the annuity provider

would have to pay annuity payments longer than expected based on the average life expectancy of the two groups. If different prices are offered to these two groups, both will pay on average the amount they expect to receive in return.

22. By allowing for prices that are more reflective of the actual risk of a specific segment, segmentation reduces the cross-subsidies from low-risk groups to high-risk groups. In the case of annuities, without segmentation, low socioeconomic groups with lower life expectancies would be subsidizing the annuities of wealthy individuals who can expect to live longer.

23. As such, segmenting the pricing of annuities between those with lower life expectancies and those with high life expectancies is intuitively appealing because the former also tends to correspond with more vulnerable and disadvantaged groups of the population. This allows these groups to receive a higher income from their annuity. However, in most types of insurance, these groups tend to be considered to be more high-risk, and therefore segmentation results in higher prices for them rather than lower prices. This is the case for life insurance and health insurance, for example. So increased segmentation does not always benefit more disadvantaged groups, and there may be some benefits for society from having a certain level of cross-subsidisation of insurance policies. In addition, insurance at its core is based on the pooling of risks, so the business model itself assumes some cross-subsidisation.

24. The conflicting aims of exact risk profiling and the pooling of risk across society give rise to the question of how much segmentation is desirable in insurance, and what, if any, limits should be imposed. On one extreme, it can be argued that even if segmentation is pushed to its limits with big data, idiosyncratic risk is still being insured (Insurance Europe, 2017[8]). That is to say even if an individual is offered an annuity whose price is completely personalised, the exact date that they will die is not known so the annuity provider is still providing insurance against their longevity risk. At the other extreme, politicians have already passed legislation limiting the variables on which segmentation can be based, most notably in Europe where insurance pricing by gender has been banned since 2012 on the grounds of gender equality and non-discrimination. Both positions clearly have consequences with respect to the ability for annuity providers to manage their risk and the cost of annuities for individuals in society.

25. Limiting the level of segmentation that annuity providers are allowed to use will have consequences in terms of adverse selection where the purchase of an annuity is voluntary. This is especially true where the differences in risk profiles are more obvious, as with gender. The fact that women live longer than men on average is well known, so generally speaking men should not be willing to pay the same price for an annuity product as women. If only women purchase annuities, the price of the annuity would ultimately be pushed up to that for women only. There has been some evidence that this has occurred in practice (OECD, 2016[2]).

26. Therefore where adverse selection is probable, increased segmentation can be welfare-enhancing as a whole. The example with gender shows that disallowing pricing by gender could exclude males from the annuity market, whereas allowing this price discrimination allows males to purchase an annuity in line with their life expectancy with little to no additional detriment to females, who would pay a similar price as they would if pricing by gender was not allowed. This would also be true for other [primarily health-related] variables where differences in life expectancy are more obvious: smoking, chronic disease, obesity, etc. This observation supports the case for encouraging the

development and supply of enhanced annuities to ensure that those with lower life expectancies have access to annuity products.

27. There are cases where the market is segmented but where adverse selection may be less obvious. In the United Kingdom, for example, annuities are commonly priced differently according to post code. This essentially represents a proxy for socioeconomic status; those living in more wealthy neighbourhoods are expected to live longer than those in poor neighbourhoods. The potential for adverse selection here comes in part from the correlation of socioeconomic status with the health-related variables mentioned earlier. Those with lower education, for example, are significantly more likely to smoke than those with medium or high education (Mackenback, 2016[9]).

28. However, even when statistically justified, such proxies could potentially lead to discrimination based on variables which are not generally accepted as fair, such as by ethnicity. In the United States, for example, neighbourhoods are highly segregated by race (Cable, 2013[10]). Zip code could therefore easily be interpreted as a proxy for race. The argument of preventing adverse selection may not hold with such variables, as race is not necessarily a key driver of differences in mortality. Rather, race can be correlated with socioeconomic advantage or disadvantage. The following section discusses the potential for the increased use of big data and algorithms to result in unfair discrimination.

The potential for discrimination

29. The potential for proxy variables correlated with the true drivers of differences in mortality to result in unwanted discrimination is not new. However, big data presents a much larger potential to identify new proxy variables which indicate correlation rather than causation of mortality differences. Furthermore, combined with the use of artificial intelligence and machine learning, the variables that are being used to determine these differences could become much less transparent, increasing the risk that discriminatory pricing could go unnoticed or be difficult to prove.

30. A big challenge of using artificial intelligence and machine learning is its capacity to learn human biases and stereotypes that are reflected in the data used to train the algorithms (Osoba and Welser IV, 2017[11]) (Caliskan, Bryson and Narayanan, 2017[12]). Rather than being neutral, such algorithms learn from historical discrimination and perpetuate that bias in the calculations and decisions going forward. Unjustified racial and gender stereotypes could therefore be enforced with learning algorithms. This could be a particular risk for insurance if the training data includes historical pricing outcomes. Underwriting tends to be seen as part science and part art, and rate adjustments have often included a certain level of human judgement which could be influenced by individual bias.

31. Another potential issue with using machine learning algorithms to price insurance is the increased risk of error for populations which are underrepresented in the training data (Osoba and Welser IV, 2017[11]). Therefore to the extent that annuity data is relied upon, markets which have historically had lower access to the annuity market - potentially the lower socioeconomic groups which could benefit the most from increased segmentation in pricing - would have a higher risk of being mispriced.

32. Given the high risk that algorithms could reinforce existing discrimination and biases that could lead to inequalities, it may be necessary to establish clear guidelines regarding how big data and algorithms can be used for pricing insurance more broadly

and annuity products in particular. The main questions that need to be addressed are what types of variables should be acceptable to use for segmentation and how the algorithms will be audited and supervised.

33. Answering the question regarding the variables that should be allowed to be used for segmentation is not easy. The stance of the European Supervisory Authorities is that the entities should at least be able to justify the variables they are using (Joint Committee of the European Supervisory Authorities, 2016[13]). The UK Institute and Faculty of Actuaries proposes that data sets used need to be "relevant, accurate, appropriate and consistent" and institutions need to ensure that the customers' needs are put first (IFoA, 2017[7]). These guidelines, however, are rather open to interpretation and it is not obvious what criteria should be used to prove that a variable is justified.

34. One justification to use certain variables could be the demonstration that the criterion has a direct causal link with the risk in question. However, the identification of casualty can be difficult even when the statistical analysis supports a strong relation. For example, there are clear differences in mortality by race in the United States. However, controlling for race shows that these differences are not necessarily biological, and may be more driven by socioeconomic disadvantage relating to access to healthcare and education. Differences could also be driven by observed historical patterns of bias and discrimination, such as the treatment that different individuals receive for the same medical condition which could result in differences in mortality.

35. The ability to identify causal links may be becoming more difficult rather than easier with the use of algorithms and big data. The use of algorithms to analyse large sets of data may pick up variables which are only correlated with the actual causal link rather than identifying the true causal factor. There are efforts being made, however, to incorporate causal reasoning in machine learning algorithms to address this problem and improve the auditability of such algorithms (Osoba and Welser IV, 2017[11]). Such improvements could potentially help to mitigate the risk that algorithms reinforce and perpetuate existing human bias and discrimination.

36. However, the acceptability of the use of certain variables is not obvious even where direct causal links are very probable, such as with the results of genetic testing. Some European Union members have allowed genetic information to be used only with consumer consent, while others have moved to restrict the use of predictive genetic testing (EIOPA, 2017[14]). In the case of annuities, the use of genetic information could have significant consumer benefit in the form of higher incomes from improved pricing of enhanced annuity offerings. Nevertheless, it would be difficult to allow for more segmentation for some insurance offers than for others, and increased segmentation could prove detrimental in the case of health insurance and could lead to certain individuals becoming uninsurable.

37. The challenges in identifying causality and the increasing opaqueness of the algorithms used to analyse big data present large challenges for the audit and supervision of pricing processes. In this sense, algorithms may not facilitate following the logic of pricing decisions, and determining whether the outcomes are biased may be no easier than judging the outcomes from human decision making.

38. While big data presents an opportunity for improving the accuracy of pricing risk and enhancing the product opportunities for annuities, it also brings many challenges and open questions as to how it should be used in practice. Similar opportunities and

challenges exist for the use of big data for the distribution of products, which are discussed in the following section.

The implications of big data for product distribution

39. Big data has the potential to improve product distribution by allowing the annuity provider to better target the appropriate markets and facilitating the product application and underwriting process. Nevertheless, as with the pricing of risk, big data also presents a risk that discriminatory practices could be facilitated.

40. The potential for big data to lead to a better match of products and their target markets through the analysis of consumer needs and preferences is positive in the sense that the needs of certain segments could be better met both in terms of product features and product uptake. Indeed, the MiFID legislation in Europe requires that financial service providers consider the general suitability of a product for the market that is being targeted. Analysis of big data could assist providers in assessing which markets to target with which products. This could help to meet the different needs of underserved demographics by improving the awareness of annuity products and highlighting their benefits in a way that is useful and appealing to that particular demographic. It could also inform distributors regarding the best channel to reach underserved consumers and get them to purchase the product. For example, some segments may be better accessed via the internet or social media, while others would prefer to meet with the distribution agent in person.

41. The use of big data to assess risk factors could also help to facilitate the process through which consumers purchase annuity products. With such improvements in risk assessment, consumers could provide some key information in a simplified questionnaire rather than having to complete a full medical exam during the underwriting process. Simplifying the process could improve the uptake of annuity products for those with lower financial education who could otherwise be more reluctant to purchase a 'complex' financial product such as an annuity.

42. Nevertheless, using big data assessment to inform distribution will only be as good as the data used to calibrate and feed the algorithms. If the data available is largely based on consumers who are very digitally connected, the needs of individuals who are less digitally connected could be missed, and this demographic would be left with the generic product option (OPSG, 2017[15]). This could be a particular risk for annuity products targeting elderly populations. Another concern is that segments could be artificially created based on consumer behaviour or sales incentives (FECIF, 2017[16]). Any bias in sales practice could be enforced, for example if sales agents generally tend to advise females to purchase more conservative products.

43. The use of big data could also increase the information asymmetry involved in the sale of an annuity product in favour of the annuity provider. This could manifest itself through higher prices offered to individuals who are more risk adverse, for example, or discouraging individuals from purchasing products which they would benefit from but which could be less profitable for the annuity provider.

44. Finally, while an increased personalisation of products could be beneficial for consumers, there will always be a trade-off between personalisation and comparability. Consumers may face increased difficulty comparing the price of different annuity products offered by the various providers as products are increasingly tailored to individual circumstances.

Concluding remarks

45. The use of big data for improving risk assessment and distribution is gaining traction with life and annuity providers, who are increasing their technical and analytical capabilities in order to be able to take advantage of the potential benefits that it can bring. For annuity providers, big data can expand the basis on which assumptions can be made, increasing their accuracy and granularity. It could also help to make distribution efforts more effective by better targeting the population segment that a particular product is aimed at through the most effective channel to do so.

46. Improving the accuracy of pricing assumptions - in particular mortality assumptions - for more granular segments has benefits that are twofold. First, it reduces the annuity provider's risk exposure coming from the uncertainty in the underlying assumptions. Secondly, it reduces the risk of adverse selection, or the risk that only individuals who expect to live longer will purchase the annuity. This would ideally lead to lower margins charged to the consumer by the annuity provider and higher overall penetration for the annuity market. Additionally, the ability to derive more accurate mortality assumptions could also lead to the development of more tailored products targeted to specific populations, and potentially increase the availability of enhanced annuities that would benefit lower socioeconomic groups having lower life expectancies.

47. Nevertheless, the increased segmentation of annuity pricing raises the question as to what extent society is willing to individualise the pricing of risk and reduce the risk pooling mechanisms on which traditional insurance is based. Increased segmentation has implications for access to the insurance market but also raises the potential that the variables on which segmentation is based could result in discrimination that society could consider unfair.

48. The movement towards the use of machine learning algorithms fed by big data could increase the opaqueness of the logic underlying the pricing of annuities and the variables which are being used. Additionally, the algorithms rely upon the data they are calibrated on, and will therefore enforce any existing bias and discrimination that has been historically prevalent rather than provide an objective and neutral assessment of risk.

49. Drawing the line where the costs of increased segmentation outweigh the benefits will be an extremely difficult task. As such, the answers to the questions of what use of big data is acceptable and how to ensure that the results of algorithms' decisions are fair will require an open and public discussion with all stakeholders.

50. Delegates are invited to provide their views on the following questions:

- What types of variables should be acceptable to use for segmentation? Have you established guidelines or limits in your jurisdiction?
- What type of guidelines would it be useful for the Secretariat to develop?
- How should algorithms used for pricing be audited and supervised?
- Should there be rules in place around how data is used for the distribution of insurance products?
- Would delegates like to pursue additional work on this topic, and if so what direction should this work take?

References

Breading, M. and D. Garth (2014), *Big Data in Insurance - Beyond experimentation to innovation*. [6]

Cable, D. (2013), *The Racial Dot Map*, https://demographics.virginia.edu/DotMap/index.html. [10]

Caliskan, A., J. Bryson and A. Narayanan (2017), *Semantics derived automatically from language corpora contain human-like biases*, pp. 183-186. [12]

Campbell, J. et al. (2014), *Modeling of Policyholder Behaviour for Life Insurance and Annuity Products*, Society of Actuaries. [5]

Edwards, M. (2017), *Big Data in Action: Wearables*. [3]

EIOPA (2017), *Sixth Consumer Trends Report*, Publications Office of the European Union. [14]

FECIF (2017), *FECIF reponse to ESAs Consulation on Big Data*. [16]

IFoA (2017), *IFoA response to European Supervisory Authority*. [7]

Insurance Europe (2017), *Insurance Europe response to ESAs joint discussion paper on big data*. [8]

Joint Committee of the European Supervisory Authorities (2016), *Joint Committee Discussion Paper on the Use of Big Data by Finanical Institutions*. [13]

Kulinskaya, E. and L. Gitsels (2016), *Use of big health and actuarial data for understanding longevity and morbidity risk*. [4]

Mackenback, J. (2016), *Health inequalities in Europe*, Erasmus Univeristy Publishing. [9]

OECD (2016), *Fragmentation of retirement markets due to difference in life expectancy*. [1]

OECD (2016), *Life Annuity Products and Their Guarantees*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264265318-en. [2]

OPSG (2017), *OPSG Opinion on Big Data and Pensions*. [15]

Osoba, O. and W. Welser IV (2017), *An Intelligence in our Image: The risks of bias and errors in Artificial Intelligence*, RAND Corporation. [11]