COVID-19 – an imperative to review your data strategy

A group of Life actuaries, working as part of the [IFoA Covid-19 Action Taskforce] (ICAT) have been considering various ways that the pandemic has affected Life insurers.

In this article, authors Alexia Jami and Roy Perlson look at potential additional data considerations due the impact of COVID-19.*

Part 1 – uncertainty, intuition, heuristics and data

COVID-19 presents life insurance companies with a new situation, full of unknowns. Because the modern world has not seen anything like it and because the patterns of mortality and long-term morbidity effects are still not clear, actuaries cannot fully rely on experience studies.

Not to mention the dramatic effects on the labour market and property markets, driven by work-from-home.

The lack of reliable data means that executives need to make tough decisions in highly uncertain conditions. In such circumstances executives tend to rely on their professional experience, intuition and heuristics (a mental shortcut we use to make decisions faster), all of which developed over many years in the business.

In their Nobel-prize-winning work on Prospect Theory, Israeli scientists Daniel Kahneman and Amos Tversky have shown us how susceptible human judgement is under uncertainty to heuristics and biases due to systematic errors in the functioning of the human cognitive machinery[1].

In 2016 Murli Buluswar, who was at the time the Chief Science Officer of AIG and is now the Head of Analytics, US consumer banks at Citi Bank, said the following in a McKinsey video interview[2]:

"The biggest challenge of making the evolution from a knowing culture to a learning culture—**from a culture that largely depends on heuristics in decision making to a culture that is much more objective and data driven and embraces the power of data and technology**—is really not the cost. Initially, it largely ends up being imagination and inertia".

Unfortunately, even when faced with the high levels of uncertainty created by COVID-19, many in the industry are still relying on heuristics.

All this, while over the last 20 years there has been a convergence of algorithmic advances, data proliferation, and tremendous increases in computing power and storage propelled Machine Learning and AI.

The French mathematician Legendre laid the groundwork for linear regression in 1805, using it in astronomy. Many consider this to be the basis of today's Machine Learning models.

---

[1] https://science.sciencemag.org/content/185/4157/1124

[2] https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-companies-are-using-big-data-and-analytics

In December 2017 Google DeepMind released its creation Alpha Zero, which within 24 hours of self-training, reached a superhuman level of mastery in three different games: Chess, Shogi and Go. Google used TPUs (Tensor processor units) for training and running the models.

In 2018 Geoffrey Hinton, Yoshua Bengio, and Yann LeCun won the Turing Award for their ground breaking work on deep neural networks, computer vision and the use of GPUs (Graphic Processor Units) for training and running machine learning models. The progress they made during the 1990s and early 2000s in the lab translated to the leap in practical applications of AI and ML in the last 5 years.

Whilst an overview of the massive progress made in algorithmic research and computing power is beyond the scope of this article, it is important to be aware of the powerful tools and techniques available today for anyone interested in exploring data.

To fulfil their roles, actuaries need to be heard when the objective interpretation of data contradicts commonly held beliefs, such as business planning assumptions, the expected development of claims, and the assessment of underwriting or reserving risk in a capital model[3].

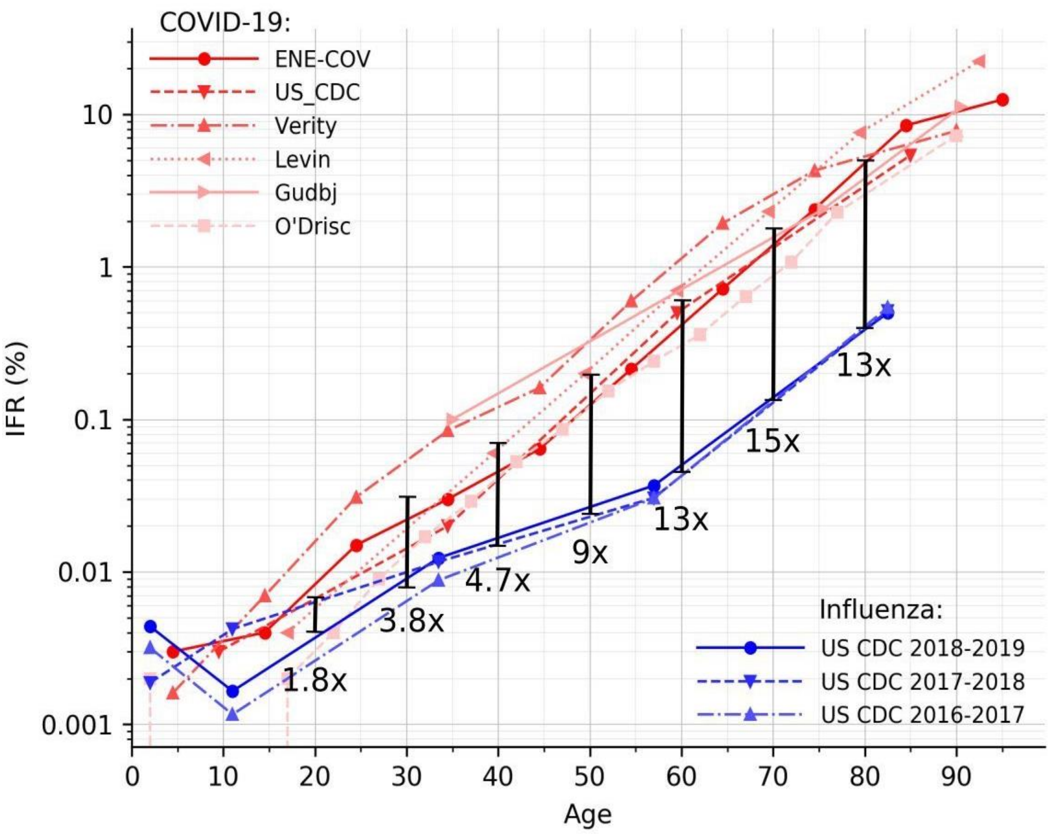
[3] https://www.actuaries.org.uk/learn-develop/attend-event/cigi-webinar-influential-actuary-identitys-effect-data-interpretation

Part 2 – Adopting a scientific approach

Many actuaries are looking at the current mortality data and argue that it is more or less ‘like the flu’ and there is no effect on any assumption. Until we have more data, they are not wrong nor right. Although mortality rates are higher than seasonal flu, the published mortality rates are still well below anything that would require a change to mortality assumptions. Is it ‘like the flu’? Well, we already see indications that COVID causes permanent damage and may stay in your system like Chicken pox, which causes shingles years later. Flu does not do this. So the fact that we don't have the data as of yet does not mean that we can simply assume that it is ‘like the flu’. We have a professional responsibility to use every means to try to start and find a way to understand what is the effect going to be on our book over the short, medium and long term.

Chart 1: Infection Death rate of COVID-19 vs. Influenza

The below chart compares the IFR of COVID-19 to the IFR of seasonal influenza. The chart shows that COVID-19 is significantly more fatal than influenza at all ages above 30 years. The vertical indicators represent the difference in fatality between COVID-19 and influenza at various ages, from 30 to 80 years at 10-year intervals. The top/bottom of the indicators are anchored at the geometric means of the COVID-19/influenza IFR estimates.



[4]

4 https://github.com/mbevand/covid19-age-stratified-ifr?fbclid=IwAR1NlTyYrKz0QWxi6kATxhlDXgYks1bVGiDAIz6nJ0WK8MzwgxgE2kicG_0#comparing-covid-19-to-seasonal-influenza

In addition, there is a bigger-picture element – perception is part of reality and the perception of the effects of COVID-19 are driving unprecedented societal and economic changes. It would be imprudent for anyone whose business is assessing risks to ignore or dismiss the idea that the world is changing in front of our eyes in ways that we cannot even imagine…there will be many consequences, and these consequences will surely find their way to knock on the doors of many life and health insurance companies.

So what would a prudent mind do?

This is not the first time in human history that humanity has been faced with seemingly insurmountable questions. Luckily for all of us, since the 17th century humanity has developed a method of procedure to deal with such problems which consists in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses. No surprise, this is the Oxford English Dictionary's definition for Scientific Method.

Lofty words indeed - put into practical terms it means you must observe the situation, try to come up with relevant questions, try to answer those questions by running experiments, collect data from those experiments and do this continuously and dynamically, taking into account new data, new questions and new experiments.

Bringing this approach to the realm of life insurance, there are so many questions that can be explored in light of COVID-19:

- Mortality rate is a fundamental element for life insurance assumption setting. We need to be able to trust the number which describes the COVID-19 related mortality rate out of total deaths. If it is below a certain point then it can be treated as a non-meaningful event, but if it is larger, we would need to start looking more seriously at reserve calculations, etc.
- But how can we know what the real mortality rate is if we don't know the real number of COVID-19 positive people? Moreover, how can we know the real number of COVID-19 positive people if many people seem to be a-symptomatic?
- How do we better understand the long term effects of COVID-19 on a person's health?
- How do we estimate the secondary effect of COVID-19 on death rates, such as heart attack at home because a person was afraid to visit a hospital?
- How do we measure the tertiary effect of COVID-19 on death rates, such as suicide caused by Mental health issues aggravated by loss of income or loss of social companionship?
- How do we assess the tertiary effect of COVID-19 on death rates due to COVID-19 related behavioural changes such as vast amounts of people spending many months away

from their urban dwelling and temporarily moving to the countryside and taking up outdoor activities - would we see more hiking accidents? More boat accidents? More skin cancer?

- How do we start to understand a tertiary effect of COVID-19 on death rates due to the sudden weight gain experienced by a large part of the population over the last few months? What kind of morbidity and mortality patterns will this generation experience in the mid and long term?

- How do all these questions and many more change by location? Are there measurable differences between countries? Between different areas in the same country? Are the changes manifested in death rates? In the number of new cases per day?

- How do we compare data across different health systems where the definition of a positive test and COVID-19 impact on mortality vary in definitions? While some systems count only the mortality impact due to COVID-19 directly others also count other mortality reasons where COVID-19 was positive as a COVID-19 impact?

- How do we explore the complex relationship between quality of medical care in a certain area and chances of dying from COVID-19?

Part 3 - Avoid analysis-paralysis. Start with exploring available data and relate it to your book of business

To show you a practical example of how to start to explore the effects of COVID-19 on your book of business, we asked ourselves where is the obvious place to start and the answer we came up with is to dive into the government COVID-19 death statistics, use this as a starting point and see where the path leads us, while keeping the effort short and focused. This is an agile approach to data exploration – doing sprints of data exploration, which take no longer than a few days of core analysis, then reviewing the results, assessing the meaning, and then doing another sprint. This method is adopted from software development agile methodology. To give you a sense of the required input – the team spent about 50 man-hours in producing the below. Basically, less than a week of focused work by a couple of experienced professionals – from ideation about how to approach the question, to final findings. We believe this is an effort any team can commit to as a first step.

Step 1 – We accessed the publicly available UK data set of March to July 2020 number of deaths from COVID-19 and deaths from all causes, found at this link:

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingcovid19bylocalareaanddeprivation[5]

We found that Table 5 included the March to July 2020 deaths by MSOA code.

MSOAs (Middle Layer Super Output Areas) are part of the Office of National Statistics for England and Wales set of granular geographical areas. MSOAs are areas in which the population range is 5,000-15,000. Based on the 2011 Census data there are 7201 MSOAs in England and Wales.

Chart 2: definition of MSOA

Population and household minimum and maximum threshold for SOAs in England and Wales				
Geography	Minimum Population	Maximum Population	Minimum number of households	Maximum number of households
LSOA	1,000	3,000	400	1,200
MSOA	5,000	15,000	2,000	6,000

[5] Source: Office for National Statistics licensed under the Open Government Licence.

To give the readers a feeling for the data we attach a snapshot of Table 5.

Chart 3: 'COVID_19_5 month - March to July' / 'ALL_CAUSES_5 month - March to July'

			COVID-19		Non COVID-19		All causes
MSOA code	**House of Commons Library MSOA Names**		**5 month - March to July**		**5 month - March to July**		**5 month - March to July**
E02000001	City of London		4		18		22
E02000002	Marks Gate		8		29		37
E02000003	Chadwell Heath East		10		28		38
E02000004	Eastbrookend		14		23		37
E02000005	Becontree Heath		7		13		20
E02000007	Central Park & Frizlands Lane		14		27		41
E02000008	Becontree East		11		29		40
E02000009	Becontree West		7		20		27
E02000010	Dagenham North		7		20		27
E02000011	Dagenham Eastbrook		12		34		46
E02000012	Longbridge & Barking Park		8		30		38
E02000013	Becontree South		4		17		21
E02000014	Dagenham Central		6		26		32
E02000015	Old Dagenham Park & Village		4		18		22
E02000016	Barking Central		1		10		11
E02000017	Barking East		6		12		18
E02000018	Mayesbrook Park & Rippleside		13		17		30
E02000019	Goresbrook & Scrattons Farm		8		20		28
E02000020	Creekmouth & Barking Riverside		6		22		28
E02000021	Rylands Estate & Dagenham Dock		6		37		43
E02000022	Gascoigne Estate & Roding Riverside		4		9		13
E02000023	Thames View		4		7		11
E02000024	High Barnet & Hadley		13		50		63
E02000025	Hadley Wood		16		20		36
E02000026	Oak Hill		7		17		24
E02000027	Ducks Island & Underhill		14		30		44
E02000028	New Barnet West		9		24		33
E02000029	New Barnet Town & East Barnet		9		23		32
E02000030	Totteridge & Barnet Gate		10		32		42
E02000031	Oakleigh Park		16		29		45
E02000032	Osidge		6		17		23
E02000033	Brunswick Park Road		11		27		38
E02000034	Whetstone East		10		27		37

Step 2 – we joined the data set with several other publicly available data[6] sets from the Office of National Statistics and ran Pearson's correlations.

Pearson's Correlation Coefficient *is a common statistical measurement to assess an interaction between parameters in the data. Pearson's coefficient is the linear* ***correlation coefficient*** *that returns a value of between -1 and +1. Where -1 means there is a strong negative* ***correlation*** *and +1 means that there is a strong positive* ***correlation****. A 0 means that there is no* ***correlation*** *(this is also called zero* ***correlation****).*

We did this for 5 different data sets and looked for interesting patterns.

Demographic related data:

- KS201EW - Ethnic group
- KS202EW - National identity

Financial status related data:

- KS403EW - Rooms, bedrooms and central heating

Employment related data:

- KS605UK - Industry

Health related data:

- KS301EW - Health and provision of unpaid care

Step 3 – based on a review of the results of step 2 we chose to focus on employment data, KS605UK - Industry, provided by NomisWeb[7].

This table provides information that classifies usual residents aged 16 to 74 in employment the week before the census by the industry in which they work, across the United Kingdom as at census day, 27 March 2011.

This data is available by MSOA (the geographical unit used in the COVID-19 death table).

Step 4 – By digging deeper into the industry employment by MSOA and deaths by MSOA we found a few interesting findings:

[6] Nomis official Labour market statistics (https://www.nomisweb.co.uk/census/2011/key_statistics_uk)
[7] https://www.nomisweb.co.uk/census/2011/ks605uk

Chart 4: Pearson's Correlation % of death by COVID-19 out of total death from March to July 2020 by Industry (% in a specific industry out of working population) by MSOA

Industry	Correlation
N Administrative and support service activities : 0.27	0.2695425
J Information and communication : 0.22	0.221070036
K Financial and insurance activities : 0.2	0.199811565
M Professional, scientific and technical activities : 0.16	0.159604531
L Real estate activities : 0.16	0.159019531
H Transport and storage : 0.15	0.150836644
R, S, T, U Other : 0.09	0.090465421
I Accommodation and food service activities : 0.02	0.024784711
P Education : -0.02	-0.018708893
G Wholesale and retail trade, repair of motor vehicles and motor cycles : -0.07	-0.071844551
D Electricity, gas, steam and air conditioning supply : -0.08	-0.075985438
Q Human health and social work activities : -0.09	-0.086251325
B Mining and quarrying : -0.12	-0.117322677
O Public administration and defence, compulsory social security : -0.12	-0.11882642
F Construction : -0.15	-0.148324561
E Water supply; sewerage, waste management and remediation activities : -0.16	-0.163739417
C Manufacturing : -0.2	-0.195938228
A Agriculture, forestry and fishing : -0.27	-0.266754487

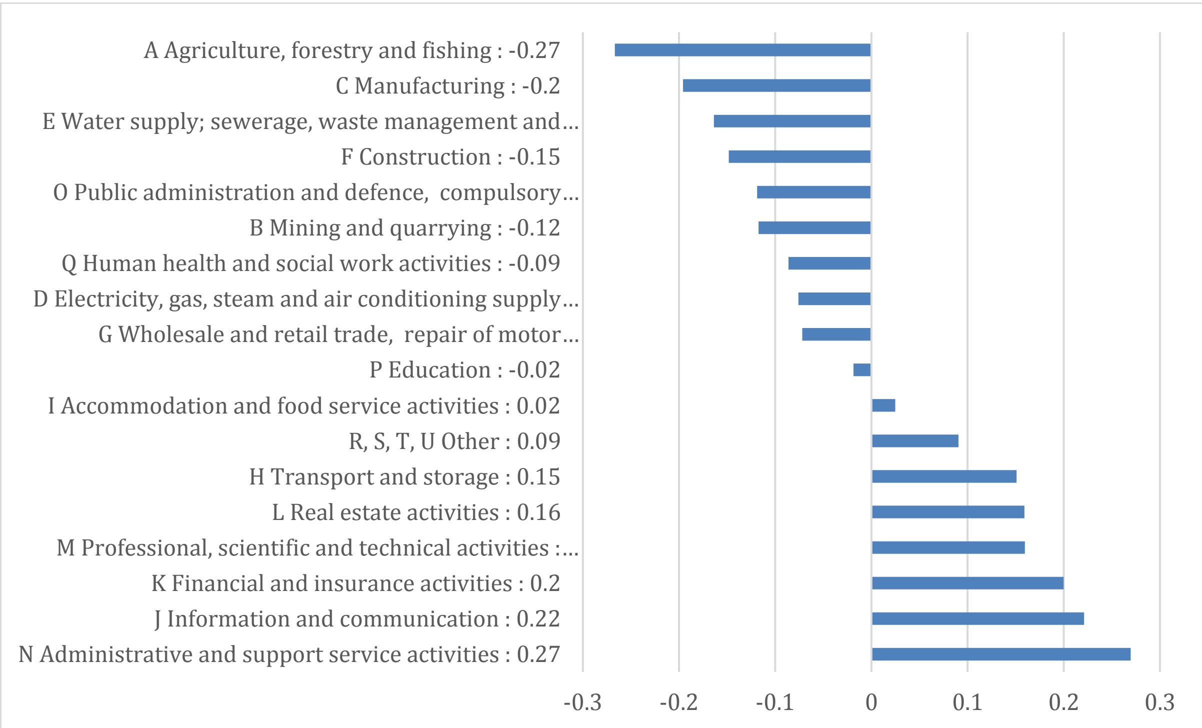


Chart 4 shows Finding #1 – office-based occupations such as administration, information, communication, financial and insurance services etc have a higher rate of COVID-19 related death than Outdoors based occupations such as agriculture, forestry, fishing.

Step 5 - For each industry in the above chart we plotted the % of COVID-19 deaths. Each dot is a MSOA. We found 2 interesting plots:

Chart 5: Negative Correlation between % of COVID-19 death out of total deaths in the MSOA and % of the population in the MSOA employed in the industry category of Agriculture, forestry and fishing

Chart 5 shows Finding #2 - If Agriculture, forestry and fishing is bigger than 5% in the MSOA then percentage of deaths from COVID-19 is lower than 40% of total deaths in the MSOA

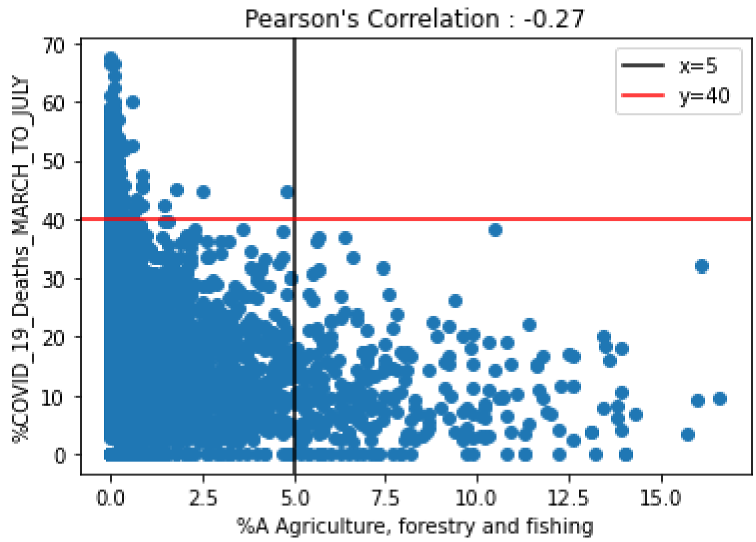
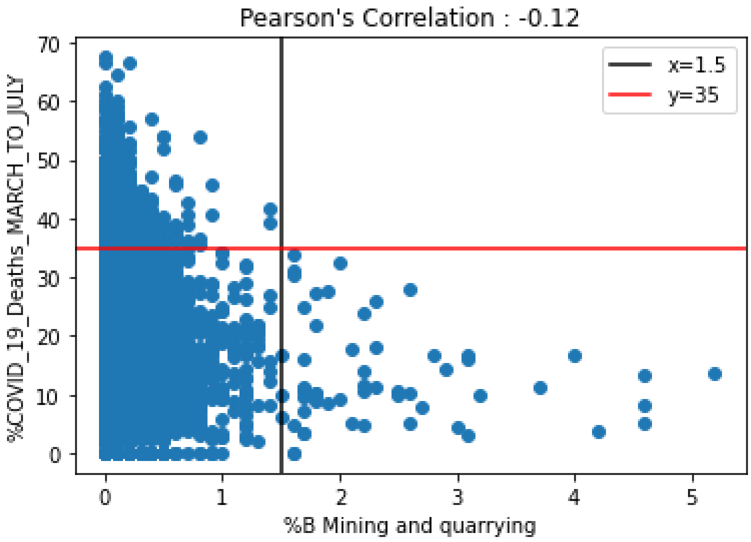


Chart 6: Negative Correlation between % of COVID-19 death out of total deaths in the MSOA and % of the population in the MSOA employed in the industry category of mining and quarrying

Chart 6 shows Finding #3 - If Mining and quarrying is bigger than 1.5% in the MSOA then percentage of deaths from COVID-19 is smaller than 35% of total deaths in the MSOA



This insight should not be a surprise as we can assume that people living in MSOAs with high focus on agriculture and outdoor occupations would have higher likelihood to spend their day in an open air environment therefore less spreading the virus. On the contrary, MSOAs with high concentration of office work occupation would probably be a more city-center type demographic, possibly higher number of day to day interactions and higher likelihood to be exposed to the virus.

Pay attention that the ability to understand the demographic, occupation and financial ability of postcodes and areas within the country was done based solely on **open and free data sets.** Therefore, leveraging such technology and know-how gives insurance companies the power to improve their understanding of the book of business to a much greater extent than traditional actuarial modelling.

Think about layering this kind of analysis on top of YOUR book of business. You would normally have the insured address as well as occupation, so joining the data should be straight forward.

This analysis is only one-dimensional. Further analysis could be made with COVID-19 data for gender, age, weight, etc., to come up with accurate risk levels by insured and even to predict behaviour and model scenarios with data science tools and cutting edge algorithms. COVID-19 taught us that a change can happen rapidly and the ability to analyse at scale in real time becomes a necessity. Continuously feeding new data into your models, and continuously coming up with new questions and running data experiments to try to find answers to these questions could be a major differentiator to companies and policyholders as this situation unfolds.

Ask yourself - is your book focused around a certain geographic area? Or a certain occupation?

Would the above analysis shed any light on your assumptions for 2021 onwards?

You might find out that your book is too small to show meaningful patterns in the data. For example, if your book (or the portion of the book relating to a certain product/waiver which you want to analyse) is smaller than 50,000 policy holders, you might want to explore the use of synthetic data – basically computer-generated avatars which, if properly structured, can be used to mimic populations and fed into your analytical models.

Part 4 – Key Takeaways

- COVID-19 uncertainty
- No experience data
- Heuristics and intuition not good enough
- Scientific approach is required
- Ask the right questions and explore available data sources
- Advanced tools and techniques are available
- Suggested starting point – look at the potential effects of COVID-19 on your book of business through two lenses: (1) granular geographic area; (2) occupation
- If you don't have enough data in your book you can use synthetic data (avatars)
- The time to act is now as the COVID-19 effects are complex, omni-present and have both short and long term implications – the life insurance industry has to better measure and understand the effects.

Authors:

Alexia Jami
Alexia is a senior data scientist at Atidot. Alexia honed her life insurance data science skills working 7 years at Siaci Saint Honore. She holds two Masters degrees from Universite Paris Dauphine, one in insurance and risk management and one in Mathematical modelling and decisions

Roy Perlson
Roy is the General Manager, EMEA of Atidot. Before joining Atidot he was the head of McKinsey London technology ventures unit. He holds an MBA from INSEAD and a law degree from Tel Aviv University.

Atidot is a software company bringing together actuarial science, data science and computer engineering to provide value to life insurance companies.

The ICAT Life 3 Workstream is researching the impact on the year-end 2020 valuation assumptions-setting process due to the COVID-19 pandemic.

This article is one in a series published by this ICAT Workstream – *Assumption setting in the current uncertainty*. Members are Anjali Mittal, Burcin Arkut, Isha Aggarwal, Isha Kulkarni, Jagrit Bagga, James Gillespie, Justine Morrissey, Mcebisi Dhlamini, Nainjeet Juneja, Natalia Mirin (Workstream Lead), Roy Perlson, Sanjoli Choudhary, Thomas Treacy.