Institute
and Faculty
of Actuaries

IFoA Member Survey on the Data Science Universe

Results and recommendations

By the Modelling, Analytics and Insights from Data (MAID) Working Party – Research Workstream

06 February 2017

IFoA Member Survey on the Data Science Universe

Authors

Members of the MAID Research workstream:

- Alicja Nocon (chair)
- Sarah Mathieson (IFoA)
- Jonathan Allenby
- Michael Crawford
- Jonathan Day
- Stuart Jarvis
- Jeanette Johnson
- Paul King
- Barry Murphy
- Richard Russell
- Simon Tam
- Ji Yao

Abstract

The Modelling, Analytics and Insights from Data (MAID) Working Party was established to identify opportunities and risks presented by the data science universe to the actuarial profession. As part of this, a survey was conducted with IFoA members to gauge current awareness of data science techniques and current practice. The findings of the research suggest there is an opportunity to increase awareness amongst actuaries, provide support to build skills and expertise, collaborate with other professionals and take a lead on the ethics of data science, as a public interest body.

Keywords

big data, data science, machine learning, predictive analytics

Correspondence details

c/o Research and Knowledge, Institute and Faculty of Actuaries, Level 2 Exchange Crescent, 7 Conference Square, Edinburgh, EH3 8RA, United Kingdom E-mail: research@actuaries.org.uk

1. Introduction

As a cross-practice working party, the IFoA's Modelling, Analytics and Insights from Data (MAID) Working Party, was established to identify the opportunities and risks presented by the data science universe for the actuarial profession. This includes, but is not limited to, big data, predictive analytics, machine learning and an ever-wider range of mathematical techniques based on operational research, probability and statistics.

The MAID Working Party has four workstreams:

- Research
- New approaches to existing actuarial work
- New areas and opportunities for actuarial work
- Implications for professional affairs

The aim of the Research workstream is to develop an understanding of thought leadership opportunities relating to the field of data science, both in the UK and internationally, that would be of interest and importance to the work of actuaries.

This report presents the results and findings from a member survey carried out by the Research workstream to build up a view of current awareness and practice of data science techniques within the IFoA membership. The objective of conducting the survey was to help inform any additional research that the IFoA may wish to undertake in this area, along with informing the work of the MAID Working Party as a whole.

2. Methodology

The research was carried out through an online survey using a combination of qualitative and quantitative techniques. The member survey was open for four weeks (17 October 2016 to 14 November 2016) and publicised through the IFoA website, practice area newsletters, social media and the IFoA's GIRO 2016 Conference in Dublin.

In total, 434 responses were received, with around 80% of participants completing the substantive questions.

Appendix A summarises the profile of the respondents. A limitation is potential bias that may have arisen from:

- Respondents self-selecting to complete the survey – it was likely to attract IFoA members already engaged or interested in the topic
- Where the survey was promoted – for example, publicity at GIRO may have introduced bias towards:
    - actuaries who have been longer qualified (Figure 1);
    - general insurance practitioners (Figure 2);
    - members working for an insurer or reinsurer (Figure 3);
    - UK or Republic of Ireland based members (Figure 4).

This limitation was considered in the analysis of the survey responses.

3. Overview of key findings

3.1. Awareness levels

Respondents were asked to self-assess their current level of awareness of the data science universe through a set of statements. Details are outlined in appendix B, with high level responses summarised in Figure 5.

- 83% of respondents have at least read about data science and are aware of what it means.
- 43% of respondents have used data science at work but only 24% have used data science techniques to produce or support actuarial findings.
- Only 9% of respondents use data science techniques regularly.
- Data science techniques appear to be most frequently used in the finance and investment and general insurance practice areas (Figure 6): 23% of respondents working in finance and investment and 12% of respondents working in general insurance regularly use data science techniques.
- In contrast, data science techniques appear to be less commonly used in the pension and life insurance practice areas. 25% of respondents working in pensions and 16% of respondents working in life insurance did not know what data science meant.
- There are also some slight geographical variations (Figure 7), where it would appear that less UK and Republic of Ireland respondents have used data science techniques when compared against other regions. However, UK and Indian respondents appear to have more experience using data science techniques to specifically support actuarial findings.

3.2. Defining the data science universe

A number of features can be associated with some of the data science universe concepts and techniques. Respondents were ask to share what they associate with “big data”, “predictive analytics” and “machine learning”, using a number of suggested terms but were also invited to include their own. The responses are summarised in appendix C.

- The term ‘big data’ is strongly associated with “extremely large datasets” and less with “incomplete data” (Figure 8). In particular, respondents who stated that they regularly use data science techniques strongly associated big data with “extremely large datasets”. On the whole, the responses were broadly similar across all groups of respondents, with the exception of those with lower awareness levels. This might suggest there is more work to be done on raising awareness of what big data actually means with that group.
- The term ‘predictive analytics’ is strongly associated with “making predictions” and less with “regression and GLMs” (Figure 9). Respondents who use data science techniques regularly showed broadly the same pattern of results as all respondents. However, respondents with low awareness levels tended to rank “regression and GLMs” higher. This tied up with a couple of comments that GLMs may be less relevant in the near future. Like big data, there is an opportunity to raise awareness with that group as well as continuing to push the boundaries through research..
- The term ‘machine learning’ is strongly associated with “learning from experience” and less with “complex algorithms” (Figure 10). Respondents who use data techniques tended to score the features differently to the group as a whole. There is an opportunity for the wider membership to learn more from that group, perhaps through various awareness-raising activities.

3.3. Learning about data science

There is a significant volume of learning material already on the big data universe. Respondents were asked to share what they are currently using as their main learning sources. Their responses are summarised in appendix D, with the following high level observations:

- The most popular approaches were online reference materials (18% of respondents), learning from non-actuaries (16% of respondents) and attending conferences (15% of respondents).
- The least popular approaches were classroom learning and participating in data science competitions, with only 4% of respondents selecting either of these options.
- The survey data suggests that students are more likely to use online reference material compared to more experienced respondents who appear to favour conferences.

3.4. Current practice

The final area that the survey explored was current practice within the IFoA membership and any actions the IFoA could be undertaking. The detailed findings are included under appendix E.

Respondents were asked how their current role compares to a given definition of a "data scientist" and a "data analyst.

- 41% view their current role as having more than 50% in common with that of a 'data analyst'
- 30% view their current role as having more than 50% in common with that of a 'data scientist'
- 33% view their current role as having more than 50% in common with that of a 'user of data science outputs'
- Those who work in general insurance appeared to identify more with the above three roles compared to other practice areas
- Those who work in consulting or business advisory appeared to identify less with the above three roles compared to those who work directly for an insurance or reinsurance company

In terms of data science techniques currently in use:

- 65% said that they use regression techniques and data visualisation techniques occasionally or regularly
- 50% said that they use generalised linear models (GLMs) occasionally or regularly
- 37% said that they use Bayesian techniques occasionally or regularly
- 20% said that they use decision trees, random forests, gradient boosting, clustering and principal components occasionally or regularly
- Fewer than 10% use neural networks, social network analysis and support vector machines occasionally or regularly

Lastly, respondents were given an opportunity to share, through an open question, comments concerning actions that the IFoA should seek to undertake. The comments were rich in content with a number of common themes, which are summarised, along with some selected quotes, in section (v) of appendix E. Education and training (pre- and post-qualification) were dominant themes, suggesting that members are open to embracing the data science universe as part of their natural career progression. Ethical and legal considerations of the data science universe also came through in the comments, which ties back the IFoA's role as a professional body and regulator. Finally, a more strategic point, raised by some respondents, was the risk that data scientists may replace the roles of actuaries in some areas.

4. Summary and conclusions

The survey results suggest that there is at least some awareness of the data science universe amongst the majority of respondents, with variations across practice areas and possibly geography. There is an opportunity for cross-practice learning and sharing of knowledge and expertise within the profession itself. At the same time, many members indicated that they learn, or would like to learn, from non-actuaries and academics, as well as collaborate with other professional bodies.

Learning was a consistent theme running through the open comment feedback. Career development was important for both actuaries of the future (students still taking the actuarial exams) and current actuaries (Fellows interested in learning more about data science in the form of CPD). Some respondents expressed a desire to move into roles in wider fields and involving data science. The IFoA's future education curriculum (to commence in 2019) will see the addition of substantive content on data analytics (IFoA, 2016a). The curriculum includes the use of R, GLMs, regression techniques and data visualisation (IFoA, 2016b), all which have been deemed important in this survey feedback.

While the new curriculum will improve the data science knowledge gap, there remains a need to improve the education of existing members, particularly those who qualified some time ago and are seeking CPD in the field of data science. The survey responses suggested that online resources (courses and reference material), along with more traditional conferences and events are the preferred learning method for that group.

In terms of data science itself, there was a general consensus amongst respondents on what they viewed as the important features of big data, predictive analytics and machine learning. This may help actuaries form a common language when it comes to building up an understanding of data science techniques as well as guiding the focus for any learning activity. As knowledge builds in this discipline, gaps will also start to be identified and potential opportunities for actuaries and others to advance the subject matter through research.

Although actuaries do not necessarily market themselves as "data scientists", a large degree of commonality was identified between respondents' current roles and the definition of "data scientist" and "data analyst". This could indicate a need to better promote the actuarial skillset using different language.

Finally, some of the qualitative feedback from respondents touched on the ethical and legal aspects of data science. While there may not be an appetite for specific actuarial standards in this area, the risks associated with working in the data science universe also need to be understood. Underpinned by a rigorous set of professional standards, this could be a unique-selling point for actuaries and the IFoA, as a professional body and regulator. It is an opportunity, as well as a risk.

5. Recommendations

Based on the evidence provided in the survey, the following recommendations are proposed for consideration:

Recommendation 1: Promote the availability of data science methods and techniques across traditional actuarial sectors, as well as the opportunities that they may present for actuaries to move into wider fields.

- Raise awareness of how different techniques and tools (such as R) are already used across the practice areas. For example, through published case studies on existing applications of data science techniques in general insurance and health and care.
- Utilise the IFoA's events and conferences to promote existing data science activities. For example, a dedicated plenary on data science at the residential conferences. To encourage cross-practice knowledge exchange, some of the cross-practice events could be targeted (such as the Momentum Conference or events organised by the Regional Societies).
- Share information through IFoA communication channels (eNewsletters, website, *The Actuary*) on data science applications within an actuarial context or where actuaries are applying similar techniques elsewhere. For example, some of the respondents to the survey may be interested in authoring articles.
- Develop an "introduction to data science" for the IFoA website, perhaps including a glossary of terms.
- Publicise the findings of this survey to members.

Recommendation 2: Ensure that pre- and post-qualification learning equips future and current actuaries with the tools and techniques required in the data science universe.

- Raise awareness of the findings of this survey with the IFoA's Education Board and Practice Boards' CPD Committees.
- Ensure that industry developments are reflected in the education syllabus and CPD opportunities (such as practical industry-based workshops or online courses).
- Encourage promotion of Curriculum 2019 its data science content.

Recommendation 3: Support the advancement of the subject matter through research, encourage its adoption by practitioners and demonstrate thought leadership outside the actuarial profession.

- Encourage the IFoA's Practice Boards' Research Committees and the Research and Thought Leadership Committee (RTLC) to consider the findings of this survey and identify any further research opportunities through working parties or commissioned projects.
- Provide a platform for key thinkers and actuarial practitioners exchange ideas and knowledge to help identify gaps in research and encourage "blue sky thinking".

Recommendation 4: Collaborate with other professionals and disciplines to share knowledge and skills and to advance techniques and methods.

- Collaboration could include joint research projects, events, publications and other activities.
- Identify the best way to work with other organisations that are already developing their own data science agenda, such as the Casualty Actuarial Society (CAS) and the Royal Statistical Society (RSS).

Recommendation 5: As part of the IFoA's public interest role, ensure that the regulatory and ethical implications arising from the big data universe and application of data science techniques are understood and relevant action taken.

- Highlight the findings of this survey to the IFoA's Regulation Board for further internal discussion and a potential topic to discuss with the Financial Reporting Council (FRC). There may be a need to develop a *Risk Alert* on the limitations of using data science methods

Acknowledgements

The authors of this report would like to thank the IFoA members who responded to the survey for their participation and making this publication possible.

The authors also thank the MAID Steering Committee for their feedback during the development of this survey:

- Michael Tripp (chair)
- Ben Canagaretna
- Alberto Chierici
- Tony Fung
- Steven Graham (IFoA)
- Alexander Hanks
- Jamie Marshall
- Alex Panlilio
- Matthew Wilson

References

IFoA. (2016a). *Curriculum 2019.* [online] Institute and Faculty of Actuaries. Available at: https://www.actuaries.org.uk/node/180291 [Accessed 3 January 2017]

IFoA. (2016b). *Syllabus 2019.* [online] Institute and Faculty of Actuaries. Available at: https://www.actuaries.org.uk/documents/syllabus-2019 [Accessed 3 January 2017]

APPENDICES

Detailed results from the member survey

APPENDIX A: About the survey respondents

(i) Career stage

Qualified actuaries with more than 10 years' experience were the most common respondents (148 respondents, 43%). Students made up 36% of respondents (125 respondents), with actuaries with less than 10 years' experience making up 21% (73 respondents) of responses.

Compared to the IFoA profile as a whole, the respondents were slightly more skewed towards more experienced actuaries with students under-represented (Figure 1).

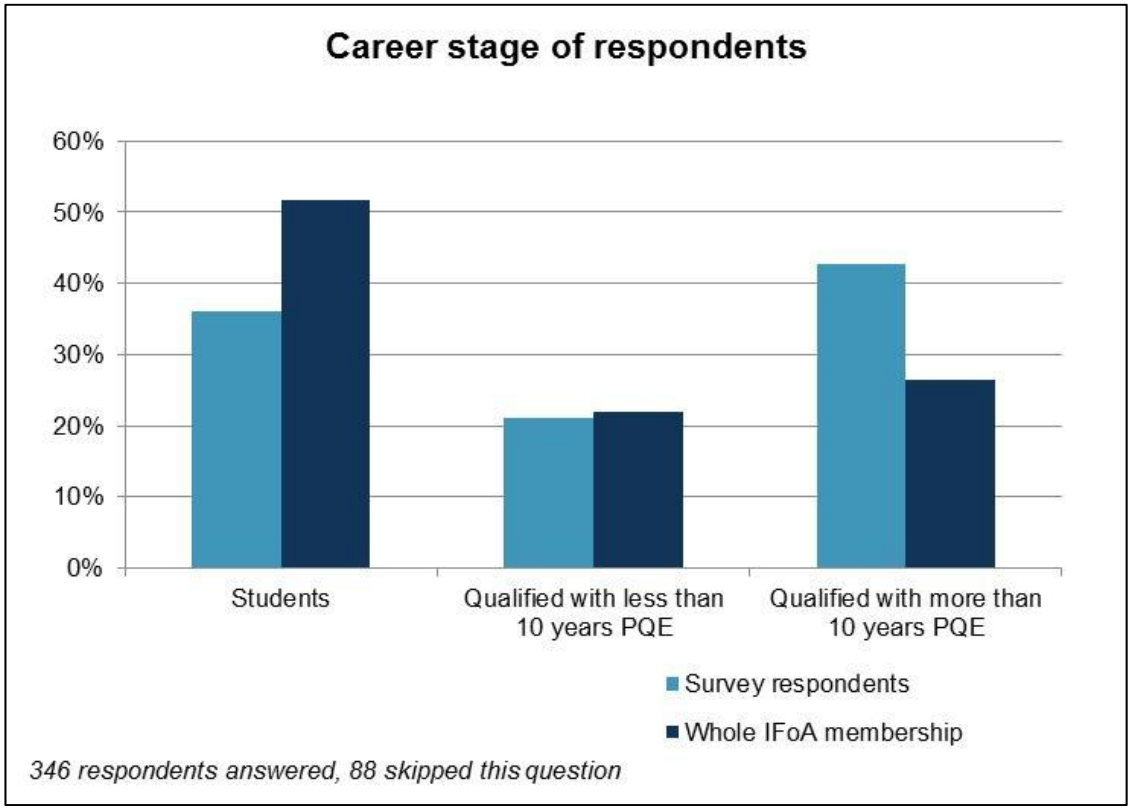

Figure 1

(ii) Practice areas

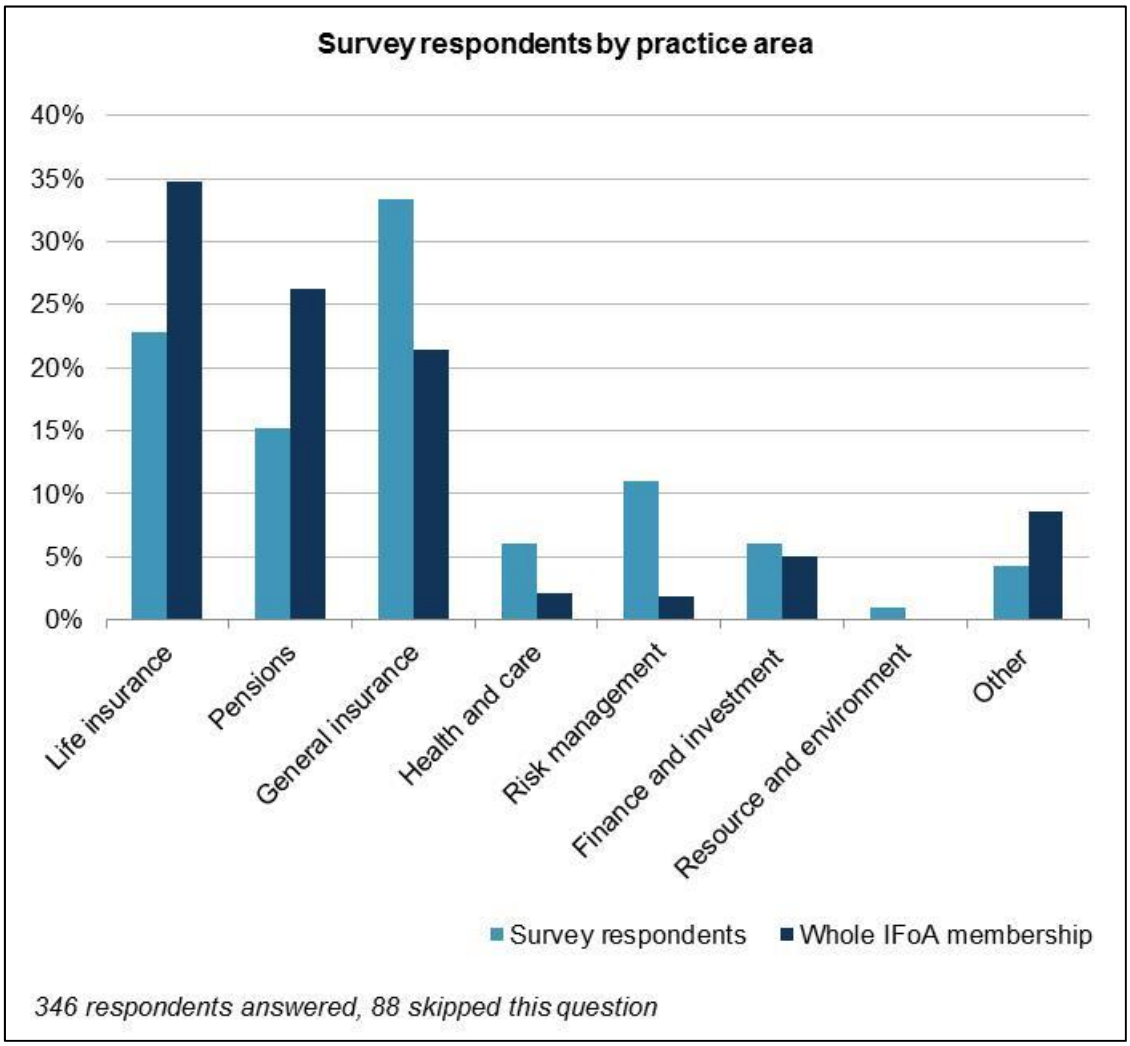

Figure 2

Retired and unspecified practice area categories were excluded from the IFoA member profile for comparison purposes. It should be noted that nearly 20% of IFoA members do not specify a practice area in the IFoA database and it is therefore assumed that they follow a similar distribution to the members who have specified a practice area.

Respondents were allowed to choose more than one practice area and no allowance has been made for this in the analysis below (the total number of respondents by practice area will be larger than the total number of respondents by other field e.g. experience).

There was a relatively high proportion of general insurance respondents (33%, 169 respondents) compared to the IFoA member profile (21%). This may be a reflection of the survey's promotion at the 2016 GIRO conference.

There was also a relatively high proportion of respondents from Health and care (6%, 31 respondents), Risk management (11%, 56 respondents) and Finance and investment (6%, 31 respondents) practice areas compared to the IFoA member profile (which is 2%, 2% and 5% respectively).

Respondents who chose 'Other' practice area and specified their own category provided the following categories:

- HR
- Workforce analytics/intelligence
- Executive compensations
- Software industry
- Software engineering/development
- Cloud computing
- IT
- Technology
- Official statistics
- Paediatric health research
- Data science and analytics
- Public sector
- Non-actuarial/not a traditional actuarial role

(iii) Type of employer

54% of the 346 respondents (185) stated that they work for an insurance or re-insurance company. A further 30% work (102) in a consulting or advisory firm, 7% work in the public sector (public service or regulator). The remaining 10% are spread across a range of investment, technology and other firms.

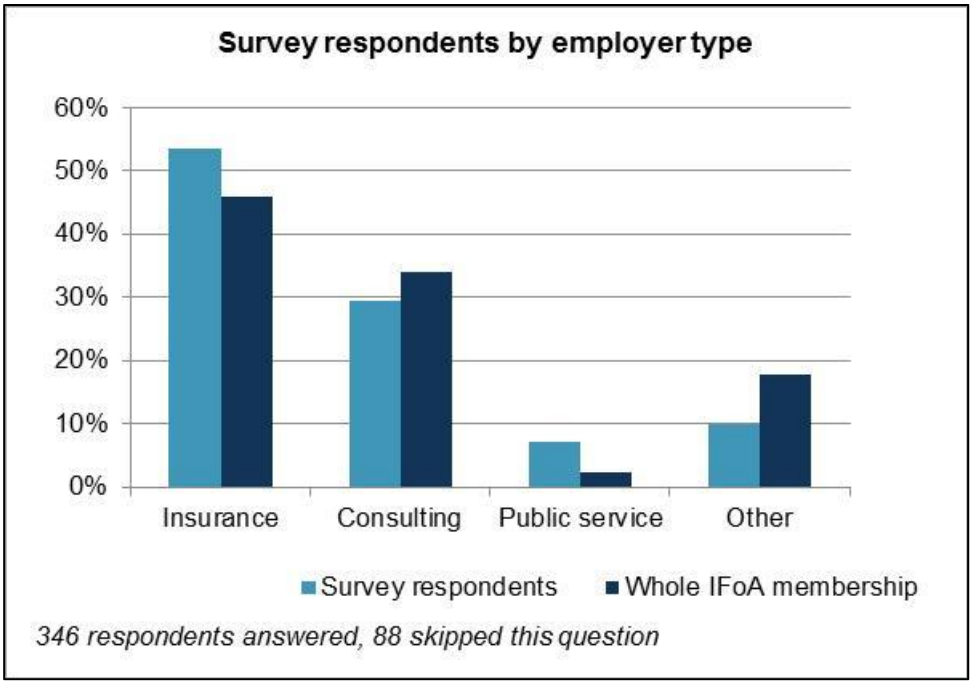

Figure 3

(iv) Geographical location

Thirty four unique countries from across the world are represented in the survey data. As shown in Figure 4, the most common location for respondents was Europe (79% of total respondents, a total of 261 respondents). The top four locations for respondents were the UK (61%, 217 respondents), India (9%, 31 respondents), Ireland (8%, 28 respondents) and South Africa (2%, 6 respondents).

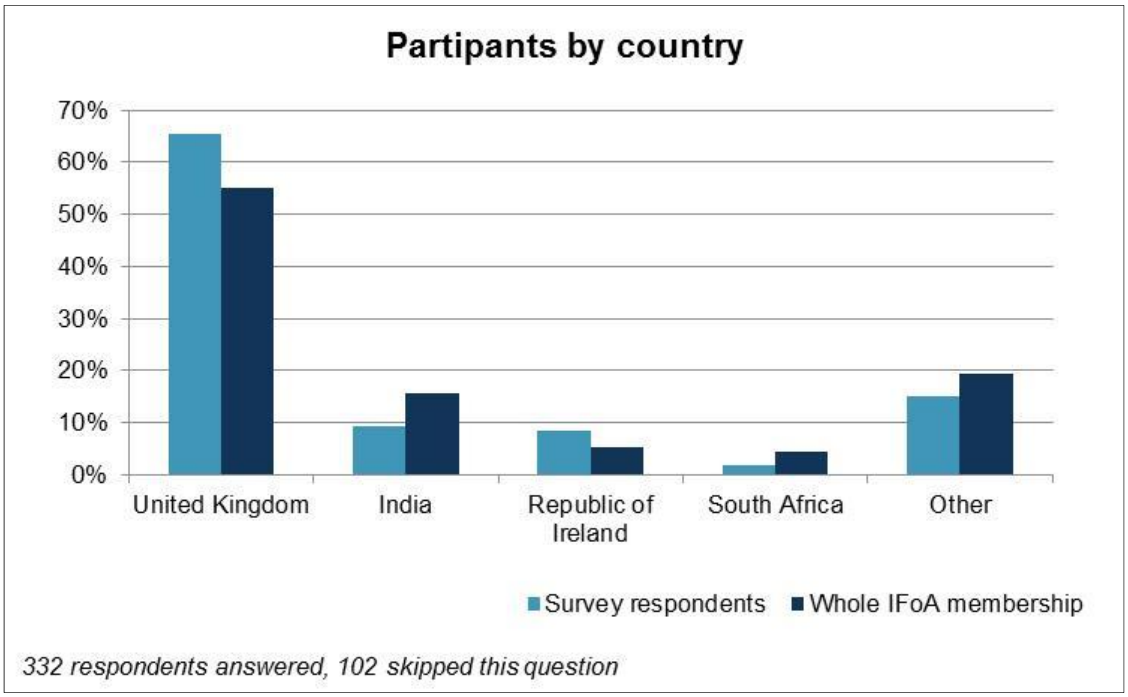

Figure 4

Note: The "Other" category in the above graph contains countries with a total of five respondents or less.

(v) Job titles of respondents

Of the 434 responses, 345 provided a job title (80%) with 113 job titles containing the word “Actuary” and a further 87 the word “Actuarial”. The ten most common job titles given were:

1. “Actuary” (30)
2. “Actuarial Analyst” (24)
3. “Pricing Actuary” (13)
4. “Chief Actuary” (12)
5. “Consultant” (9)
6. “Senior Actuarial Analyst” (9)
7. “Senior Consultant” (7)
8. “Consulting Actuary” (7)
9. “Actuarial Consultant” (7)
10. “Manager” (6)

APPENDIX B: Awareness levels

(i) Familiarity with the data science universe

Respondents were asked to choose one of the following five statements that describe their familiarity with data science universe:

1. I have heard of “data science” as a buzz word but I don’t know what it means.
2. I have read about data science and I am aware of what it means but I have not used any of its key techniques or technologies.
3. I have tried out some basic data science techniques on a data set at work but have not used them to produce or support any actuarial findings.
4. I have used data science techniques on a real data set for work in order to produce/support actuarial findings but do not use them regularly.
5. I regularly use data science techniques in the production of my actuarial findings.

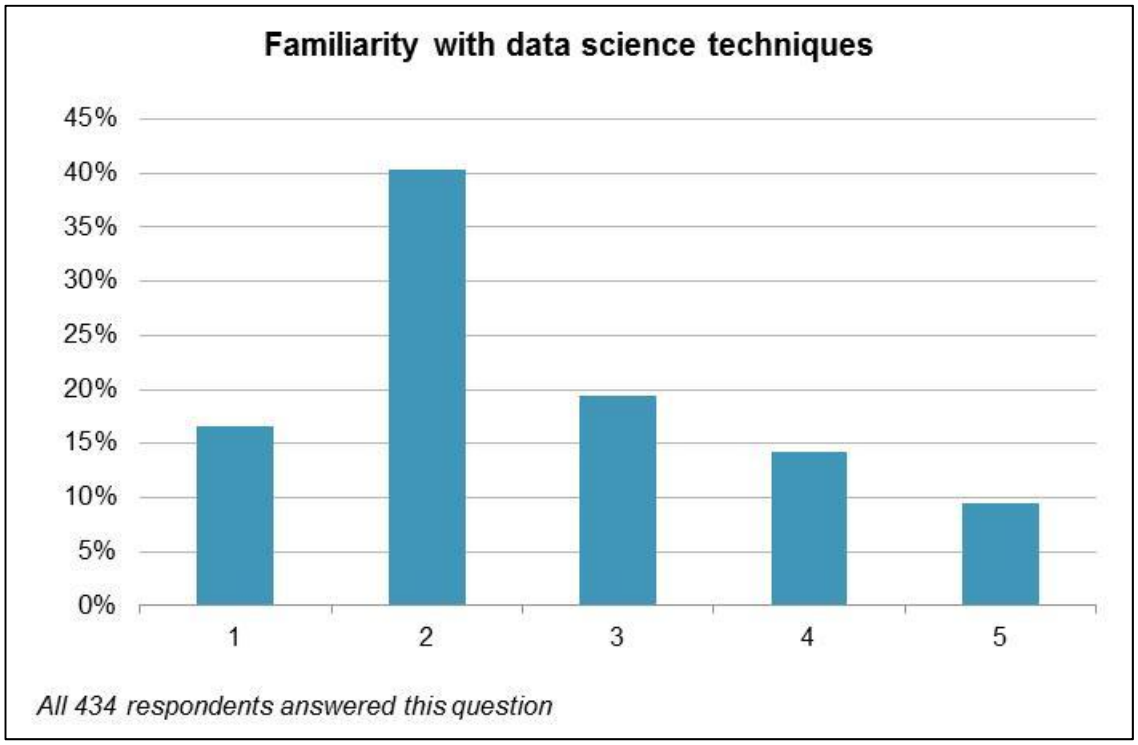

Figure 5

362 (83%) of respondents show a good level of awareness of data science techniques (option 2 to 5). However, 175 (40%) of respondents are aware of what data science means but have not used any of its key techniques or technologies (option 2).

187 (43%) of respondents have used data science techniques at work (option 3 to 5) but only 103 (24%) used them to produce or support actuarial findings (option 4 to 5).

(ii) Awareness levels by practice area

41 (9%) of respondents use data science techniques in the production of actuarial findings on a regular basis (option 5), out of which 21 (40%) are from general insurance, 10 (19%) are from life insurance, 7 (14%) are from Finance and investment, and the remaining 27% are broadly evenly spread between Pensions (4%), Health and care (3%), Risk management (3%) and 'Other' (4%) practice areas. We note that the total number of practice areas is 52 which is larger than the number of respondents (41) and is due to some respondents selecting multiple practice areas.

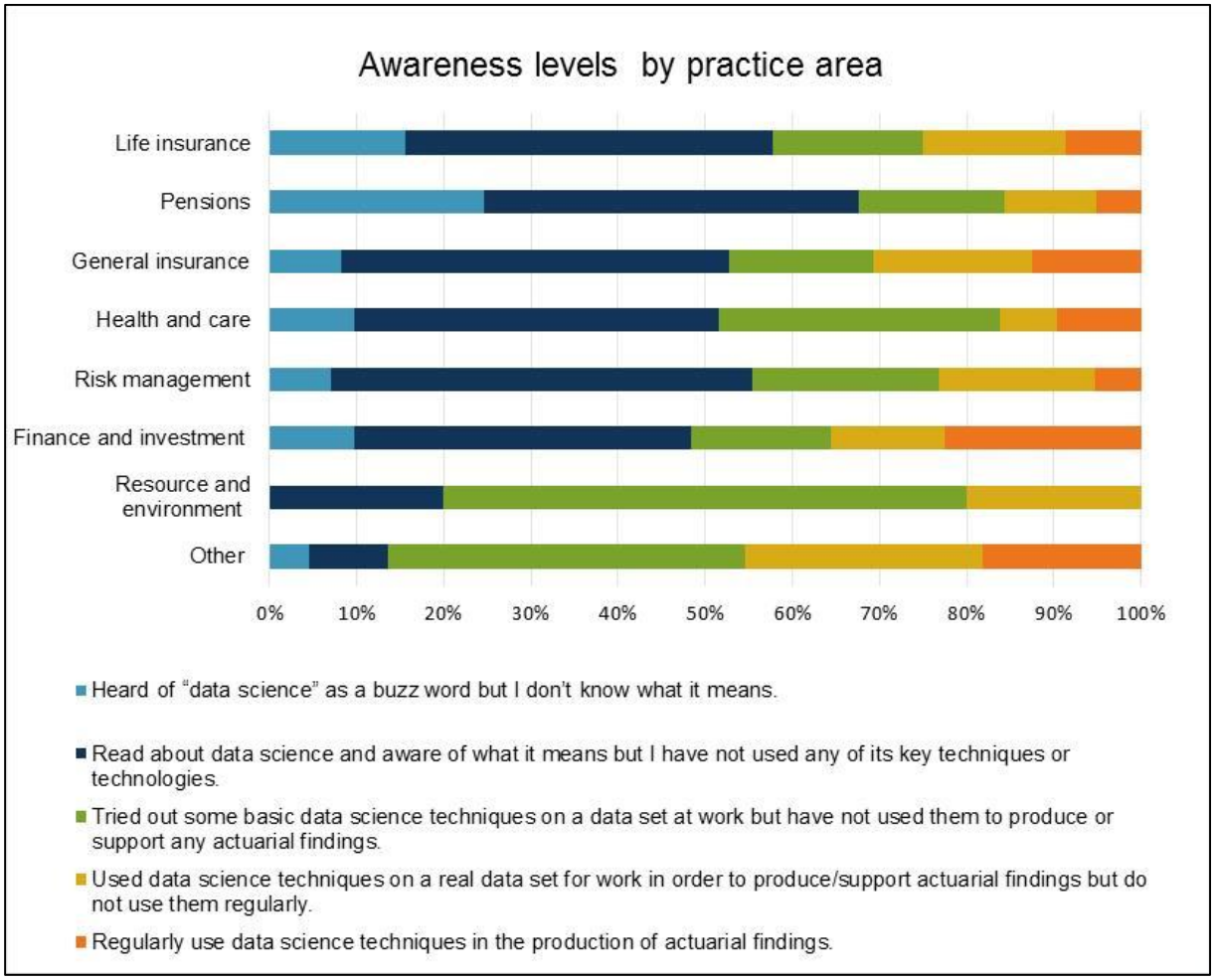

Figure 6

(iii) Awareness levels by country

Responses from different regions of the world where there are at least 20 in the response group.

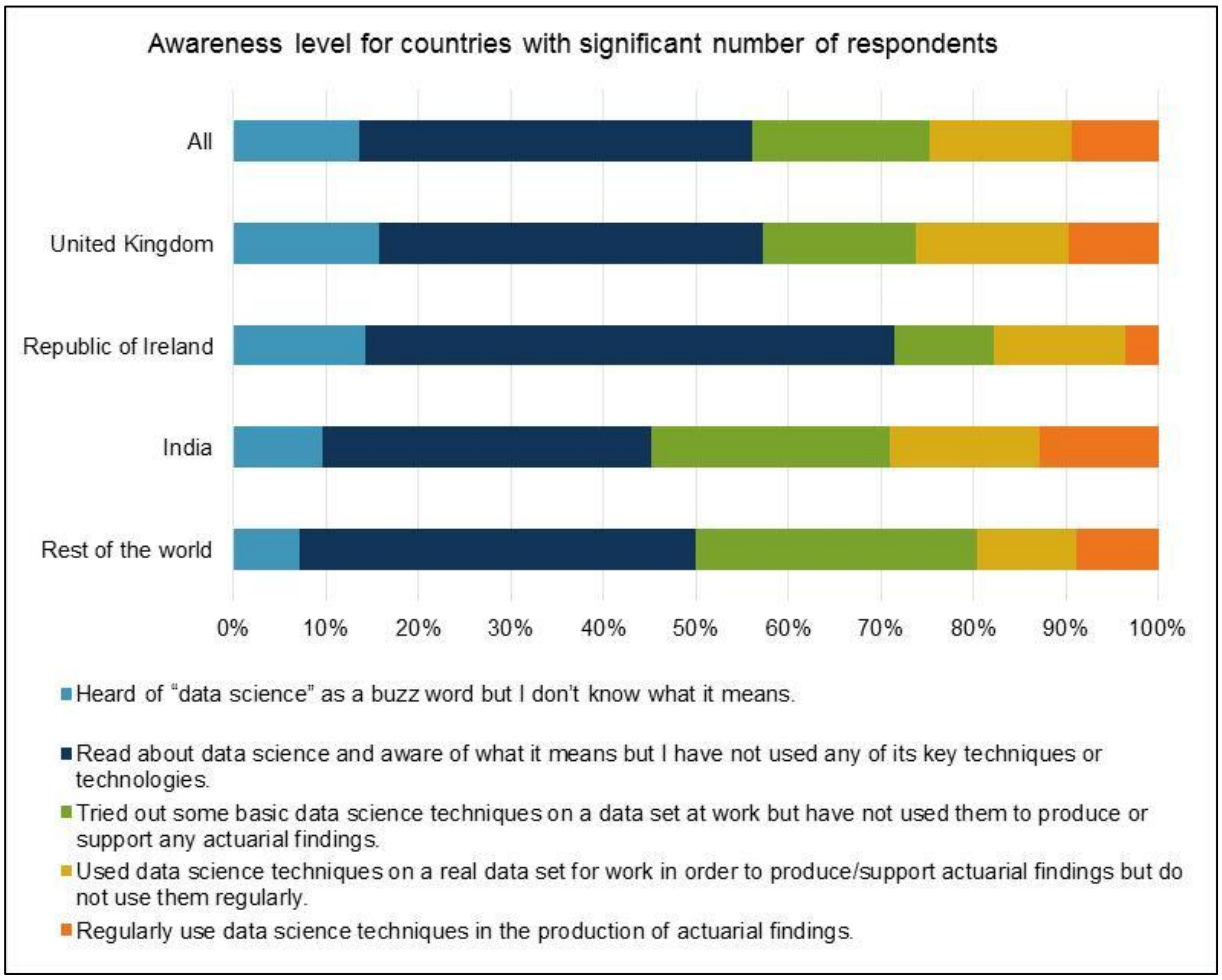

Figure 7

APPENDIX C: Defining the data science universe

(i) Big data

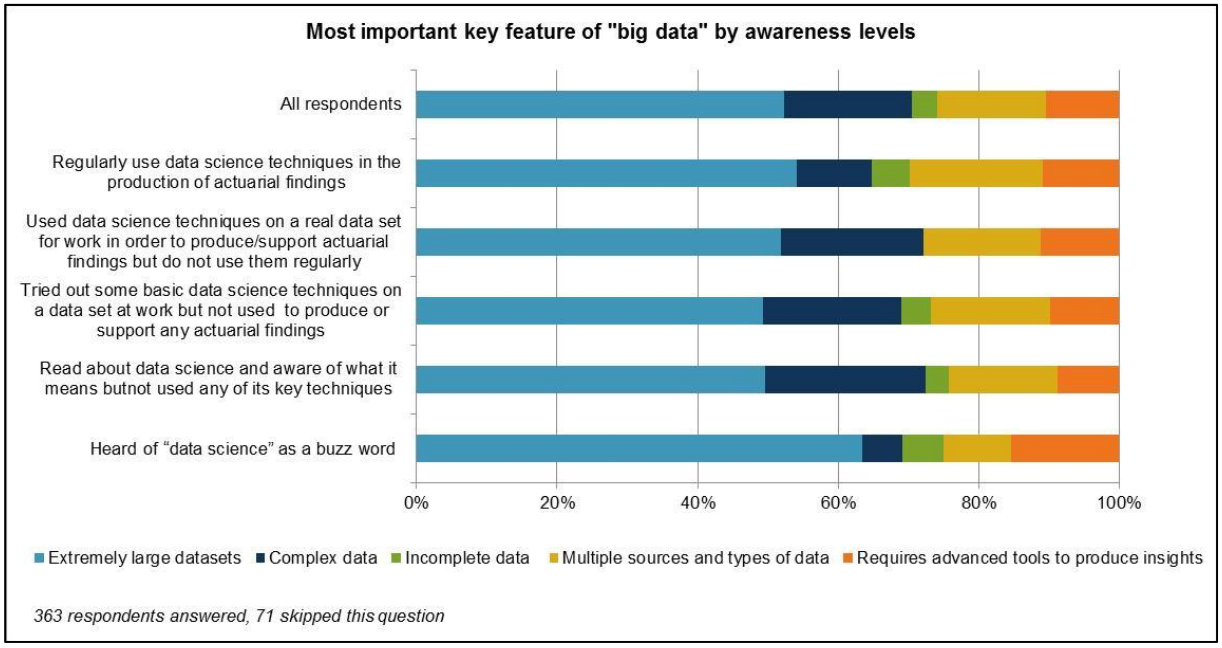

Figure 8

Based on those surveyed:

- 190 respondents (52%) viewed "Extremely large datasets" as the most important characteristic of "big data" overall
- 66 respondents (18%) viewed "Complex (e.g unstructured) data" as the most important characteristic of "big data" overall
  56 respondents (15%) viewed "Multiple sources/types" as the most important characteristic of "big data" overall
- 38 respondents (11%) viewed "Advanced tools" as the most important characteristic of "Big Data" overall
- 13 respondents (4%) viewed "Incomplete data (e.g. missing data)" as the most important characteristic of "big data" overall

"Extremely large datasets" was also deemed the most significant feature of "big data" by both those who identified themselves as "working regularly" with data science techniques and also across each industry.

Only 13 respondents (4%) deemed the most significant characteristic of "big data" to be "Incomplete data (e.g. missing data)." This was also true when considering responses from those who identified themselves as "working regularly" with data science techniques, of which only 3 respondents (5%) deemed it to be the most important characteristic.

Based on responses split out by practice area, "Incomplete data (e.g. missing data)" was also considered most significant characteristic by the fewest respondents across all practice areas, apart from Finance and Investment.

"Extremely large datasets" was clearly deemed the most popular choice, from the options presented, whereas, the relative significance of each of the four other characteristics is much closer.

Other characteristics proposed by respondents included:

- The 3 Vs: Velocity, Volume, Variety
- New technology with high processing capabilities, which is very accessible
- High quantity of data, without high quality of data.
- Tools and techniques to deal with complex datasets from many sources in different format (i.e. numerical and non-numerical)
- Continuously growing, in relation to data size and capabilities
- Real-time updating / live feeds and analysis
- Use of all data, not just a subset and therefore reduction of extremely large data sets to a smaller group of summary statistics
- Exceeds capacity of basic tools (including Excel manipulation) and is too big for humans to grasp
- Cloud computing
- Data mining
- IT architecture

(ii) Predictive analytics

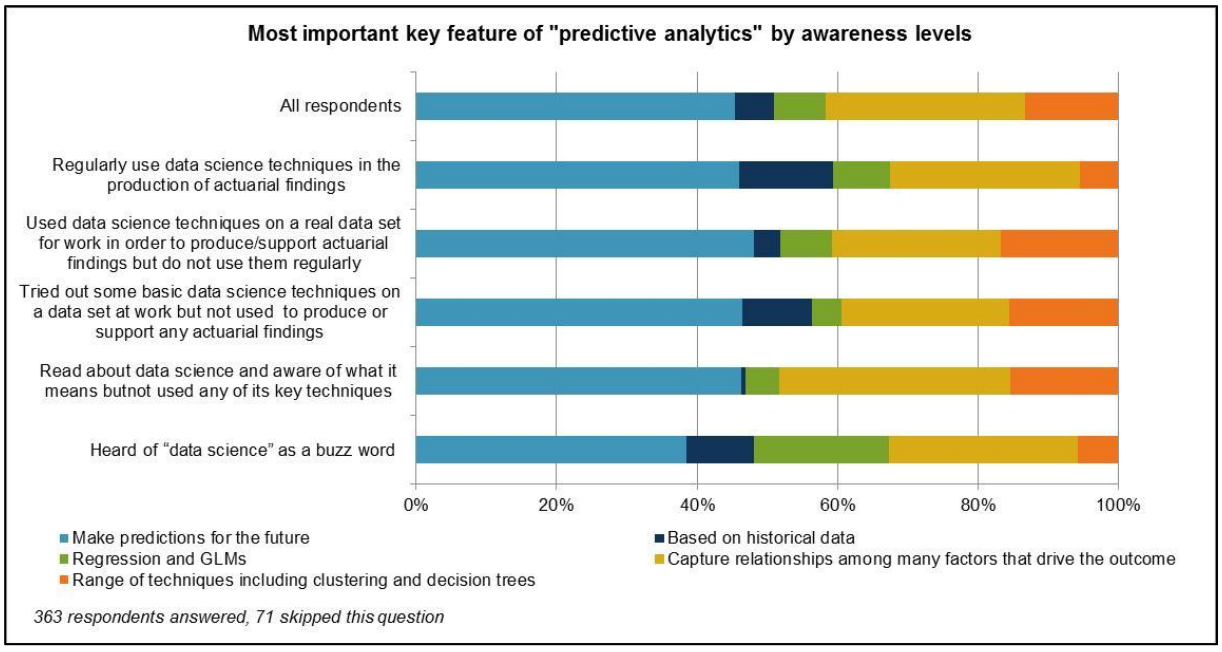

Figure 9

Based on those surveyed:

- 165 respondents (46%) deemed "Make predictions" to be the most important characteristic of "predictive analytics"
- 103 respondents (28%) deemed "Capture relationships among many factors that drive the outcome" as the most important characteristic of "predictive analytic"
- 48 respondents (3%) deemed "A range of techniques including clustering and decisions trees" as the most important characteristic of "predictive analytics"
- 27 respondents (7%) deemed "Regression and GLM models" as the most important characteristic of "predictive analytics"
- 20 respondents (6%) deemed "Based on historical data" as the most important characteristic of "predictive analytics"

"Make predictions" was also the most common response from respondents who "worked regularly" with data science techniques and also appeared to be broadly consistent across all practice areas.

Other characteristics proposed by respondents:

- Output prediction through regression and also clustering and classification
- Regression of output on the independent variables
- Techniques that deliver greater insights relative to traditional GLMS
- Predictions, rather than just extrapolation of trends
- Analysis of historical data to make predictions about future events
- Mechanised search for correlations that might improve predictions
- Less assumptions (but still dependent on assumptions)
- Problem solving
- Construct verifiable hypothesis from analysis and modelling
- An outcome based model

(iii) Machine learning

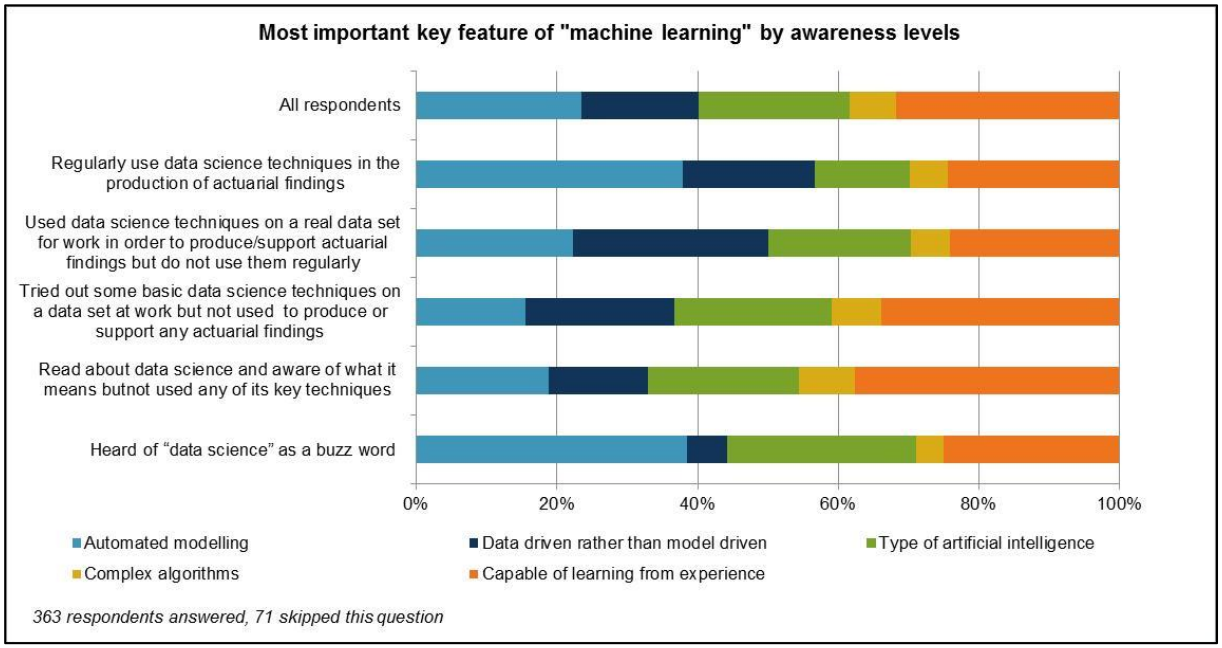

Figure 10

Based on those surveyed:

- 115 respondents (32%) deemed "Capable of learning from experience" the most important characteristic of "machine learning"
- 85 respondents (23%) deemed "Automated modelling" the most important characteristic of "machine learning"
- 78 respondents (21%) deemed "Type of Artificial Intelligence" the most important characteristic of "machine learning"
- 61 respondents (17%) deemed "Data Driven" the most important characteristic of "machine learning"
- 24 respondents (7%) deemed "Complex algorithms" the most important characteristic of "machine learning"

However, when considering the responses of those who deem themselves as “working regularly” with data science techniques, 14 respondents (38%) deemed “Automated modelling” to be the most important feature. “Capable of learning from experience” was deemed most important characteristic by 9 respondents (24%).

In general, the responses by practice areas followed a similar pattern to the results above, with the exception of members who classified themselves as “Other”, who deemed “Automated modelling” to be the most significant feature of “Machine learning”.

Other characteristics proposed by respondents:

- Supervised and unsupervised learning
- Iterative improvement of model using semi-automated algorithms
- Machine led feature discovery, rather than human led - Identifies features that humans are unable to.
- Computer science subject meets pure statistics
- Data used to parameterise model & also choose between competing models.
- Robotics
- Feature engineering
- Data mining
- Unstructured analysis

APPENDIX D: Learning about the data science universe

(i) Methods used to find out more about data science or big data

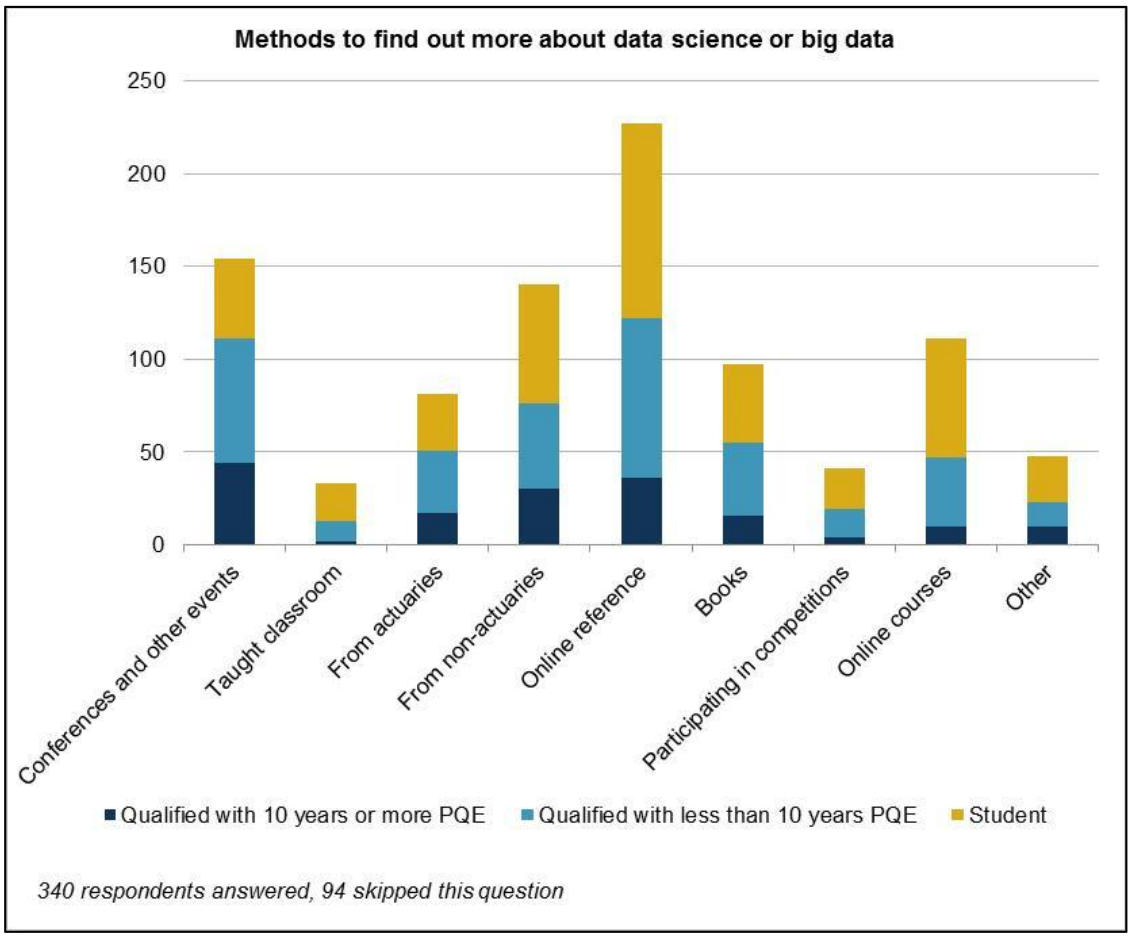

Figure 11

The most popular method of learning more about data science and big data is online reference material (including blogs, YouTube etc.). Conferences, seminars or other events were also a popular method of learning more about data science and big data according to survey respondents.

Fewest respondents were taught about data science or big data in a classroom learning environment or participated in competitions.

Overall, students were the most common respondents for learning through the different methods. The exceptions to this were conferences, seminars and other events and learning from actuaries, where qualified actuaries with less than 10 years' experience were the most common respondents learning through these methods.

Actuaries with more than 10 years' experience have the lowest levels of participation across all methods of learning.

Other methods used to find out more about data science or big data:

- The Actuary Magazine
- University / Post-graduate courses (e.g. Masters in Statistics / PhD in Machine Learning)
- Technical user groups
- Articles from consultancies

- Learning from statisticians and software engineers
- Journals
- News articles

Online courses used to find out more about data science or big data specified by respondents:

- Coursera
- Udemy
- DataCamp
- Udacity
- Big Data University
- MOOCs
- Edx

APPENDIX E: Current practice and actions from the IFoA

(i) Role overlap with common data analytics job titles

Survey respondents were asked to assess their current role against definitions of a data analyst and a data scientist, given the following:

Data analyst: A person whose job is to inspect, clean, transform, visualise, and model data with the goal of discovering useful information, suggesting conclusions, and supporting, rather than leading, decision-making.

Data scientist: A person whose job is to extract knowledge and actionable insights from data in various forms, using a combination of techniques from multiple disciplines, especially computer science, mathematics and business. They take responsibility for the full data science process, from the identification of suitable sources of data to address a specified problem, to the development and communication of an insight or data product that solves the problem. They may manage the work of data analysts.

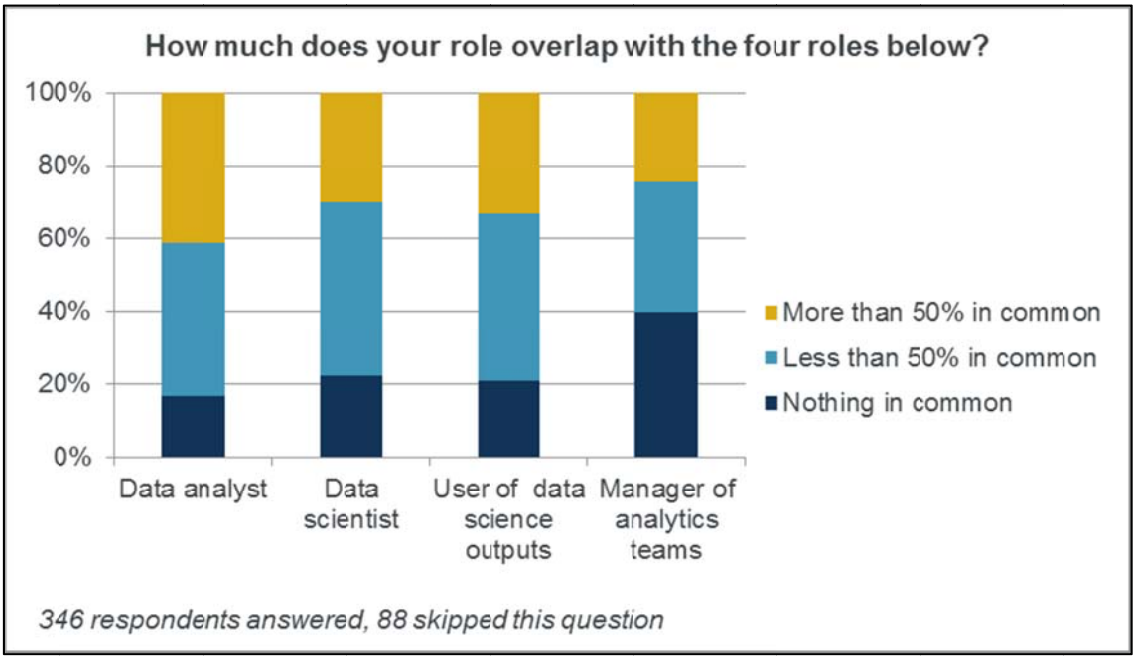

Figure 12

Based on those surveyed:

- 41% (142 respondents) view their current role as having more than 50% in common with that of a 'data analyst'
- 30% (103 respondents) view their current role as having more than 50% in common with that of a 'data scientist'
- 33% (113 respondents) view their current role as having more than 50% in common with that of a 'user of data science outputs'
- In contrast, 40% (133 respondents) view their current role as having nothing in common with 'manager of analytics teams'

(ii) Analysis Split by Practice Area

Consistent with the responses to other questions, those who work in general insurance and health and care appeared to identify more with the provided job titles compared to other practice areas.

For example, 48% (80 respondents) of general insurance actuaries view their role as having more than 50% in common with ‘data analyst’. This compares to just 37% of life actuaries (43 respondents), 39% of pension actuaries (30 respondents) and 38% of risk management actuaries (21 respondents). The other definitions showed a similar trend.

Similarly only 10% (3 respondents) of those who work in health and care view their roles as having nothing in common with ‘data analyst’. However, 21% of life actuaries (24 respondents), 17% of pension actuaries (13 respondents) and 24% of risk management actuaries (13 respondents) view their role as having nothing in common with ‘data analyst’.

Those who work in consulting or business advisory appeared to identify less with the provided roles compared to those who work directly for an insurance or reinsurance company. For example, 45% (84 respondents) of those who work directly for a (re)insurer view their role as having more than 50% in common with ‘data analyst’. However just 31% of consulting actuaries (31 respondents) have the same view. The other definitions showed a similar trend.

(iii) Analysis Split by Career Stage

53% of students (79 respondents) view their role as having more than 50% in common with ‘data analyst’. However, only around 30% of students view their role as having more than 50% in common with ‘data scientist’ or ‘user of data science outputs’ (42 respondents and 45 respondents, respectively).

Only 29% of those with more than 10 years post-qualification experience (21 respondents) view their role as having more than 50% in common with ‘data analyst’. However, 41% of those with more than 10 years post-qualification experience (29 respondents) view their role as having more than 50% in common with ‘manager of analytics teams’, which is perhaps not unexpected given they are likely to be in more senior positions in in their organisations.

(iv) Use of data science techniques for actuarial purposes

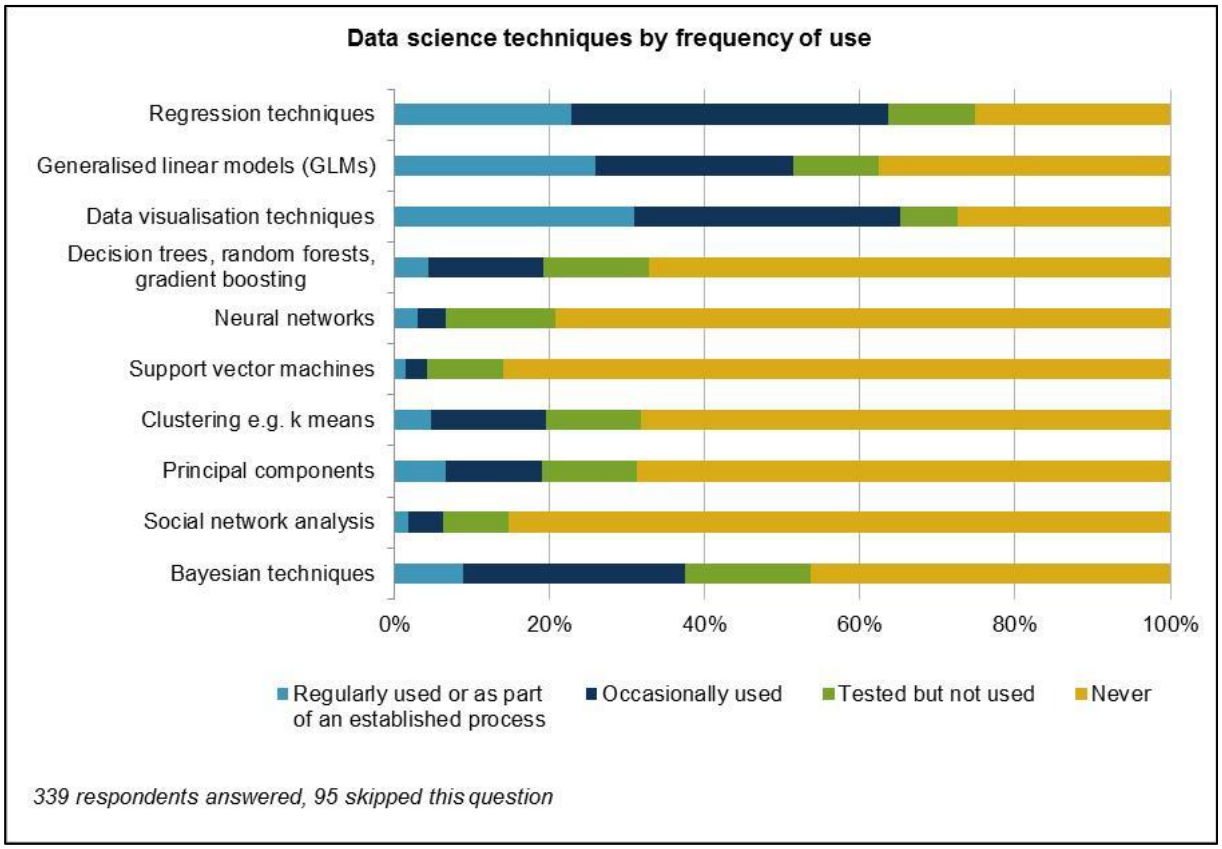

Figure 13

64% of respondents (215) said that they use regression techniques and data visualisation techniques occasionally or regularly whereas 51% of respondents (173) said that they use generalised linear models (GLMs) occasionally or regularly. These results are unsurprising as regression and GLM techniques have been used in insurance for several decades.

37% of respondents (126) said that they use Bayesian techniques occasionally or regularly.

Only 19% of respondents said that they use decision trees, random forests, gradient boosting (65); clustering (66) and principal components (64) occasionally or regularly and less than 10% said that they use neural networks (22), social network analysis (21) and support vector machines (14) occasionally or regularly.

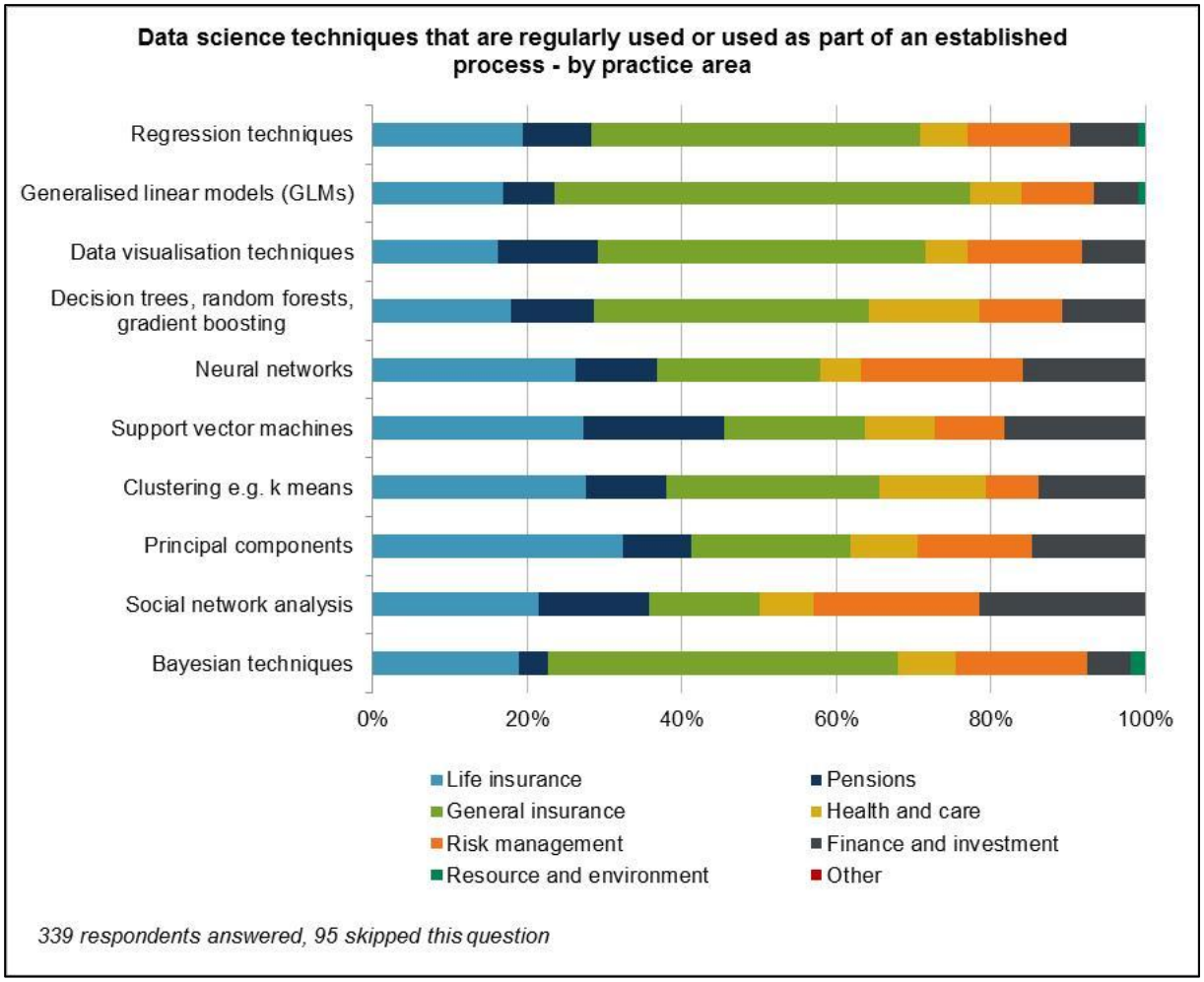

Figure 14

The data science techniques listed are used regularly or as part of an established process by respondents from virtually all practice areas.

The most common data science techniques used by General insurance practitioners are generalised linear models (GLMs) (64 responses), data visualisation techniques (63 responses), regression techniques (48 responses) and Bayesian techniques (24 responses).

The most common data science techniques used by Life insurance practitioners are data visualisation techniques (24 responses), regression techniques (22 responses), generalised linear models (GLMs) (20 responses), principal components (11 responses) and Bayesian techniques (10 responses).

(v) Additional comments or ideas on what the IFoA should be doing in respect of the data science universe

95 survey respondents left a comment. A key theme in many of the comments was providing more education to pre- and post-qualified actuaries. The major themes, with exemplar comments, identified were (number of comments relating to theme shown in brackets):

- **A call for the IFoA to facilitate training, seminars and learning material in data science** (35)
    - *"Perhaps offering online training courses and making suggestions about how actuaries can use data science in their day to day work."*
    - *"This is definitely a subject that the profession should embrace and train the members on. Not only the future qualified actuaries but also current actuaries who will have to adapt in the short term."*
    - *"Can IFOA help its students with online courses or job openings on data science technology?"*
    - *"Raise awareness, introduce real world examples of uses and facilitate more education."*

- **Introduction of data science into the IFoA curriculum, i.e. pre-qualification training** (28)
    - *"Data Science should be brought as a specialism in the later stages of the qualification process! Programming should be made a mandatory CT subject!"*
    - *"Include as an ST subject"*
    - *"Consider updating the syllabus and open up training courses for experienced members."*
    - *"Introduce more data science into the exam syllabus. Also data visualisation and 'story telling' to reach conclusions from data"*
    - *"I think it's essential that we engage with this. I think we need to expect current and future generations of actuaries to be comfortable with analytics and statistical computing in general. The education syllabus should support this - and we should consider "update" course for fellows."*

- **Training in R or other programming languages** (7)
    - *"The profession needs to stop using excel. It is not a statistical piece of software. Promote use of R"*
    - *"Data Science should be brought as a specialism in the later stages of the qualification process ! Programming should be made a mandatory CT subject !"*
    - *"I would like to see examples provided if these methods in R to encourage actuaries to try out these new approaches"*
    - *"Coding skills in R or Python needs to be learnt to apply these data science skills. We also need to teach actuaries the IT/Data architecture bits such as Hadoop - so computer science needs to become a part of the syllabus, fast!"*

- **Ethical, legal and governance issues of data science and working with data** (6)
    - *"I think we need to consider further the moral use of big data. When I say moral use, I mean the risk of locking people out of insurance due to their risk factors established from big data and predictive analytics. My worry is that we narrow risk factors down too much to individuals, when the whole insurance business works on grouping of risk.*
    - *"Courses on data science, teaching computing languages, making it easier for non-specialists to enter this field, collaboration with other industries, bearing in mind need to link data science work to other important aspects such as governance and documentation, case studies on good use of data analytics in insurance, etc."*
    - *"Provide guidance on data protection in relation to big data"*

- **The risk that actuaries are replaced by data scientists** (5)
    - *“I think Data Science could be seen as an opportunity for actuaries to enhance their skill set or as a threat to traditional actuarial work if we don't manage to take that opportunity. I think the IFoA is right to have set up MAID. I wonder if there would be merit in opening discussions between IFoA and other practitioners of Data Science to discuss benefits of sharing our skill sets with each other. It may be worth Data Science becoming a separate Specialist Subject, like ERM.”*

Institute
and Faculty
of Actuaries

London

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100 · Fax: +44 (0) 20 7632 2111

Edinburgh

Level 2 ·Exchange Crescent · 7 Conference Square · Edinburgh ·EH3 8RA
Tel: +44 (0) 131 240 1300 · Fax +44 (0) 131 240 1311

Oxford

1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD
Tel: +44 (0) 1865 268 200 · Fax: +44 (0) 1865 268 211

Hong Kong

2202 Tower Two · Lippo Centre · 89 Queensway · Hong Kong
Tel: +11 (0) 852 2147 9418 · Fax: +11 (0) 852 2147 2497

Beijing

6/F · Tower 2 · Prosper Centre · 5 Guanghua Road · Chaoyang District · Beijing · China 1000020
Tel: +86 (10) 8573 1000

Singapore

163 Tras Street · #07-05 Lian Huat Building · Singapore 079024
Tel: +65 6717 2955

www.actuaries.org.uk