Morningstar Risk Profiler
FinaMetrica Psychometric Risk Tolerance Methodology

Morningstar Inc.
01 June 2021

Contents

Shawn Brayman
Director of Financial Planning Methodology
Financial Planning
Advisor Client Experience (ACE)

Nicki Potts
Director of Product Management
Financial Planning,
Advisor Client Experience (ACE)

1. Introduction

This report reviews the risk-tolerance scale and the scoring algorithm of the Morningstar Risk Profiler, which uses the FinaMetrica psychometrics risk-tolerance assessment. Risk tolerance references the psychological trait of a consumer that measures their willingness to take investment risks, measured by the FinaMetrica methodology. This is a foundational aspect of the consumer's overall risk profile, which combines the psychological attribute with other financial and factual considerations, such as risk required, risk capacity, time horizon, prior investment experience, financial knowledge, and more, to arrive at an aggregate treatment for a portfolio or group of accounts, called a Suitability Score (the Morningstar Risk Profiler). A consumer has one tolerance but can have different portfolios of funds dedicated to different goals and therefore has different Suitability Scores for each. This document should be read as a component of the overall Morningstar Risk Profiler Methodology document and specifically addresses the FinaMetrica psychometric testing of the psychological tolerance for financial risk.

Risk tolerance is measured following completion of either a 25-question self-reported questionnaire or a 10-question version.[1] The test approach is norm-referenced, therefore, the scores are interpreted relative to other people's scores. In other words, the test reports how much lower or higher the level of risk tolerance of a respondent is compared with the average level of risk tolerance of the population of interest. In this report, the population of interest is defined as adults between the ages of 20-80 who seek to understand their risk tolerance either directly or have sought financial advice from a financial advisor/planner. To estimate the average level of risk tolerance in the population, a sample of FinaMetrica data is used (this is the norm group).

Currently there is a world scale that is used for all countries. The data for the norm group is predominately from the U.K., U.S., Australia, New Zealand, Canada, India, DACH (Germany, Austria, and Switzerland), South Africa, and Ireland, for the period September 2011 to July 2016. The scoring algorithm uses 22 of the 25 questions (questions 11, 23, and 24 are new questions being evaluated and excluded from scoring) and consists of the following steps:

1. After the norm group/data has been specified, the mean and standard deviation for 22 questions are computed to be used in step 2.
2. The z-scores for the 22 questions for all individuals in the norm group are computed, where z-score in question A of individual X = (answer of individual X to question A – mean of question A) / standard deviation of question A.

3. For all individuals in the norm group, the sum of their z-scores is computed as well as the standard deviation of the sums of the z-scores. The latter is used in step 4.
4. For all individuals, the z-score of their sum of z-scores is computed by dividing their sum of z-scores with the standard deviation[2] of the sums calculated in step 3. Let us refer to the z-score of the sum of the z-scores as the total z-score.
5. For all individuals in the norm group, the total z-score is multiplied by 10 and then 50 is added to the product. Let us refer to the transformed total z-score as the total score. The last step is done so that the total scores have mean 50 and standard deviation 10.

The distribution of the total scores of the norm group is used in order to interpret the total score of a new respondent. For example, if a new respondent scores 60, we can say that this person exhibits higher risk tolerance by one standard deviation compared with the average level of the norm group (which is 50); if the score is 45, then it is lower by half a standard deviation compared with the average level of the norm group. For this, the norm group needs to be representative of the potential future test-takers.

The target audience of this report is financial advisors and those undertaking due diligence of the Morningstar Risk Profiler. The report, in addition to the Introduction, consists of eight sections. Each section starts with a summary of the section's findings followed by a detailed discussion of the analysis and the results. A brief outline of each section is given below.

- Section 2 summarizes the main findings of all of the sections.
- Section 3 introduces the available data and the part of which is kept for analysis.
- Section 4 describes the demographic profile of the analyzed data (gender, age, education level, income, marital status, value of net assets, number of family members financially dependent on the respondent, country of residency, and year of taking the risk-tolerance test).
- Section 5 focuses on identifying associations of risk tolerance with demographic characteristics. This is mainly to a) spot major differences across countries that will lead to separate risk-tolerance scales, b) investigate differences over time so that we will select the data for the norm group accordingly, c) determine the cases to be included in the norm group data, and d) describe associations of risk tolerance with different demographics that may inform financial advice and decisions.
- Section 6 carries out a construct validity analysis to confirm that the questions measure risk tolerance.
- Section 7 conducts a reliability analysis to assess whether and to which degree the questionnaire consistently measures risk tolerance.
- Section 8 confirms the use of the norm groups and validity of the scoring algorithm.
- Section 9 compares the risk tolerance scores based on the 10-question version of the questionnaire with the scores derived by the full version of the questionnaire (25 questions) to assess whether the shorter version of the questionnaire, which is already administered to clients, measures the risk tolerance as accurately as the full version of the questionnaire.

2. Executive Summary

The data utilized covers the period September 2011 through July 2016. Duplicated cases, nonrelevant cases, and cases with nonlegitimate variable values (such as, outside of the age range of 20-80) were removed. We were left with 407,016 cases from the initial 541,549 cases.

Describing the data at high level

The data originates geographically from the U.K., Australia/New Zealand, U.S., India, DACH, Canada, South Africa, Ireland, Hong Kong, China, Sweden, Finland, Malaysia, and Kuwait (the countries are listed in order of data size) and is predominantly sourced from clients of financial advisors/planners who have completed a FinaMetrica psychometric risk-tolerance assessment. The bulk of the data, 91%, comes from the U.K., Australia/New Zealand, and the U.S., while 8.7% is from India, DACH, Canada, South Africa, and Ireland. Eighty-two percent of the data spans just four years, from 2012-15, and is almost evenly distributed to the respective years.

The data is:

- Slightly male-dominated (56.7% male respondents).
- Seventy percent are 40-69 years old, and 60% are university graduates or postgraduates.
- Three quarters are married or in a relationship.
- Sixty-three percent financially do not support (fully or partially) anyone or just one family member.

It is worthwhile noting that about one third of the respondents did not provide their demographic characteristics and subsequently they are not included in the above numbers. It is also worth noting that India has a distinctive demographic profile where the respondents appear the youngest, most educated (about 90% have a university degree or higher) and are male in majority (78.5%).

For further analysis, only data from the U.K., Australia/New Zealand, U.S., India, DACH, Canada, South Africa, and Ireland, that is 99.7% of the 407,016 cases, was retained. Based on the analysis, there are no practically significant differences in the distribution of risk-tolerance scores across countries and over time. Zooming in on the country score averages, they all hover around 50, with the U.K. having the lowest average score, 47.8, and India and South Africa having the highest average scores at 53.5.

With respect to the rest of the demographics, higher levels of risk tolerance are found to be associated with men, younger ages, and people with higher income and higher education. Men, on average, score higher than women by around five points, and this difference roughly remains after controlling for income (for a given income category). The average score of the highest income category is around nine points higher than the average score of the lowest income category. This difference roughly remains the same when the analysis is done by country, with the exception of Ireland and India, where the differences of the average scores are 14 and 5, respectively.

The average risk-tolerance score of age group 20-29 is about seven points higher than the average score of age group 70-80. The average score of the most educated people is higher by six points compared with the average score of the least-educated people.

The construct validity analysis shows that all scored questions, (excludes questions 11, 23, and 24), measure the same concept that can be interpreted as risk tolerance. All scored questions are found to be reliable indicators measuring risk tolerance consistently.

A general world scale is considered sufficient for scoring purposes because the maximum absolute difference between a score on a regional subscale and a score on the world scale is 5.66. The largest difference mean occurs in India and it is 4.10 (for example, a score on India scale is, on average, 4.10 higher than the same score on the world scale). The regional differences are considered small for practical application purposes and so further subscales were not developed.

Exhibit 1 below gives a brief summary of the seven risk groups.

Exhibit 1 Seven Risk Group

Risk Group	Score Range	Percentage in Risk Group	Brief Description of Risk Group
1	0 – 24	1%	People with **extremely low** risk tolerance.
2	25 – 34	6%	People with **very low** risk tolerance.
3	35 – 44	24%	People with **low** risk tolerance.
4	45 – 54	38%	People with **average** risk tolerance.
5	55 – 64	24%	People with **high** risk tolerance.
6	65 – 74	6%	People with **very high** risk tolerance.
7	75– 100	1%	People with **extremely high** risk tolerance.

The scoring algorithm gives equal weights to the z-scores of all 22 questions. A weighted scoring algorithm with larger weights to the z-scores of questions with higher reliability to measure risk tolerance was also examined, but the conclusion was that it is not needed as, on average, the weighted score was the same as the unweighted one.

The scores produced using a 10-question short form (questions 1, 2, 3, 6, 10, 13, 14, 16, 18, and 21) are found to be very close to the scores based on the 22 scored. The 10-question score is within 5 units from the 22-question total score for 90.6% of cases of the world norm group, while 76.7% of cases are assigned to the same risk group by both scores (22-question form and 10-question form).

3. Cleaning the Data

The data set covers the period September 2011 through July 2016 and consists of 541,549 registers. Out of the complete data, we retained only cases that meet the following criteria:

- Clients of financial advisors.
- Completed the 25-question version of the questionnaire.
- For better data integrity, age range is limited to 20-80, inclusive.
- Ensured legitimate values in all variables. Only one case was found with nonlegitimate value for variable combined before-tax income and removed.
- Ensure Client Code is not missing and appears only once in the data set. Client Codes appearing more than five times in the course of five years, September 2011 to July 2016, were regarded as problematic because multiple measurements were taking place too close in time (for example, within the same year) or the answers to the questions differed substantially.

After the cleaning procedure, there were 407,016 cases (from 541,549) left.

4. Demographic Profile

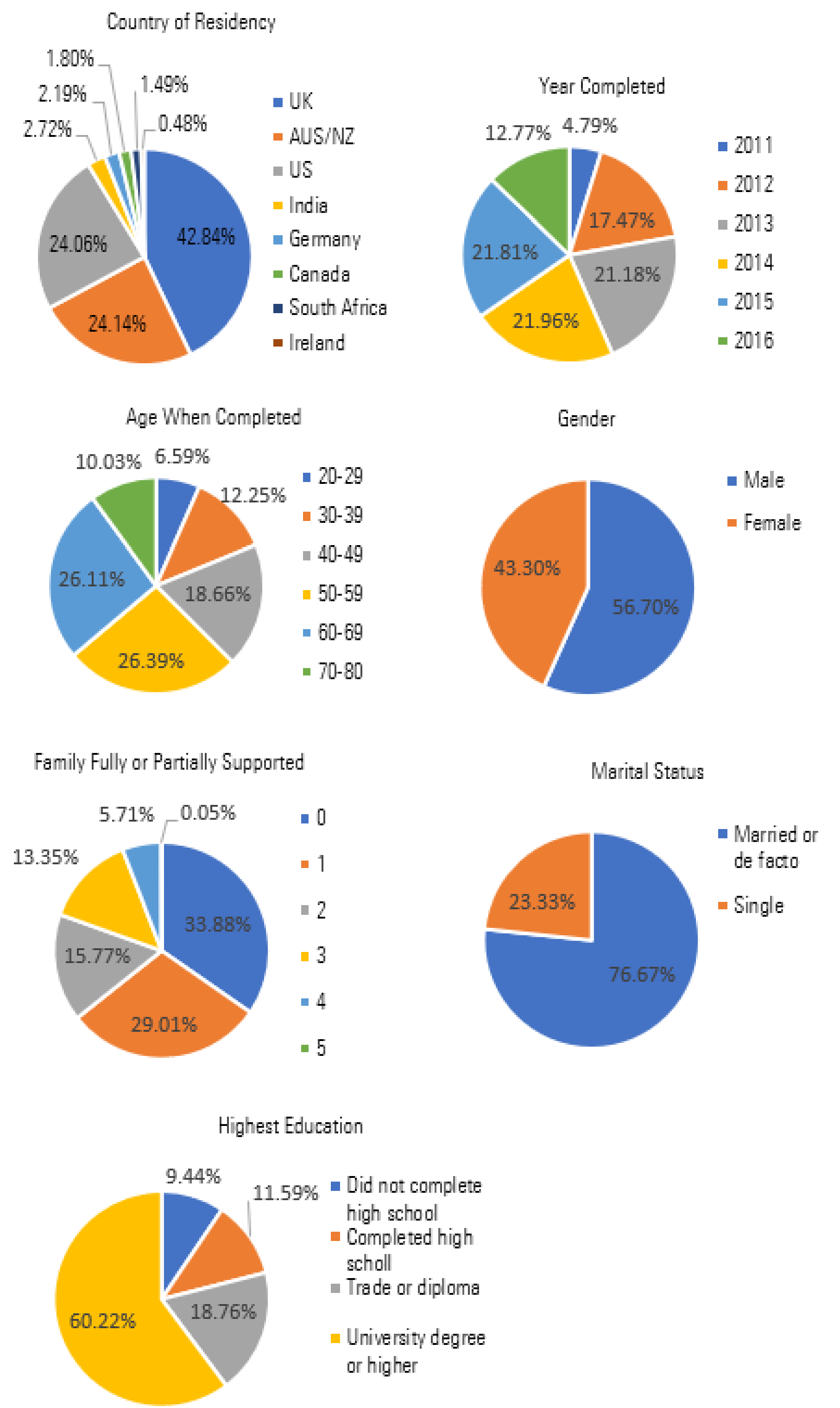
This section describes the demographic profile of the data by providing graphs accompanied by brief explanations.

Before proceeding with a detailed description, a snapshot of the demographic profile of the data is given below.

- The bulk of the data (91%) comes from the U.K., Australia-New Zealand (ANZ), and the U.S.
- Eighty-two percent of the data relates to 2012-15 and is almost evenly distributed to the respective years.
- Seventy percent of the respondents were between 40-69 years old when they completed the questionnaire.
- Gender: Male slightly outnumber female (56.7% men).
- Sixty percent of the respondents have a university degree or attained a higher education level.
- Seventy-seven percent report that they are married or in a relationship.
- Sixty-three percent of the cases support financially (fully or partially) at most one person in their family while 92% support at most three people.

Note that all aforementioned percentages and the graphs below are with respect to the total number of cases for which the related demographic has been reported. The percentage of cases out of the complete data set with missing value in each of the demographic variables is as follows: 32% for age, 31% for gender, 39% education level, 33% for marital/relationship status, and 40% for number of family dependents.

Exhibit 2 Demographic Profile of the Data

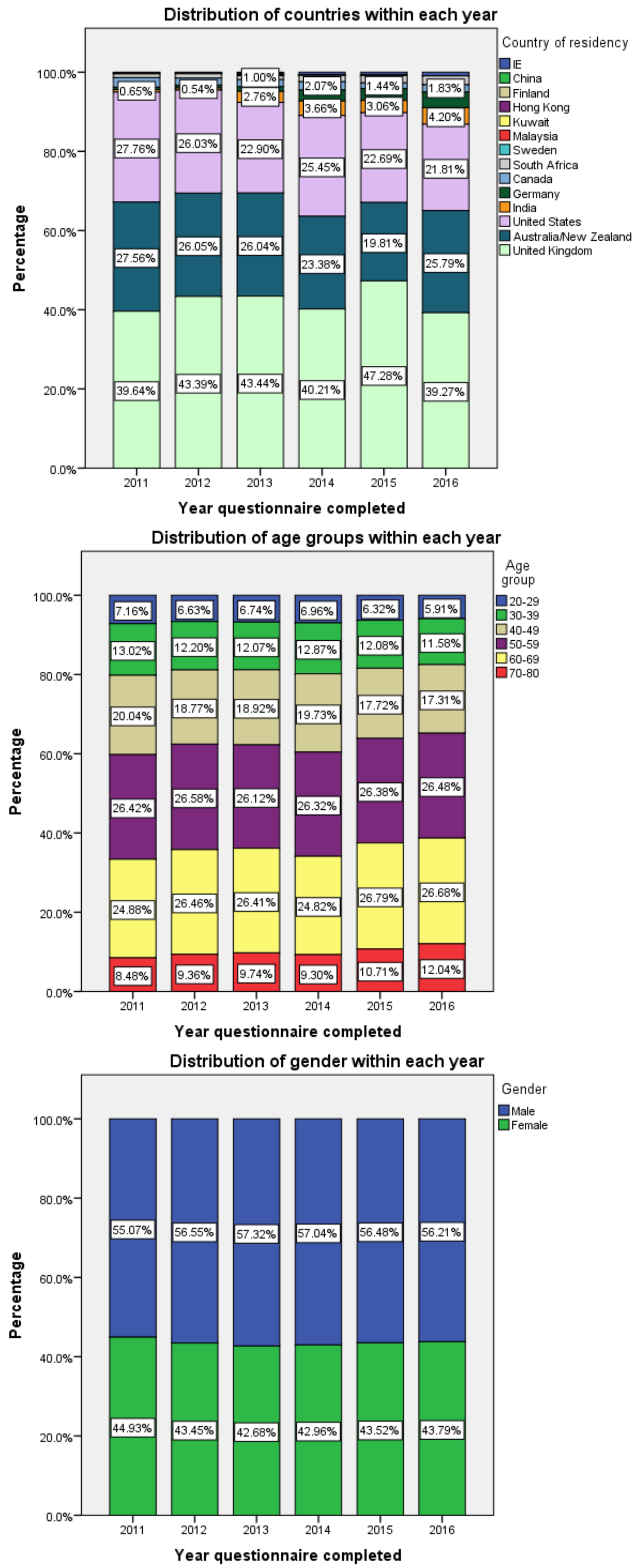


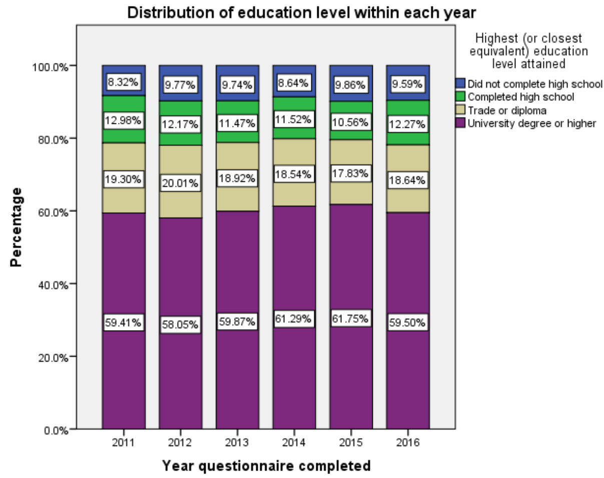
- Country: Ninety-one percent of the data comes from the U.K., Australia/New Zealand (ANZ), and the U.S., with U.K. being the largest group (43.0%), while the ANZ and U.S. data is of similar size (24.1% each). The rest of the data comes from India (2.7%,11,058 cases), DACH (2.2%, 8,903 cases), Canada (1.8%, 7,307 cases), South Africa (1.5%, 6,083), Ireland (0.5%,1,939 cases), Hong Kong, China, Sweden, Finland, Malaysia, and Kuwait. For the last six countries, the sample size ranges from 611 cases (for Hong Kong) to five (Kuwait).
- Year: Eighty-two percent of the data refers to 2012-15 and is almost evenly distributed to the respective years (2012--17%, 2013--21%, 2014--22%, 2015--22%) while, as expected, less data comes from 2016 (13%) and 2011 (5%). Recall that 2016 covers approximately the first half of the year (up to July) and 2011 covers only September through December.
- Age: Seventy percent of the cases in the data set are people who are 40-69 years old. The age groups 50-59 and 60-69 are the largest ones (26% each) followed by the 40-49 group (19%). The smallest group is that of age 20-29 (7%), while the groups 30-39 and 70-80 are of similar size (12% and 10%, respectively).
- Gender: Males (56.7%) slightly outnumber females.
- Education: Sixty percent of the respondents have at least a university degree while, together with those who have a trade or diploma, they make up 79% of the data. The remaining 21% is divided almost evenly into those who completed high school and those who did not.
- Marital/Relationship Status: Three quarters of the respondents in the data (77%) report that they are married or in a relationship.
- Number of Family Members Financially Supported (Fully or Partially) by Respondent: Sixty-three percent of the respondents support financially (fully or partially) none or one person in their family while 29% of respondents support two to three family members. Hence, 92% of the data supports at most three people.

Demographic Profile Over Time

Within each year, the distribution of the data to the three major countries—U.K., Australia/New Zealand, and U.S.—is about the same. The main difference over the years is that the data sets of India and DACH seem to grow but still represent less than 5% of the total data in a given year. Moreover, the distributions of age, gender, and education remain about the same over time. Hence, with respect to these variables, the demographic profile of the data remains roughly the same in the course of the period September 2011 through July 2016. Note that these results are based only on the cases for which the related demographic characteristics have been provided.

Exhibit 3 Demographic Profile Over Time





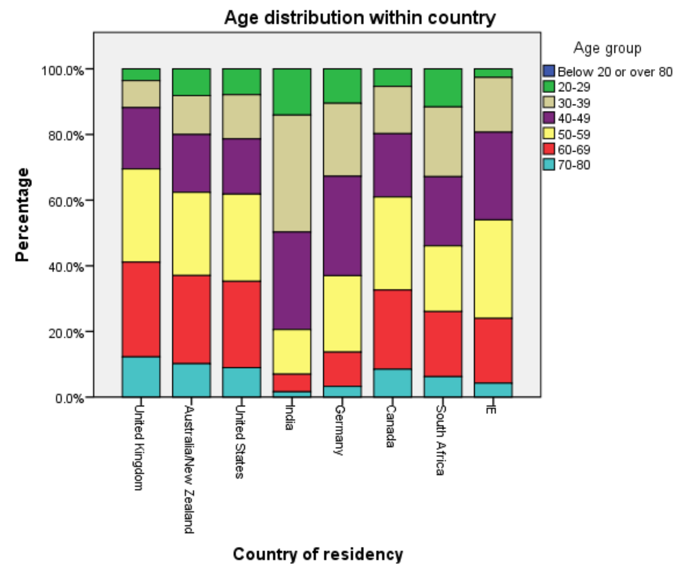
Demographic Profile Across Countries

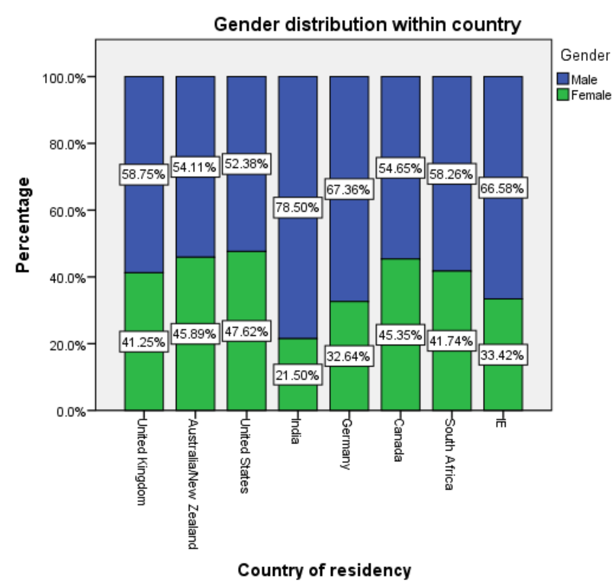
It is important to be aware of any differences in the demographic profile of the countries. This section presents the age, gender, personal income, and education profile of the countries that comprise 99.7% of the data set, such as the U.K., Australia/New Zealand (ANZ), U.S., India, DACH, Canada, South Africa, and Ireland.

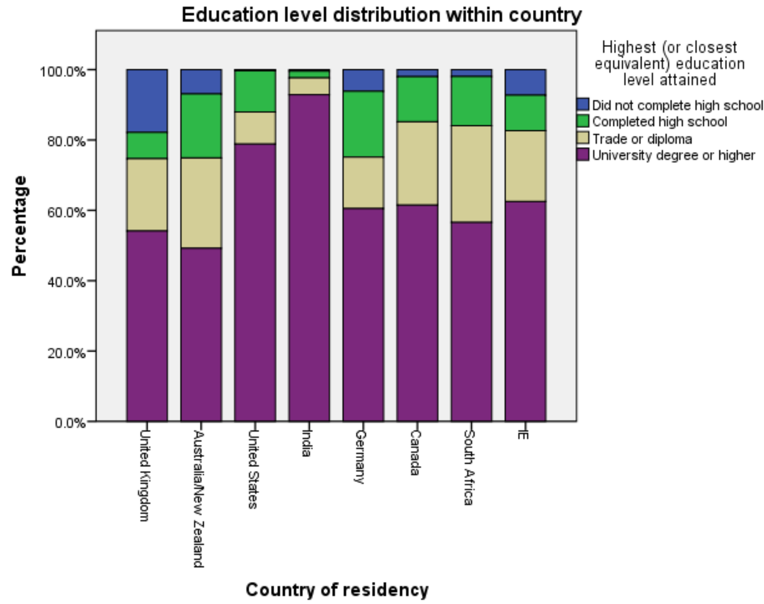
The graphs below show that the profile of Indian data is fairly different to that of the rest of the countries. Respondents from India are in comparison the youngest, most educated, and the majority are male, but with the lowest average personal income in face value (note that caution is needed when comparing the distribution of personal income across the countries, as the purchasing power may differ substantially). For the remaining countries, we observe the following:

- ANZ, U.S., and Canada have almost the same age profile, while the U.K.'s profile is slightly older, Ireland's slightly younger, and much younger than that of South Africa and DACH.
- Regarding gender, U.K., ANZ, Canada, and South Africa are similar, while Ireland's and DACH's data is more male-dominated.
- When it comes to education level, U.K. presents the largest percentage in the lowest education level (high school not completed), while the U.S. presents the largest, after India, percentage in the highest education level (university degree or higher).

Exhibit 4 Demographic Profile Across Countries







5. Data Analysis - Association of Risk Tolerance with Demographics

To determine the norm group, we needed to investigate whether the distribution of risk tolerance differed over time (from September 2011 to July 2016) and across countries and to identify the demographic characteristics that are associated with risk tolerance.

For the data set (September 2011 to July 2016) we keep the data from the U.K., Australia/New Zealand, U.S., India, DACH, Canada, South Africa, and Ireland as these countries make up 99.7% of the data (405,828 cases in total). The data size for each of the aforementioned countries is at least 1,900 cases. It ranges from 1,939 cases for Ireland to 174,383 for the U.K. We refer to this data as the Working Data Set.

The data of Hong Kong, China, Sweden, Finland, Malaysia, and Kuwait has been excluded from the Working Data Set. Hong Kong's data size is 611 cases, but it does not seem to be representative of the region, as 97% of Hong Kong respondents are believed to be expats. The data sets of China, Sweden, Finland, Malaysia, and Kuwait are still fairly small with 344, 162, 36, 30, and five cases, respectively.

The main results of the analysis presented in this section are:

- The differences in average risk-tolerance score across countries are of no practical importance. They all hover around 50, the mean value of the risk tolerance scale. India and South Africa exhibit the highest average score, 53.5, while U.K. exhibits the lowest average with 47.8.
- There is essentially no difference in the annual average score for the time period September 2011 through July 2016. The annual average score is 49.
- Men, those who are younger, and people with higher education tend to have a higher level of risk tolerance, on average.
- Men, on average, score higher than women by about five points.
- The average score of the highest income category is around nine points higher than the average score of the lowest income. This difference roughly remains the same when the

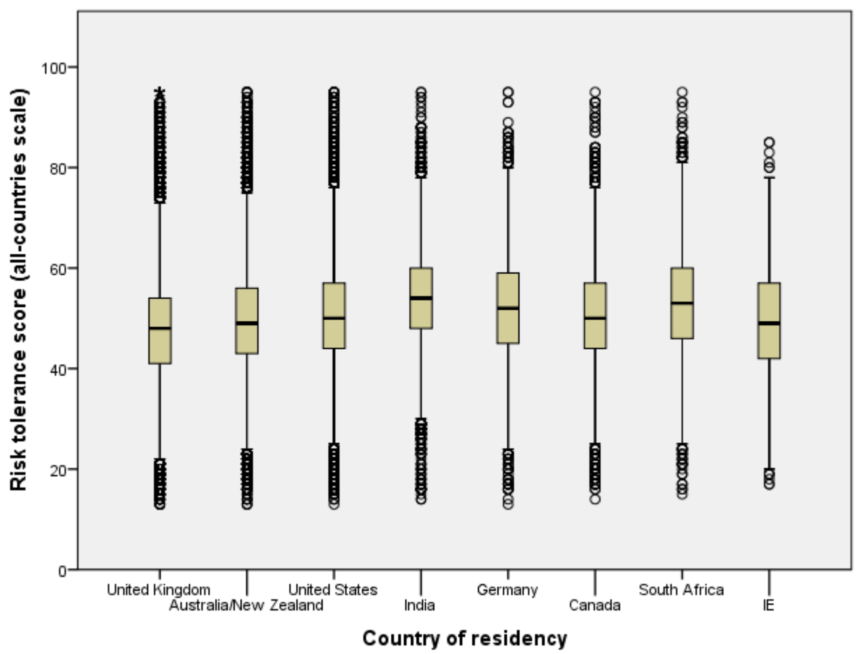
analysis is done by country, with the exception of Ireland and India, where the differences of the average scores are 14 and five points, respectively.

- The average risk-tolerance score of age group 20-29 is about seven points higher than the average score of age group 70-80.
- The average score of the most-educated people is higher by six points compared with the average score of the least-educated people.
- For most subgroups defined by the value of various demographic characteristics, the average score falls into the middle risk group four.

Differences in Risk Tolerance Across Countries

Our analysis below suggests that there are no differences of practical importance in the distribution of risk tolerance across countries. The country distributions can be considered roughly the same.

Exhibit 5 Distribution of Risk-Tolerance Scores by Countries



The graph above (Exhibit 5) presents the boxplot[3] of risk scores for each country. The distributions look similar with small differences in the score means. Exhibit 6 below summarizes the mean, median, standard deviation, minimum, maximum, skewness, and kurtosis for the observed distribution of risk tolerance by country. The standard errors of skewness and kurtosis are provided in parenthesis. There are small differences in the score means, with U.K. exhibiting the smallest mean, 47.79, and India the largest (53.34), but the difference is barely of six points. Such differences may be considered of no practical importance when regarding the investment implications, especially if it is taken into account that the score means are close to 50 (the mean of the scale) and fall into the same risk group, group 4, which represents people with average level of risk tolerance.

The country means are almost identical to the country medians, suggesting that the distribution of risk scores is close to symmetrical for all countries. The standard deviation for all countries is very close to 10, which is the standard deviation of the scale. The differences of the standard deviations are less than a unit. The country minimum and maximum values of risk score are almost identical. The skewness is a very small positive number for U.K., Australia/ New Zealand, U.S., Canada, and South Africa. This indicates an almost symmetrical distribution with a slightly longer tail to the right—such as, slightly more values to the top end of the risk-tolerance scale rather than to the bottom-end of the scale. India, DACH, and Ireland have a symmetrical distribution. The kurtosis is a small positive number for all countries, implying that there are more observations in the ends of the risk scale (for example, fatter tails) compared with a normal distribution with mean 50 and standard deviation 10. The Indian data seems to have more observations in the ends of the risk scale than in any other country, while the data of Australia/New Zealand practically follows a normal distribution with mean 50 and standard deviation 10.

Exhibit 6 Descriptive Statistics for Distribution of Risk Tolerance Scores by Country

Country	Data Size	Mean	Median	Std. Dev.	Min	Max	Skewness		Kurtosis	
UK	174383	47.79	48	10.053	13	95	.130	(.006)	.388	(.012)
AUS/NZ	98244	49.46	49	9.916	13	95	.155	(.008)	.377	(.016)
USA	97911	50.55	50	9.699	13	95	.128	(.008)	.359	(.016)
India	11058	53.75	54	9.838	14	95	-.047	(.023)	.633	(.047)
DACH	8903	51.81	52	10.550	13	95	.088	(.026)	.180	(.052)
Canada	7307	50.57	50	10.607	14	95	.131	(.029)	.429	(.057)
South Africa	6083	53.34	53	10.568	15	95	.113	(.031)	.155	(.063)
Ireland	1939	49.32	49	10.838	17	85	-.082	(.056)	.026	(.111)

standard errors in parenthesis

An ANOVA analysis (where the independent variable is risk score and factor is country) rejects the hypothesis of equality among country score means at any conventional significance level (p-value=0.000). However, we should bear in mind that large data sets like ours will almost always reject the hypothesis of equality as the power of the ANOVA test to detect that small differences across groups increases substantially with the size of the data. In other words, the data size is so large that the test picks up on very minor differences. Same reasoning applies when comparing pairs of countries where differences of less than 1.67 units are found to be statistically significant. In particular, the mean differences of the following pairs of countries are found to be statistically significant (p-value<0.001): U.K. and ANZ, ANZ and U.S., U.K. and Ireland, U.S. and DACH. The mean difference of India and South Africa, 0.41 units, is found to be statistically significant at 5% significance level (p-value=0.01). Only the mean differences of ANZ and Ireland, and the U.S. and Canada are not found statistically significant (p-value>0.5), but these differences are only 0.13 and negative 0.03, respectively.

Differences in Risk Tolerance Over Time

Distribution of risk-tolerance score over time shows no practical difference. Exhibit 7 provides the boxplots of risk score per year and Exhibit 8 the corresponding descriptive statistics. We observe that in all years the score mean and median coincide and are between 48 and 50. There is no observable trend over time.

Exhibit 7 Distribution of Risk-Tolerance Scores by Year

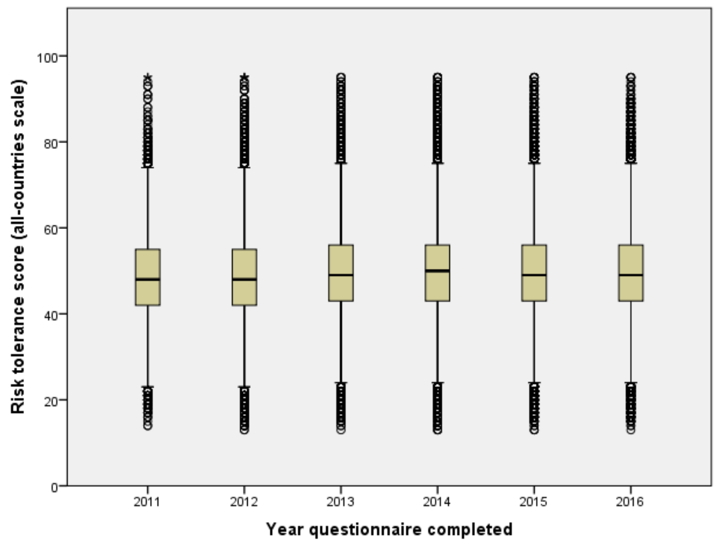


Exhibit 8 Descriptive Statistics for Distribution of Risk-Tolerance Scores by Year

Year	Data size	Mean	Median	Std. Dev.	Min	Max	Skewness		Kurtosis	
2011	19414	48.44	48	10.244	14	95	.171	(.018)	.313	(.035)
2012	70819	48.27	48	10.034	13	95	.131	(.009)	.282	(.018)
2013	85638	49.11	49	10.013	13	95	.131	(.008)	.330	(.017)
2014	89361	49.87	50	10.112	13	95	.114	(.008)	.408	(.016)
2015	88678	49.61	49	10.018	13	95	.107	(.008)	.382	(.016)
2016	51918	49.42	49	10.102	13	95	.130	(.011)	.402	(.021)

standard errors in parenthesis

Although Exhibit 9 reports the mean differences of adjacent years being statistically significant (p-value<0.001 except for 2011-12 where p-value=0.033, thus difference is significant at 5% significance level), they are extremely small, less than a unit, and subsequently of no practical importance. The distributions of all years are fairly symmetrical, with slightly heavier tails than a normal distribution (for example, a bit more extreme observations than expected under a normal distribution).

Exhibit 9 Mean Differences in Risk-Tolerance Scores Between Adjacent Years

Pairwise Comparisons	Means Difference	Std. Error	p-value
2011 vs 2012	.17	.082	.033
2012 vs 2013	-.84	.051	.000
2013 vs 2014	-.75	.048	.000
2014 vs 2015	.25	.048	.000
2015 vs 2016	.19	.056	.001

Differences in Risk Tolerance by Gender

The mean of males' scores is higher than females' scores by 5.68, and the difference is statistically significant (p-value<0.001). The result is based on the analysis of the cases for which gender is registered--69% of the Working Data Set.

Exhibit 10 Distribution of Risk-Tolerance Scores by Gender

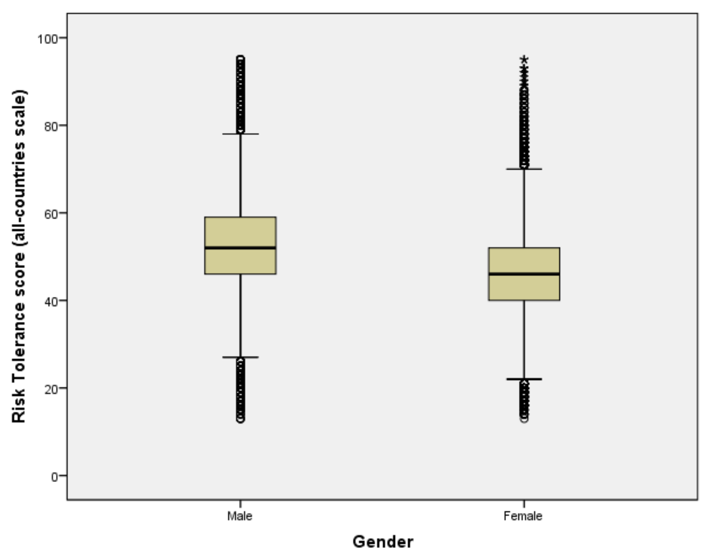


Exhibit 10 presents the boxplots of risk-tolerance scores by gender, and Exhibit 11 gives the descriptive statistics as well as the test of equal means. The mean score is 52.11 and 46.44 for men and women, respectively. The medians are almost identical to the means, and along with the skewness values, indicate that the distributions are symmetrical for both genders. The standard deviations are very close, and the kurtosis is a bit higher for women (slightly more extreme observations for women than men).

Exhibit 11 Descriptive Statistics for Distribution of Risk-Tolerance Scores by Gender

Gender	Data size	Mean	Median	Std. Dev.	Min	Max	Skewness		Kurtosis	
Male	158691	52.11	52	9.965	13	95	.074	(.006)	.379	(.012)
Female	121189	46.44	46	9.245	13	95	.086	(.007)	.434	(.014)

Difference in means = 5.68, p-value=0.000, standard errors in parenthesis

Differences in Risk Tolerance Across Age Groups

The score mean seems to increase as we move from age group 20-29 to group 30-39 but decreases as age increases beyond 39. A similar pattern is observed in score variance. The average risk-tolerance score of group 20-29 is about 7 points higher than the average score of group 70-80. The means for all age groups fall into the middle risk group 4 and the differences of the means between adjacent age groups are of a maximum 3 points. The differences in standard deviation are minimal as well, 1.3 at most. The results come from 67.9% of the Working Data for which age is reported. All these can be seen in Exhibits 12 and 13. Exhibit 14 shows that the difference in score means between adjacent age groups are statistically significant (p-value<0.001), but these differences are negative 0.67, 0.78, 2.87, 2.9, 1.51, respectively. Group 20-29 and 40-49 have the same mean score; the difference in mean is just 0.10 and is not found statistically significant (p-value=0.21).

Exhibit 12 Distribution of Risk-Tolerance Scores Across Age Groups

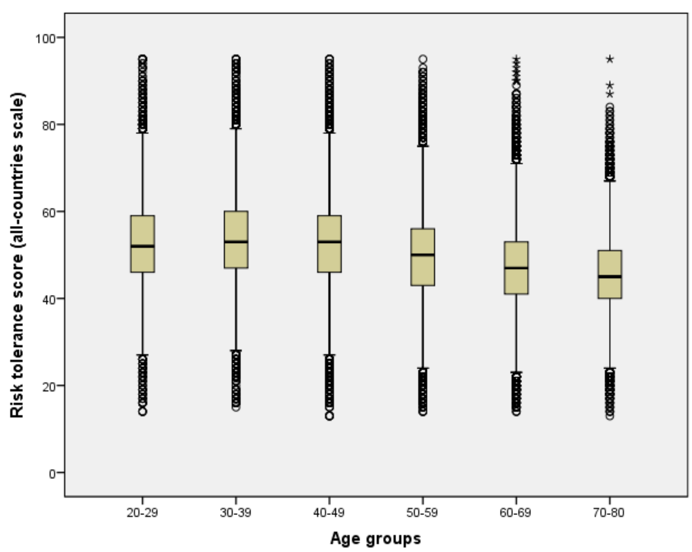


Exhibit 13 Descriptive Statistics for Risk-Tolerance Scores Across Age Groups

Age Groups	Data size	Mean	Median	Std. Dev.	Min	Max	Skewness		Kurtosis	
20-29	18104	52.69	52	10.049	14	95	.216	(.018)	.503	(.036)
30-39	33670	53.36	53	10.127	15	95	.138	(.013)	.312	(.027)
40-49	51391	52.58	53	9.983	13	95	.027	(.011)	.355	(.022)
50-59	72703	49.71	50	9.755	14	95	.025	(.009)	.309	(.018)
60-69	72139	46.82	47	9.296	14	95	.069	(.009)	.380	(.018)
70-80	27717	45.31	45	8.795	13	95	.077	(.015)	.433	(.029)

standard errors in parenthesis

Exhibit 14 Mean Differences in Risk-Tolerance Scores Between Adjacent Age Groups

Pairwise Comparisons	Mean Difference	Std. Error	p-value
20-29 vs 30-39	-0.67	.089	.000
30-39 vs 40-49	0.78	.068	.000
40-49 vs 50-59	2.87	.056	.000
50-59 vs 60-69	2.90	.051	.000
60-69 vs 70-80	1.51	.068	.000
20-29 vs 40-49	0.10	.083	.210

Differences in Risk Tolerance Across Education Levels

Risk tolerance is found to be positively associated with education level, like higher level of education is associated with higher level of risk tolerance. The increase in the mean score between adjacent education levels is not more than 2.6 and the difference in means of the highest and lowest education levels is 6.1. The means for all education levels fall into the average risk group 4. The result is based on 61.3% cases of the Working Data, which have reported their education level.

Exhibits 15, 16 and 17 below detail the analysis. The mean differences for the first two education levels (high school not completed versus completed) and for the last two (trade or diploma versus university degree or higher) seem to be similar--2.6 in absolute value--while the difference between the two middle levels (completed high school versus trade/diploma) is smaller, 0.93 in absolute value.

Exhibit 15 Distribution of Risk-Tolerance Scores Across Education Levels

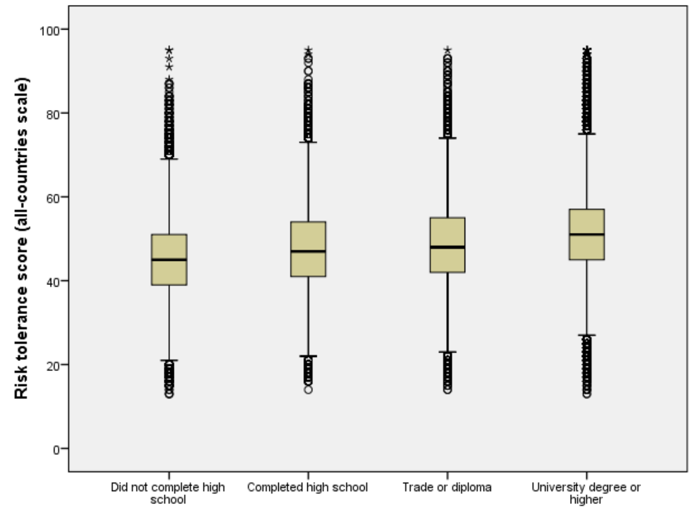


Exhibit 16 Descriptive Statistics for Risk-Tolerance Scores Across Education Levels

Education Level	Data size	Mean	Median	Std. Dev.	Min	Max	Skewness		Kurtosis	
High school not completed (1)	23512	45.14	45	10.124	13	95	.154	(.016)	.389	(.032)
High school completed (2)	28850	47.72	47	9.869	14	95	.189	(.014)	.360	(.029)
Trade/diploma (3)	46675	48.65	48	10.029	14	95	.171	(.011)	.310	(.023)
University degree or higher (4)	149539	51.24	51	9.761	13	95	.158	(.006)	.370	(.013)

standard errors in parenthesis

Exhibit 17 Mean Differences in Risk-Tolerance Scores Between Adjacent Education Levels

Pairwise comparisons	Mean Difference	Std. Error	p-value
Category 1 vs 2	-2.57	.087	.000
Category 2 vs 3	-0.93	.074	.000
Category 3 vs 4	-2.59	.052	.000

6. Construct Validity

In this section we studied the degree to which the 22 questions (11, 23, and 24 are not scored) measures risk tolerance using two approaches. First, we compute the associations[4] among the 22 questions. High levels of association indicate that there is a common factor in all questions. Second, factor analysis will be conducted to confirm whether the information conveyed by the 22 questions can be effectively summarized in one factor: risk tolerance. The Working Data set is used for both approaches. The main results of the section are summarized below.

- Among all pairwise associations of the 22 questions, half of them are moderate or strong.
- Questions 1-10, 12-22, and 25 measure one factor, which can be interpreted as risk tolerance.
- Questions 1, 9, 10, 14, 16, 20, and 25 are found to be the most reliable indicators of risk tolerance.[5]

Associations of the 25 Risk-Tolerance Questions

To consider the ordinal nature of scored questions, the Goodman and Kruskal's gamma coefficients are computed for all pairs of questions.[6] Their Pearson correlations are also reported to keep continuity with the previous technical report. The Goodman and Kruskal's gamma coefficient measures the strength of association of the cross-tabulated data of two ordinal variables. The coefficient ranges from negative 1 to positive 1, with zero indicating absence of association and 1 implying 100% positive association; higher values represent higher levels of risk tolerance. If the variables are reversely coded, a perfect agreement will be reflected by negative 1. Absolute values between 0 and 0.25 indicate lack of association, while between 0.7 and 1 denote strong association.

Exhibit 18 provides the value of Goodman and Kruskal's gamma coefficient for all pairs of the 22 questions measuring risk tolerance. If we summarize the values of gamma coefficient, the minimum gamma value is 0.17, and the median and the mean are 0.40 and 0.42, respectively. Therefore, 50% of the associations range from 0.4 to 0.83, suggesting moderate to strong associations. The pairs of questions that exhibit strong association (value larger than 0.7) are question 1 with questions 3, 10, and 16; question 16 with questions 10 and 20; and question 10 with question 205.

Exhibit 19 gives the Pearson correlations for all pairs of questions. The minimum value is negative 0.14, the maximum is 0.70, the mean is 0.25, and the median is 0.28. The values need to be interpreted with caution as Pearson correlation assumes that both variables are continuous. However, the qualitative results remain roughly the same as the ones presented above for the gamma coefficient. Fifty percent of the correlations are moderate or strong. As above, there is almost a lack of correlation for questions 23, 24, and 11, and the correlations of question 1 with questions 10 and 16, and of question 16 with questions 10 and 20 are the strongest (between 0.6 and 0.7).

Exhibit 18 Goodman and Kruskal's Gamma Coefficient for 25 Risk-Tolerance Questions

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22
Q2	0.58																				
Q3	0.72	0.55																			
Q4	0.53	0.32	0.49																		
Q5	0.50	0.35	0.44	0.37																	
Q6	0.60	0.46	0.60	0.41	0.43																
Q7	0.47	0.48	0.46	0.23	0.31	0.50															
Q8	0.49	0.33	0.46	0.44	0.52	0.40	0.34														
Q9	0.69	0.41	0.54	0.58	0.40	0.47	0.40	0.44													
Q10	0.83	0.57	0.70	0.48	0.49	0.60	0.47	0.50	0.68												
Q12	0.44	0.42	0.39	0.32	0.30	0.34	0.56	0.32	0.43	0.43											
Q13	0.47	0.37	0.44	0.32	0.34	0.38	0.36	0.35	0.36	0.50	0.31										
Q14	0.55	0.48	0.49	0.35	0.33	0.42	0.37	0.36	0.40	0.61	0.33	0.41									
Q15	0.34	0.26	0.32	0.26	0.31	0.29	0.22	0.31	0.21	0.39	0.22	0.31	0.29								
Q16	0.72	0.49	0.58	0.40	0.42	0.50	0.41	0.41	0.52	0.78	0.40	0.44	0.58	0.34							
Q17	0.57	0.41	0.51	0.36	0.35	0.45	0.34	0.37	0.41	0.66	0.29	0.41	0.54	0.33	0.58						
Q18	0.47	0.35	0.39	0.24	0.33	0.38	0.38	0.30	0.38	0.50	0.32	0.37	0.38	0.24	0.49	0.39					
Q19	0.55	0.39	0.46	0.28	0.34	0.40	0.31	0.29	0.36	0.64	0.29	0.35	0.43	0.28	0.57	0.46	0.39				
Q20	0.62	0.41	0.50	0.36	0.34	0.42	0.34	0.35	0.44	0.71	0.33	0.37	0.51	0.28	0.73	0.55	0.42	0.50			
Q21	0.44	0.29	0.36	0.26	0.29	0.31	0.33	0.27	0.38	0.47	0.32	0.30	0.31	0.17	0.45	0.32	0.40	0.33	0.39		
Q22	0.53	0.37	0.46	0.37	0.38	0.43	0.34	0.38	0.41	0.58	0.31	0.40	0.44	0.35	0.52	0.54	0.42	0.40	0.48	0.36	
Q25	0.68	0.44	0.54	0.38	0.37	0.45	0.37	0.37	0.49	0.65	0.37	0.36	0.43	0.27	0.58	0.45	0.38	0.44	0.51	0.34	0.42

Exhibit 19 Pearson Correlations for 25 Risk-Tolerance Question

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22
Q2	0.44																				
Q3	0.46	0.33																			
Q4	0.30	0.17	0.25																		
Q5	0.41	0.28	0.30	0.23																	
Q6	0.44	0.32	0.36	0.21	0.33																
Q7	0.30	0.28	0.23	0.10	0.21	0.30															
Q8	0.38	0.25	0.29	0.25	0.45	0.29	0.21														
Q9	0.54	0.30	0.33	0.32	0.32	0.34	0.24	0.35													
Q10	0.65	0.40	0.42	0.26	0.38	0.42	0.28	0.37	0.51												
Q12	0.32	0.29	0.23	0.17	0.23	0.23	0.33	0.24	0.31	0.29											
Q13	0.39	0.29	0.29	0.18	0.30	0.29	0.22	0.28	0.28	0.38	0.23										
Q14	0.46	0.37	0.33	0.22	0.29	0.33	0.25	0.30	0.33	0.48	0.26	0.36									
Q15	0.26	0.19	0.19	0.14	0.26	0.20	0.13	0.23	0.16	0.27	0.15	0.25	0.24								
Q16	0.61	0.39	0.40	0.26	0.38	0.40	0.28	0.35	0.43	0.63	0.32	0.39	0.52	0.28							
Q17	0.42	0.28	0.30	0.19	0.27	0.30	0.19	0.27	0.29	0.46	0.20	0.31	0.42	0.23	0.45						
Q18	0.37	0.25	0.24	0.13	0.27	0.27	0.23	0.23	0.28	0.36	0.22	0.30	0.32	0.18	0.41	0.29					
Q19	0.44	0.28	0.29	0.18	0.28	0.29	0.20	0.23	0.30	0.48	0.21	0.28	0.36	0.21	0.49	0.34	0.30				
Q20	0.56	0.36	0.38	0.26	0.34	0.36	0.25	0.33	0.41	0.60	0.30	0.35	0.49	0.26	0.70	0.45	0.38	0.46			
Q21	0.38	0.23	0.24	0.16	0.27	0.25	0.22	0.24	0.32	0.38	0.25	0.27	0.29	0.14	0.42	0.25	0.33	0.28	0.40		
Q22	0.38	0.25	0.27	0.19	0.29	0.28	0.19	0.27	0.29	0.39	0.21	0.29	0.34	0.24	0.40	0.36	0.29	0.29	0.40	0.28	
Q25	0.58	0.37	0.39	0.25	0.36	0.38	0.27	0.34	0.43	0.54	0.31	0.34	0.41	0.24	0.56	0.38	0.34	0.40	0.54	0.35	0.35

Factor Analysis

Two techniques will be used for factor analysis, principal component analysis, and a one-factor latent variable model will be fitted to the data. The first approach is descriptive; there is no formal statistical test to test whether one or more factors are needed to summarize the data, and it assumes that all questions are continuous variables. The second approach assumes a statistical model so that we are able to conduct a hypothesis, testing whether one factor is enough to summarise the information of the 22 questions, and it treats the questions as ordinal variables except for the very last one, question 25, which is treated as continuous. Also, the second approach acknowledges that the questions are not perfect measurement instruments of risk tolerance, but they measure the factor with some error. We present the first technique because it is commonly used and has been used in the previous technical report.

Principal Component Analysis

Exhibit 20 presents the results of principal component analysis for the first five components with the highest percentage of data variance explained. The total number of components is 22. The first component explains 36% of the total variance of the data while the rest of the components, when added, explains an additional 5% of the data-variance maximum. This may be interpreted that one component is enough to summarize effectively the information of questions 1-10, 12-22, and 25.

Exhibit 20 Principal Component Analysis for First Five Components

Component	Eigenvalue	Data Variance Component Explained	Cumulative Data Variance Explained
1	7.895	35.888%	35.888%
2	1.111	5.049%	40.937%
3	1.050	4.772%	45.709%
4	0.979	4.449%	50.158%
5	0.889	4.041%	54.199%

Exhibit 21 gives the correlation of each question with the first component. Nearly all the correlations are fairly strong and larger than 0.5. This indicates that all questions measure the component in an adequate way. Since all correlations are positive, the first component can be interpreted as a measure of risk tolerance. The largest correlations are between 0.75 and 0.8 for questions 16, 1, 10, 20. This is not a surprise since these questions exhibit the highest gamma coefficients (see previous subsection).

Exhibit 21 Correlation of Risk-Tolerance Questions with First Component

Q16	.792	Q9	.630	Q2	.551	Q12	.471
Q1	.791	Q19	.591	Q13	.549	Q7	.430
Q10	.784	Q17	.586	Q22	.549	Q4	.403
Q20	.757	Q6	.572	Q8	.543	Q15	.384
Q25	.705	Q3	.569	Q21	.533		
Q14	.645	Q5	.563	Q18	.531		

One-Factor Latent Variable Model

A one-factor model is fitted to the data where questions 1-10 and 12-22 are treated as ordinal variables and question 25 as continuous. The device that enables us to treat ordinal variables as such is the assumption that for each ordinal variable there is an underlying continuous variable that we partly observe through its ordinal counterpart. Indeed, it is not unreasonable to assume that a Likert scale comes from a continuous scale that has been chopped into categories.

The fit indexes CFI and RMSEA for the one-factor model are 0.991 and 0.048, respectively, suggesting that the model has a good fit to the data.[7] In other words, one factor summarizes the 22 questions effectively.

Exhibit 22 presents the correlation of the underlying continuous variables with the factor of the model. They roughly indicate the associations of the questions with the factor. All questions have moderate-to-strong positive associations with the factor. Hence the factor can be interpreted as a measure of risk tolerance. As before, questions 1, 9, 10, 14, 16, 20, 25 exhibit the highest associations with the factor, while questions 4, 7, 12, 13, 15, 21 exhibit the lowest associations with the factor. Questions with higher associations are considered more reliable indicators of risk tolerance, as they measure the construct with smaller error.

Exhibit 22 Correlation of Risk-Tolerance Questions with Factor in One-Factor Latent Variable Model

Q10	0.855	Q9	0.647	Q2	0.575	Q4	0.495
Q1	0.837	Q3	0.635	Q5	0.566	Q7	0.476
Q16	0.832	Q17	0.622	Q18	0.544	Q12	0.473
Q20	0.788	Q19	0.608	Q8	0.544	Q15	0.411
Q25	0.693	Q6	0.605	Q21	0.529		
Q14	0.651	Q22	0.585	Q13	0.520		

7. Reliability Analysis

Two types of reliability are studied: the internal-consistency reliability (also known as equivalence reliability) and the test-retest reliability (also known as stability reliability). The internal consistency is connected to the inter-relatedness of the questions and provides an assessment of the amount of measurement error in the test. The test-retest reliability is used to assess the degree to which the questionnaire consistently measures risk tolerance over time. The main conclusions of this section are that the questionnaire

- exhibits high level of internal-consistency reliability, and
- consistently measures risk tolerance.

Internal-Consistency Reliability

There are many reliability measures proposed in the literature, with each one having pros and cons. However, if they all point toward the same direction, the conclusion can be considered reliable. For this analysis, we consider Cronbach's alpha, Guttman's lambda coefficients, split-half coefficients, Spearman-Brown coefficient, and Cronbach's alpha when one question at a time is deleted from the set. To compute all these coefficients, we use the standardized[8] form of question 25 because its standard deviation is much larger (14.7) compared with the standard deviations of the rest of the items (0.6-2).

Exhibit 23 provides Cronbach's alpha and Guttman's lambda coefficients, all suggesting high level of reliability. Split-half approach can be viewed as a one-test equivalent to test-retest reliability. The split-half coefficients in Exhibit 24 refer to the case where the questions are split into two parts: part one includes the first 10 questions plus question 12, and part two includes the rest of the questions. The Cronbach's alpha for each part remains high. The high correlation of the two parts (0.774) also indicates high level of reliability. Spearman-Brown coefficient and Guttman split-half coefficient point to the same direction. Note that the grouping of the questions into two sets determines the results. However, the Cronbach's alpha given in Exhibit 24 is equivalent to the average of all possible split-half Cronbach's alpha estimates.

Exhibit 23 Reliability Coefficients

Cronbach's Alpha	.902
Guttman's lambda 1	.861
Guttman's lambda 2	.913
Guttman's lambda 3	.902
Guttman's lambda 4	.842
Guttman's lambda 5	.912
Guttman's lambda 6	.916

Exhibit 24 Split-Half Reliability Coefficients

	Part 1	Part 2
	Questions 1-10, 12	Questions 13-22, 25
Cronbach's alpha	.830	.838
Correlation Between Forms	.774	
Spearman-Brown Coefficient	.873	
Guttman Split-Half Coefficient	.842	

Exhibit 25 gives Cronbach's alpha after a question is deleted from the initial set of the 22 questions. Since the coefficients are quite consistent, they all contribute similarly to the overall reliability.

Exhibit 25 Split-Half Reliability Coefficients

Q1	.892	Q7	.901	Q14	.896	Q20	.899
Q2	.898	Q8	.898	Q15	.901	Q21	.900
Q3	.899	Q9	.896	Q16	.891	Q22	.899
Q4	.901	Q10	.894	Q17	.898	Q25*	.894
Q5	.898	Q12	.900	Q18	.898	*unstandardised	
Q6	.898	Q13	.898	Q19	.897		

Test-Retest Reliability

To carry out the test-retest reliability analysis, we focus on the cases where the test has been taken twice,[9] and the second time has been marked as a retest. There are 9,902 such cases that have not been included in the working data set we have been analyzing above. This gives more value to the test-retest analysis, as the test-retest data can be considered as independent to the Working Data set.

The essence of the analysis is to study the correlation of the risk-tolerance scores[10] measured at the two different time points. If the scores do not change substantively--for example, they are highly correlated--it indicates that the questionnaire consistently measures risk tolerance. The shorter the interval, the higher the correlation, but this may be due to memory effects. The longer the interval, the lower the correlation since risk tolerance is not set in concrete and certain life events may have an effect. There are no rules about the best time lag; it depends on the specific concept measured.

Exhibit 26 presents the percentage of cases with different time lags between the two assessments of risk tolerance. The time lag is measured in years and ranges from 0 (for example, less than a year) and 5.

Exhibit 26 Percentage of Cases With Different Time Lags Between Assessments

Time Between Assessments (Years)	Cases
0	19.4%
1	19.7%
2	25.3%
3	25.4%
4	8.5%
5	1.6

Exhibit 27 gives the correlation of the scores for a given time lag (for example, the data is split by time lag) and for the complete data set. The correlation for the complete data set is 0.77, fairly high, indicating high level of the test-retest reliability, which is consistent with similar studies.[11] As expected, the correlation is higher when the time lag is small—such as, 0.81 for a one-year interval—and slightly decreases as the time interval increases (for example, 0.68 for a five-year lag, which is still a strong correlation). It is interesting to note that for the cases with less than a year interval, the correlation is not the highest. A qualitative analysis of these cases may cast light as to what has happened at the time of taking the test.

Exhibit 27 Correlation of Risk-Tolerance Scores Measured at Two Time Points

Time Between Assessments (Years)	Correlation of the Scores
0	0.71
1	0.81
2	0.79
3	0.78
4	0.77
5	0.68
0-5 (all cases)	0.77

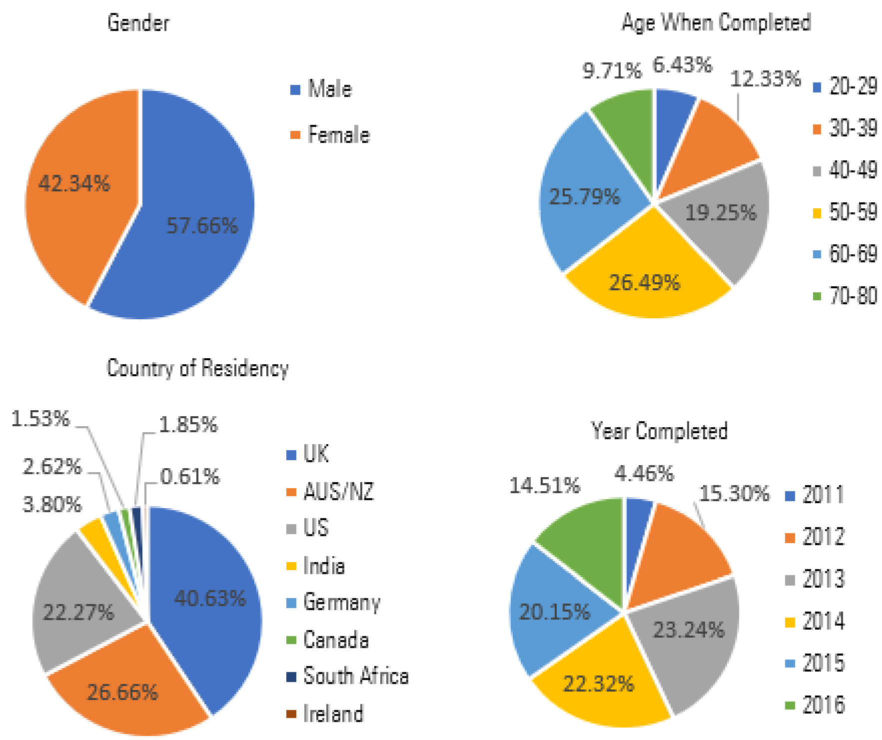
8. Risk-Tolerance Scales and Scoring Algorithm

Having found in Section 5 that risk-tolerance score is associated with gender, age, and education, the norm group consists only of cases for which the aforementioned demographics are registered (241,470 cases out of 405,828 in the Working Data set). The demographic profile of the norm group is reported in Exhibit 28. It is actually very similar to that of the complete data set presented in Section 2.

In Section 5, we have also found that the differences in the distribution of risk-tolerance scores across countries are of no practical importance in that the implications to investment decisions are minimal. Thus, a common norm group and scale for all countries (world) is fine from a practical perspective. The Working Data set covers U.K., Australia/New Zealand, U.S., Canada, DACH, Ireland, India, and South Africa for the period September 2011 to July 2016. Data from the whole period can be used since there is no difference in the distribution of risk tolerance over time (see Section 5).

A weighted total score where different weights are given to the questions depending on their reliability to measure risk tolerance does not differ substantively from the total score (where all questions are given the same weight).

Exhibit 28 Demographic Profile of the Norm Group



Global Norm Group and Scoring Algorithm--Equal Weighting

The world norm group is based on the data presented in the beginning of this section for U.K., Australia/New Zealand, the U.S., Canada, DACH, Ireland, India, and South Africa for the period September 2011 to July 2016 for which gender, age, income, and education have been registered.

The total score is computed following the steps detailed in Exhibit 29.

Exhibit 29 Scoring Algorithm Using Questions 1-10, 12-22, and 25

1. The mean and standard deviation for questions 1-10, 12-22, and 25 are computed from the norm group data to be used in step 2.
2. The z-scores for all questions and all individuals in the norm group are computed, where z-score in question A of individual X = (answer of individual X to question A – mean of question A) / standard deviation of question A.
3. For all individuals in the norm group, the sum of their z-scores is computed as well as the standard deviation of the sums of the z-scores. The latter is used in step 4.
4. For all individuals, the z-score of their sum of z-scores is computed by dividing their sum of z-scores with the standard deviation of the sums calculated in step 3. Let us refer to the z-score of the sum of the z-scores as the total z-score.
5. For all individuals in the norm group, the total z-score is multiplied by 10 and then 50 is added. This is done so that the total scores have mean 50 and standard deviation 10. Let us refer to the transformed total z-score as the total score.

The distribution of the world score is very close to a normal distribution with mean 50 and standard deviation. Exhibit 30 displays the histogram of the total. The mean and standard deviation are indeed 50 and 10, respectively, as they should by construction. The solid line in the histogram denotes a normal distribution with mean 50 and standard deviation 10. We see that the shape of the histogram is very close to the graph of the normal distribution. The skewness of the histogram is 0.115, indicating that it is almost symmetrical with a slightly longer tail to the right. The kurtosis is 0.322, suggesting slightly more values to the ends of the risk-tolerance scale compared with a normal distribution. Exhibit 31 compares the score values at different percentiles[12] of the histogram with the score values at the same percentiles of a normal distribution with mean 50 and standard deviation 10. The differences between the values are minimal, all less than 0.7.

Exhibit 30 Histogram of the World Scores

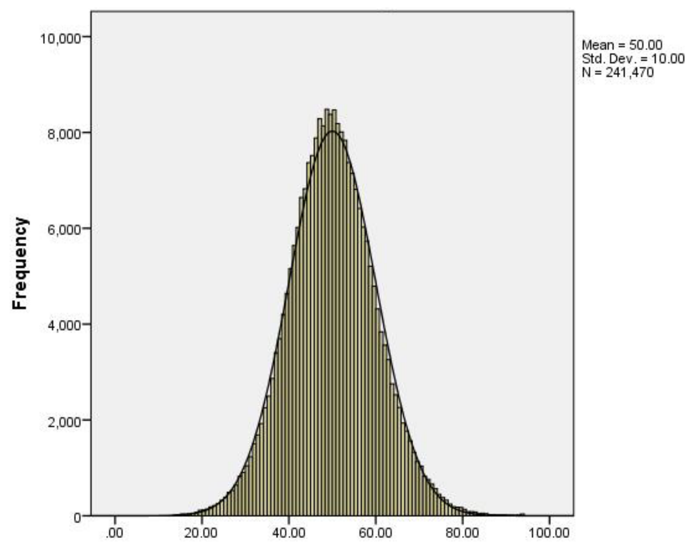


Exhibit 31 Empirical Percentiles of World Distribution Against the Normal Distribution

Percentile	Score Value Based on World Score Distribution	Score Value Based on Normal Distribution (mean=50 and SD=10)
10	37.56	37.18
20	41.83	41.58
30	44.83	44.76
40	47.39	47.47
50	49.81	50.00
60	52.24	52.53
70	54.92	55.24
80	58.11	58.42
90	62.72	62.82

The range of scores defining each risk group remains the same and are detailed on page 5. The mean, median, and standard deviation of the world scores for each risk group are provided in Exhibit 32.

Exhibit 32 Mean, Median, and Standard Deviation of Risk-Tolerance Scores by Risk Group

Group	Mean	Median	Std. Deviation
1	21.73	22.38	2.671
2	31.48	32.07	2.631
3	40.91	41.29	2.741
4	49.94	49.92	2.823
5	59.12	58.74	2.764
6	68.66	68.10	2.670
7	79.44	78.11	4.266

Global Scoring Algorithm--Different Weightings

In Section 6 we have found that some questions are more reliable indicators of risk tolerance than others. Here we analyze how much the world scores would change if we assigned larger weight to the more reliable questions. The weights for each question obtained by principal component analysis are given in Exhibit 33, and they are sorted from the largest to the smallest one in Exhibit 34. In Exhibit 34 we see that the order of the questions is exactly the same as in Exhibit 25, which presents the correlation of each question with the factor risk tolerance. The weights range from 0.099 to 0.031 and their values are such that the computed weighted total score will have 0 mean and standard deviation 1.

Exhibit 33 Question Weights by Principal Component Analysis

Q1	.100	Q7	.054	Q14	.082	Q20	.096
Q2	.070	Q8	.069	Q15	.049	Q21	.068
Q3	.072	Q9	.080	Q16	.100	Q22	.070
Q4	.051	Q10	.099	Q17	.074	Q25	.089
Q5	.071	Q12	.060	Q18	.067		
Q6	.072	Q13	.070	Q19	.075		

Exhibit 34 Question Weights by Principal Component Analysis Sorted from Largest to Smallest

Q1	0.100	Q9	0.080	Q2	0.070	Q12	0.060
Q16	0.100	Q19	0.075	Q22	0.070	Q7	0.054
Q10	0.099	Q17	0.074	Q13	0.070	Q4	0.051
Q20	0.096	Q6	0.072	Q8	0.069	Q15	0.049
Q25	0.089	Q3	0.072	Q21	0.068		
Q14	0.082	Q5	0.071	Q18	0.067		

The weighted world score is computed as follows:

1. Having computed the z-scores for all 22 scored questions in step 2 of the scoring algorithm (see Exhibit 29), we calculated the weighted sum of the z-scores. This is equal to w1 * z-score of question 1 + w2 * z-score of question 2 + ... + w22 * z-score of question 22 + w23 * z-score of question 25, where w1, w2, ..., w22, w25 are the weights of the question given in Exhibit 34. As already said, the weighted sum of the z-scores has mean 0 and standard deviation 1.
2. The weighted sum of the z-scores is multiplied by 10 and 50 is added to the product in order to obtain the weighted total score with mean 50 and standard deviation 10.

Comparing the equal weighting world score computed in the previous subsection with the weighted world score, the mean and median difference between the two are 0 and negative 0.0065, respectively. This implies that, on average, there is no difference between the two scores. The equal weighting world score is preferred since it is simpler to calculate and conveys the same information as the weighted total score.

9. Comparing the 10-Question Score With the 25-Question Score

This section confirms that the score based on the 10 questions (10Q score) is very close to the 22-question total score presented in the previous section. The 10Q score is within five units from the 22-question total score for 90.6% of the cases of the world norm group, while 76.7% of the cases are assigned to the same risk group by both scores (total and 10Q score).

We compute the 10Q score following the same steps as for the total score detailed in Exhibit 35. The difference is that, in step 3, we use only the z-scores of the 10 aforementioned questions to compute a sum of z-scores and subsequently, the standard deviation of the latter sum needs to be computed for step 4. Second, we calculate the differences between the 10Q scores and the total scores. The mean and median of the differences are 0 and negative 0.06, respectively, implying that, on average, there is no difference between the two scores. The standard deviation of the differences is 2.99 and the minimum and maximum values are negative 15.48 and 17.87, respectively. However, such extreme values occur in less than 2% of the cases of the world norm group. Exhibit 35 shows that the 10Q scores and the total scores are different at most by five for 90.6% of the cases and at most by seven for 97.8% of the cases.

Exhibit 35 Difference Between 10Q Scores and Total Scores

Difference Between 10Q Scores and Total Scores	Cases of World Norm Group
Within 5 units	90.6%
Within 7 units	97.8%

We also compared the risk group a respondent is assigned to based on their total score and the 10Q score using the score ranges of the risk groups presented on page 5. Exhibit 36 reports that 76.7% of the cases of the world norm group are assigned to the same risk group regardless of which score is used.

The rest of the cases are assigned to the adjacent risk groups--half of them the 10Q score is more conservative and half is more aggressive.

Exhibit 36 Difference Between 10Q Scores and Total Scores by Risk Group

Risk Group Assigned by 10Q Score Compared With Total Score	Cases of World Norm Group
Previous less-risky group	11.5%
Same risk group	76.7%
Next riskier group	11.7%

Exhibit 37 provides Cronbach's alpha and Guttman's lambda coefficients for the 10 questions. All suggest a fairly high level of reliability for the shorter version of the questionnaire.

Exhibit 37 Reliability Coefficients of the 10Q Score

Cronbach's alpha	.841
Guttman's lambda 1	.757
Guttman's lambda 2	.850
Guttman's lambda 3	.841
Guttman's lambda 4	.800
Guttman's lambda 5	.839
Guttman's lambda 6	.842

22 West Washington Street
Chicago, IL 60602 USA

1 http://www.riskprofiling.com/Downloads/Questionnaire_AUSNZv3.pdf, this is the Australian version of the questionnaire. Other versions are identical except for terminology differences, such as, stocks vs. shares.
2 The variance (which is the square root of the standard deviation) of the sum of the z-scores is equal to the sum of the variances of the 22 questions' z-scores plus the covariance of all pairs of the 22 z-scores.
3 A boxplot displays the distribution of a variable where the median is reported by the solid line inside the box and the lines on the bottom and top of the box show the value range of the middle 75% of the data. The circles represent extreme values. Very roughly a boxplot can be seen as a "flat" histogram.
4 The term "association" is preferable to "correlation" because 22 out of the 25 questions are ordinal variables--they are measured on a Likert scale where the number of response categories ranges from 2 to 11.
5 This result is in line with the result of a previous analysis looking into a possible three-question version.
6 An ordinal variable is a categorical variable for which the possible responses can be ordered. The responses are usually denoted by arbitrary numeric values which need to reflect the ordinality of the possible responses, but otherwise their absolute value can be any number (such as, 1,2,3, or negative 2, negative 1,0,1). One cannot perform arithmetic operations on ordinal variables, such as, category 2 does not have double the characteristic of category 1.
7 CFI values range from 0 to 1 with 1 denoting very good fit. RMSEA is a positive number with 0-0.05 indicating good fit, 0.05-0.1 moderate fit, and larger than 1 poor fit of the model.
8 A variable is standardized when each observed value of the variable is transformed as follows: (observed value – variable mean)/variable standard deviation. Often the standardized values are referred to as z-scores.

9 There are 12,603 cases with distinct Client Codes that have been marked as they have retaken the test. However, a 12.6% of these cases have actually taken the test only once and 8.9% more than twice (from 3 to 12 times). For the latter cases, the extra issue of which measurements to select arises, so we decided to omit them from the analysis.
10 The variable labelled as "Score ALL" in the data set is used.
11 http://riskprofiling.com/Downloads/intertemporalRTscores.pdf
12 A percentile indicates the value below which a given percentage of observations falls. For example, if the 10% percentile is 37, this means that 10% of the scores are below 37. The value of α% percentile may differ between the empirical distribution (for example, the histogram) of the scores and an assumed theoretical distribution (for example, a normal distribution with mean 50 and standard deviation 10).

Thank you to Dr. Myrsini Katsikatsou, research fellow at the London School of Economics, Department of Statistics, and Stuart Erskine, commercial economist, for their assistance with this analysis. For more information on any of content and methodology outlined in this document, contact:

Nicki Potts
Director of Product Management
Financial Planning,
Advisor Client Experience (ACE)