SOA Research Institute

Actuarial Innovation and Technology

Redesigning the Life Insurance Underwriting Journey with Behavioral Science

May | 2024

Redesigning the Life Insurance Underwriting Journey with Behavioral Science

Experimental Evidence showing the Impact of Underwriting Questionnaire Design on Disclosures

AUTHORS

Aisling Bradfield

Tiernan Caulfield

Denis Charles

Pamela Granzin

Michelle Lerch

Jenny McDonnell

Katherine McLaughlin

Davy Moore

Caitlyn Parsons

Robyn Wallner

SPONSORS

Actuarial Innovation and Technology Strategic Research Program Steering Committee

SOA Product Development Section

SOA Reinsurance Section

SOA Smaller Insurance Company Section

CONTENTS

Redesigning the Life Insurance Underwriting Journey with Behavioral Science

Experimental Evidence Showing the Impact of Underwriting Questionnaire Design on Disclosures

Executive Summary

This study demonstrates that applying the principles of behavioral science to the design of underwriting questions improves disclosure across a range of conditions and behaviors that are relevant to underwriting decisions. In the context of the shift towards accelerated underwriting in the insurance industry, behavioral science question design offers a way to collect better information through the application form.

A key element of the re-design of questions by applying behavioral science principles is to focus on the applicant perspective. However, we note that it is important to keep in mind the perspective of those who will be analyzing the information and using the application form answers to make an underwriting assessment of the risk. Considering both perspectives strives to achieve the optimum design that increases disclosure, improves customer experience, and maintains or improves ease of processing for the insurance company.

The industry survey shows that about a third of organizations want to learn more about applying behavioral science to life insurance underwriting question re-design. The two biggest difficulties identified with using behavioral science are 'Not enough expertise in the organization' and 'Not enough concrete results,' and this study addresses both concerns by providing a framework for applying behavioral science to underwriting question design and evidence of the impact on disclosure.

The study employed a randomized control trial (RCT) or A/B testing framework to test the impact on disclosure of different wording and design of the questions typically asked on life insurance application forms. Participants recruited from the U.S. population were randomly assigned to one of five experimental groups, each facing a different health questionnaire design. The outcome variable of interest was the rate of disclosure of various conditions and lifestyle choices.

In the experiment, many behavioral science principles were investigated and, for each application form question, two different question designs were tested. For each question, the disclosure rates were compared across the five experimental groups to identify the features of question design that most impact disclosure for each question topic. Full results by question are available in the interactive dashboard that supports this report.

In this report, we discuss the key themes that have emerged in terms of approaches to question design that are particularly effective at improving disclosure:

1. Separating conditions or behaviors of different severity improves disclosure. This was demonstrated for mental health conditions, tobacco products, and substance use support questions where allowing applicants to admit to only the relevant part of the question, via separate checkboxes, significantly increased disclosure rates.

2. Removing Yes/No questions that highlight negative behaviors and framing the question to assume the behavior exists was shown to improve disclosure. This often involves removing a Yes/No question and looking to the first drill-down question on the topic which, with the inclusion of a 'None' or 'Never' option, could be asked as the initial question. Evidence of a significantly increased disclosure of tobacco use and alcohol use is observed when questions are framed to assume the behavior exists.

3. Implementing social norms through anchoring shows promise for improving disclosure. Open-response boxes are difficult for applicants to answer. The experiment tested the use of closed-option responses and the use of slider bars, which provide anchors that allow applicants to interpret what the social norms or acceptable answers to the question might be. Closed-option responses increased disclosure compared to open response style for alcohol consumption, but a slider bar with a higher anchor was shown to be even more successful at increasing disclosure of high-risk alcohol consumption. Learnings from the experiment could inform the design of closed-option response style questions and slider bars to optimize disclosure impacts.

4. Taking the applicant perspective, it is possible to design questions that resonate better with applicants and capture additional information as a result. Traditional weight change questions ask about gain or loss, but what about those whose weight fluctuates? The experiment tested the inclusion of the option 'My weight went up and down,' and this was selected by 20-24% of respondents, highlighting that this option resonates with a significant proportion of respondents (at least 1 in 5). Having relevant options for frequency of use for alcohol and tobacco consumption was shown to increase disclosure of high-risk behaviors. A tobacco question that appears burdensome to answer was shown to be less effective at eliciting disclosure than a series of reflexive questions that appear based on the prior response, highlighting that reflexive style questioning works better than all-at-once tables of required information.

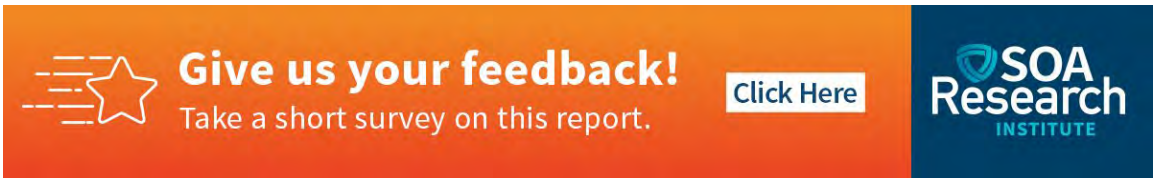
5. Long lists of medical conditions are often subject to skimming, and the experiment provided evidence that those who have had conditions, such as diabetes, cancer and heart disease, didn't always select these from the list of medical conditions asked. The study found that a confirmation question, re-asking about these three conditions, significantly increased disclosure of each condition. The experiment also showed that medical condition disclosures can be collected through experience-style questions asking about surgery, medication, and hospital stays. Asking about experiences in this way can be considered to reduce the number of medical conditions that need to be specified in lists.

6. The inclusion of an honesty pledge at the start of the questionnaire was shown to increase disclosure. The honesty pledge wording was designed using behavioral science techniques. The study tested the placement of the honesty pledge and showed that when an honesty pledge was seen at the end of the application form, it had limited impact, with very few participants going back to change answers at that stage. In the questionnaire with the honesty pledge placed at the start, statistically significant improvements in disclosure were observed for specific conditions – mental health, cancer, and high-risk BMIs. A behaviorally-designed honesty pledge placed at the start of the questionnaire improves disclosure on topics that are most impacted by intentional non-disclosure.

These findings provide evidence that behavioral science techniques in question design can improve disclosure.

In addition to the promising results, the experimental framework applied in this study provides a tool for insurance companies to quantify the impact on disclosure of different application form designs. The study aims to motivate companies to implement behavioral science principles in their application process and to monitor disclosures to

quantify the impact. This is an important next step to test the approaches that have shown strong impacts on disclosure in this experiment in the real-world setting of the life insurance application process.



Section 1: Introduction

1.1 BACKGROUND

1.1.1 BEHAVIORAL SCIENCE

Behavioral science is a branch of social science which incorporates insights from psychology, economics, and other subject areas to explain how a person's decision-making or behaviors can be affected by psychological, cognitive, emotional, cultural, and social factors. Behavioral economics is a subset of behavioral science concerned specifically with human decision-making. Both terms are widely used. For the purpose of this study, we discuss behavioral science, which encompasses behavioral economics.

Behavioral science does not assume that humans are perfectly rational actors who always make careful, calculated decisions. Instead, it recognizes that humans are fallible, with limited capacities for cognitive processing, self-control, and attention. Behavioral science studies the ways in which humans can make mistakes in a complex and demanding world. It also investigates the strategies that humans rely on to survive in this world and avoid making such mistakes.

Far from being a purely academic study, behavioral science sheds light on human decision-making in a variety of real-world domains. Why do many individuals fail to save for retirement? How do humans organize their household budgets? What is the best way to promote healthy eating choices? All of these questions and more are the purview of behavioral science.

If behavioral science is the study of human decision-making, it should be no surprise that it is relevant to many questions in the insurance industry. After all, at the heart of the entire concept of insurance are decisions about which risks to take, what protections to use against those risks, and how much to pay for those protections.

Behavioral science can be applied to all stages of the insurance journey. When insurance products are being developed and marketed, behavioral science can help insurers understand the biases that affect consumers' demand for various products. As consumers are purchasing insurance, behavioral science can inform the design of tools to help them choose the products that are right for them. Later on, continued engagement with policyholders can be fostered through behaviorally-designed programs, and the claims process can be optimized with behavioral science.

However, this study focuses on the application stage of the life insurance journey – the process faced by consumers as they apply for their chosen insurance plans. Specifically, we investigate how to use behavioral science to improve this process and produce better outcomes for applicants, underwriters, and the industry as a whole.

1.1.2 DISCLOSURE

As part of the process of applying for life insurance, the first step usually involves the applicant filling out an application form, sometime known as a questionnaire. This questionnaire asks about the applicant's medical history – diagnoses of medical conditions and previous medical experiences. It also asks questions about the applicant's lifestyle – their alcohol and tobacco consumption, for instance.

The insurance company to which the applicant is applying then uses the answers provided in the application to determine the applicant's level of risk and appropriately price the insurance contract offered to the applicant.

Traditionally, this questionnaire was often part of a bigger application process that might also include medical exams, blood tests, and interviews. However, in recent years, the insurance industry has been experiencing a trend towards fully digital or automated application processes, which can be completed entirely by the applicant, alone, online. This means that the information the company can gather about the applicant and their lifestyle from the application form is even more important because that may be the only information they can use to price the applicant's risk.

Previous studies (Lyons, 2022) have shown that approximately 9% to 23% of life and health insurance applications submitted contain some amount of non-disclosure, with variation by type of benefit. This is when the applicant either gives false answers to the questions on the form or leaves out information that is asked about on the form.

Anything less than full disclosure on an application form is a problem because full disclosure means that both parties entering into the insurance contract have the same information about the risks and the benefits involved. It means that the contract is valid and fair for everyone.

This valid contract means that, later on, when the policyholder makes a claim on their insurance, the insurance company will be able to fulfill that claim and pay out the benefits of the contract. So, full disclosure benefits both the insurance company and the applicant.

Given the importance of disclosure as a measure of the quality and efficacy of the application process, this study focuses on improving disclosure rates in life insurance application forms by applying behavioral science principles to the design of the questions. Consideration is also given to improving the customer experience to make it easier for applicants to disclose all relevant information. Improved disclosure rates of medical diagnoses, alcohol and tobacco consumption, and other information benefit the insurance company as complete information allows for more accurate risk assessment, and this also benefits the customer by increasing the certainty of claim payment.

1.2 INDUSTRY SURVEY

As a starting point for the research, we sought to gain some insights into how the SOA membership considers behavioral science in terms of application to life insurance products and processes. In late October 2023, the SOA distributed a survey designed in partnership with the Behavioral Science Team at SCOR, which sought to understand where insurance companies are using or considering using behavioral science in life insurance business.

There were 95 respondents from across the world who completed the survey. Respondents categorized where they work into three types of organizations: life insurance companies, reinsurance companies and consulting companies servicing the life or reinsurance industry. To prevent multiple respondents from the same organization skewing results for questions focusing on an organization's views on behavioral science, 87 individual organizations were identified. There were two separate conditions for multiple responses from the same organization to be identified as individual organizations. These were if respondents were from the same organization but in different locations, and if one respondent stated their organization used behavioral science for life insurance products and processes while another stated that they didn't.

1.2.1 USAGE OF BEHAVIORAL SCIENCE IN ORGANIZATIONS

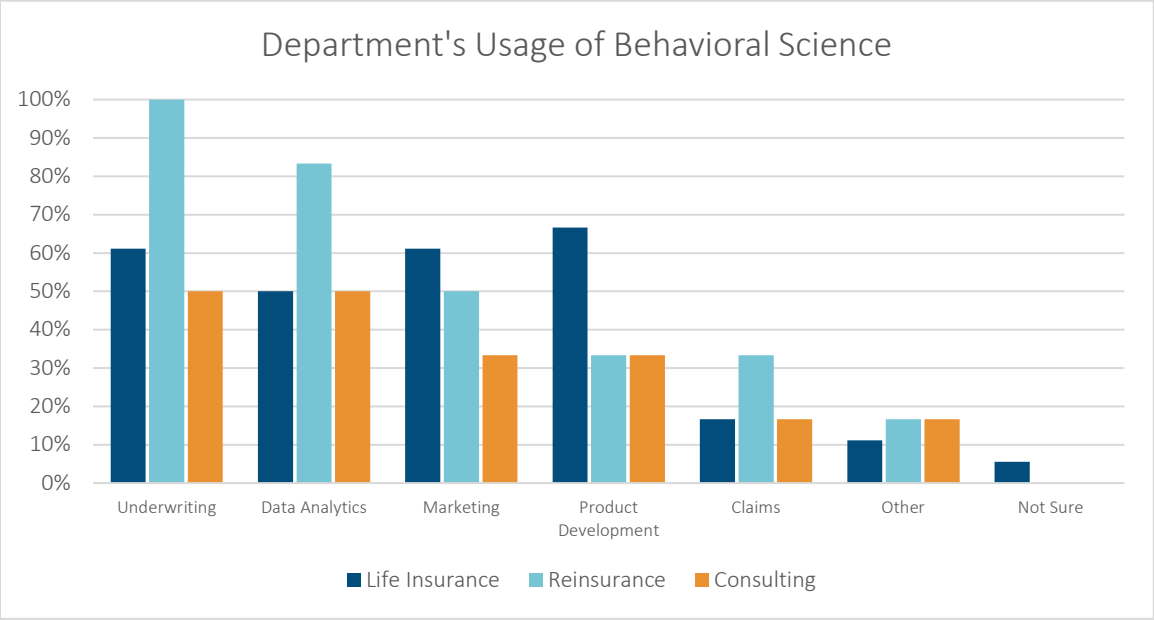
The initial focus for the survey explored where behavioral science was being used within the 87 separate organizations. Respondents were asked if their organizations used behavioral science for life insurance products and processes. Those who said yes were then asked how often their organization used behavioral science on projects and by which departments it was used.

Globally, 34.5% of organizations use behavioral science in life insurance products and processes. A significantly higher proportion of respondents in the U.S. (45%) said their organization uses behavioral science compared with the rest of the world (22.5%).

From the 30 organizations that were identified to be using behavioral science for life insurance, more than 80% used it on at least a third of their products and processes, with 16.5% using it all the time.

Figure 1

DEPARTMENTS APPLYING BEHAVIORAL SCIENCE



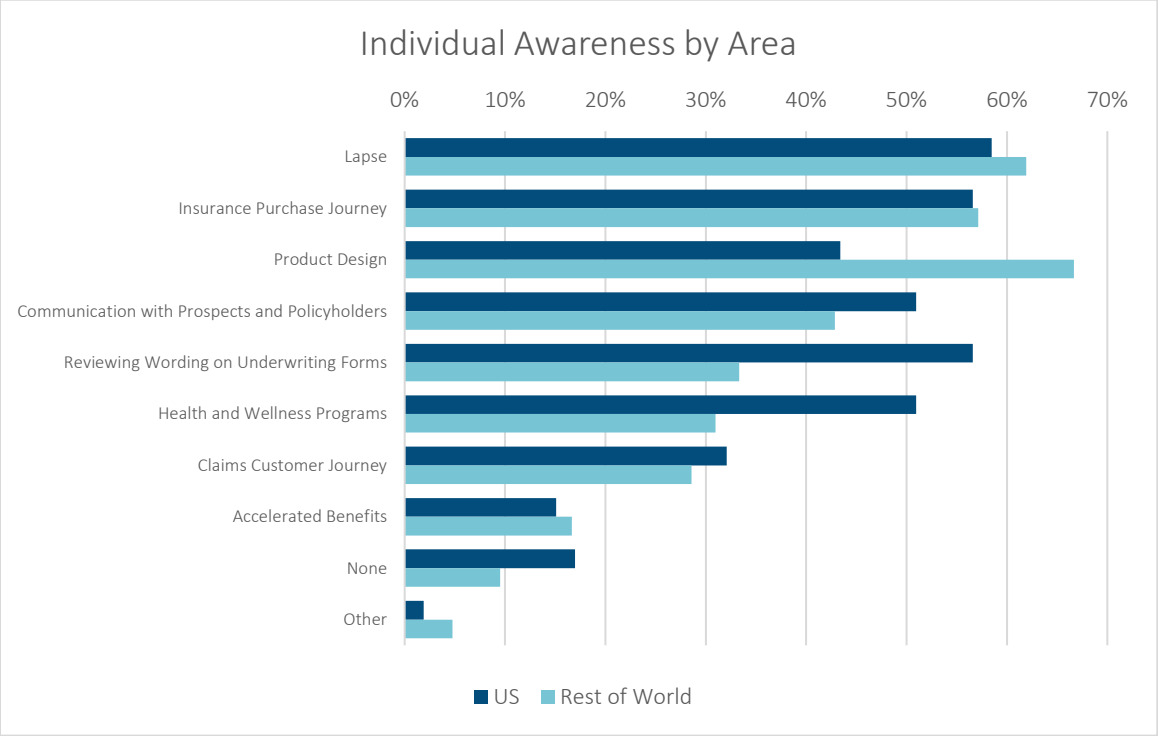
The top three departments that use behavioral science for life insurance products and processes are underwriting, data analytics, and marketing. A significantly higher proportion of underwriting, data analytics, and claims departments in reinsurance organizations use behavioral science. For respondents in life insurance organizations a significantly higher proportion of product development departments uses behavioral science.

1.2.2 INDIVIDUAL AWARENESS OF BEHAVIORAL SCIENCE

The survey explored respondents' individual awareness of behavioral science in applications and areas where they would like to learn more. They were provided with a list of areas that incorporate behavioral science where they would select all that applied.

Figure 2

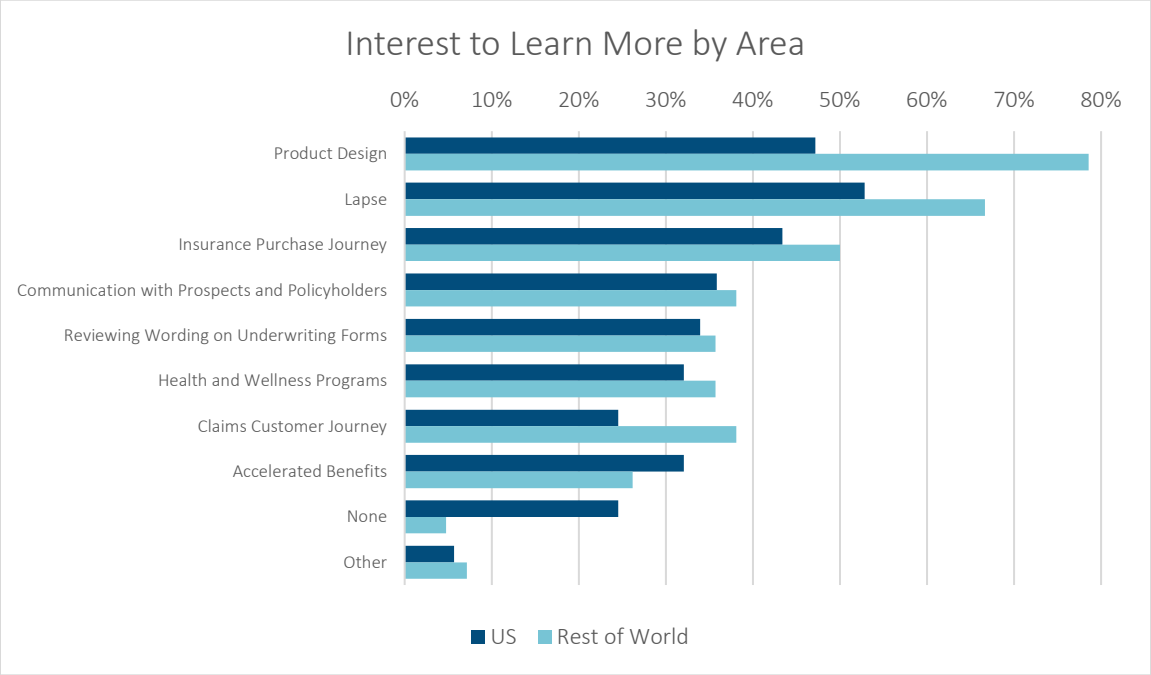
INDIVIDUAL AWARENESS BY AREA



Behavioral science in lapse, insurance purchase journey, and product design had the highest levels of awareness globally. In the U.S., there is significantly higher awareness of reviewing wording on underwriting forms (57%) compared to the rest of the world (33%). For countries outside the U.S., including Canada, there is a greater awareness of behavioral science in product design (67%) compared to the U.S. (43%). If we look at the proportion of respondents who stated that they were not aware of any areas behavioral science was used in ("None" above), we can see that significantly more U.S. respondents are unaware compared with the rest of the world.

Exploring areas where respondants wanted to learn more about behavioral science, we observed similar results.

Figure 3
INTEREST TO LEARN MORE BY AREA



Product design, lapse, and insurance purchase journey are the three areas about which most respondents want to learn more. The results almost mirror what was observed in respondents' awareness, highlighting the correlation in responses between awareness and interest in learning more. Over one-third of respondents want to learn more about reviewing wording on underwriting application forms, highlighting that this study will be educational for a significant proportion of the industry. For almost all areas, there is a higher proportion of respondents from outside the U.S. who want to learn more about them; the only exception is accelerated benefits. The proportion of respondents from the U.S. who selected 'None,' and so were not interested to learn more about any of these areas, is about 1 in 4 compared to about 1 in 20 for the rest of the world. This suggests that respondents from the rest of the world have a greater appetite to learn more about behavioral science.

1.2.3 ORGANIZATION'S ENGAGEMENT WITH BEHAVIORAL SCIENCE ON PROJECTS

Respondents from organizations that use behavioral science for life insurance products and processes were asked what methods their organizations have used in projects involving behavioral science.

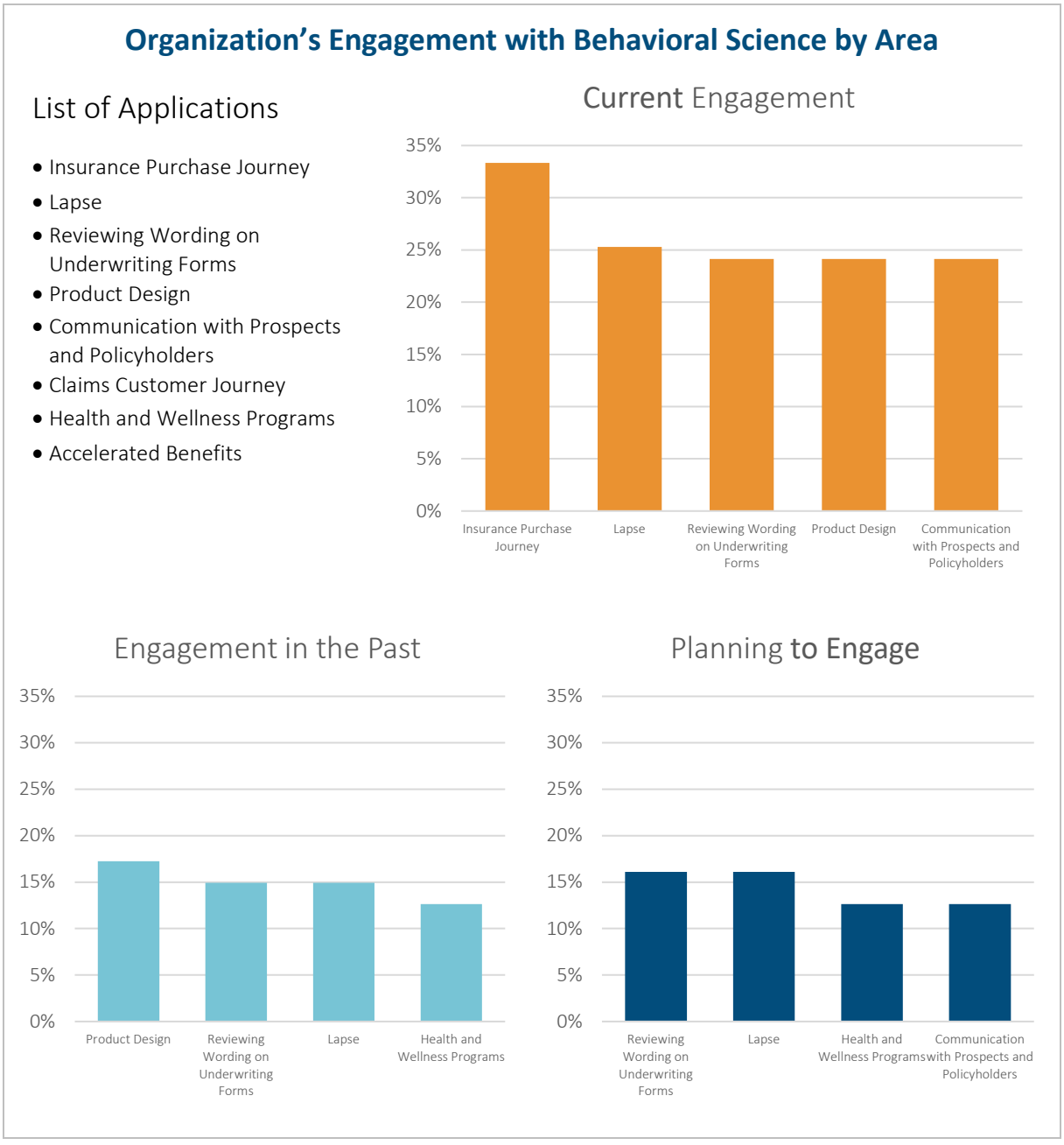
Figure 4

BEHAVIORAL SCIENCE ENGAGEMENT TYPES

Over 56% of organizations have attended conferences, webinars, seminars, or expert talks. Fifteen organizations (50%) have run an experiment, nine of which were life insurers and four were reinsurers. Fourteen organizations (47%) have created and/or run a model based on already collected data. From the list provided in the survey, collaboration with academic researchers was the least used method at 23%.

The survey expanded on areas organizations have engaged with behavioral science, focusing on when and if they plan on engaging with it in the next 12 months. Respondents from all organizations were asked, for each area listed, when they had engaged with behavioral science or if they planned to introduce it in the next 12 months. The list of areas and results highlighting the areas with the highest engagement with behavioral science are shown below.

Figure 5
ENGAGEMENT – PAST, PRESENT AND FUTURE



Currently, 21 organizations (24%) are engaging with behavioral science in reviewing wording on underwriting forms. In the past, 13 organizations (15%) engaged with reviewing wording on underwriting forms. For those planning to engage with behavioral science in new areas in the next 12 months, at 14 organizations (16%) reviewing wording on underwriting forms was one of the most popular areas.

Some organizations plan to introduce behavioral science in life insurance products and processes for the very first time in the next 12 months. These organizations were asked how they intend on engaging with behavioral science.

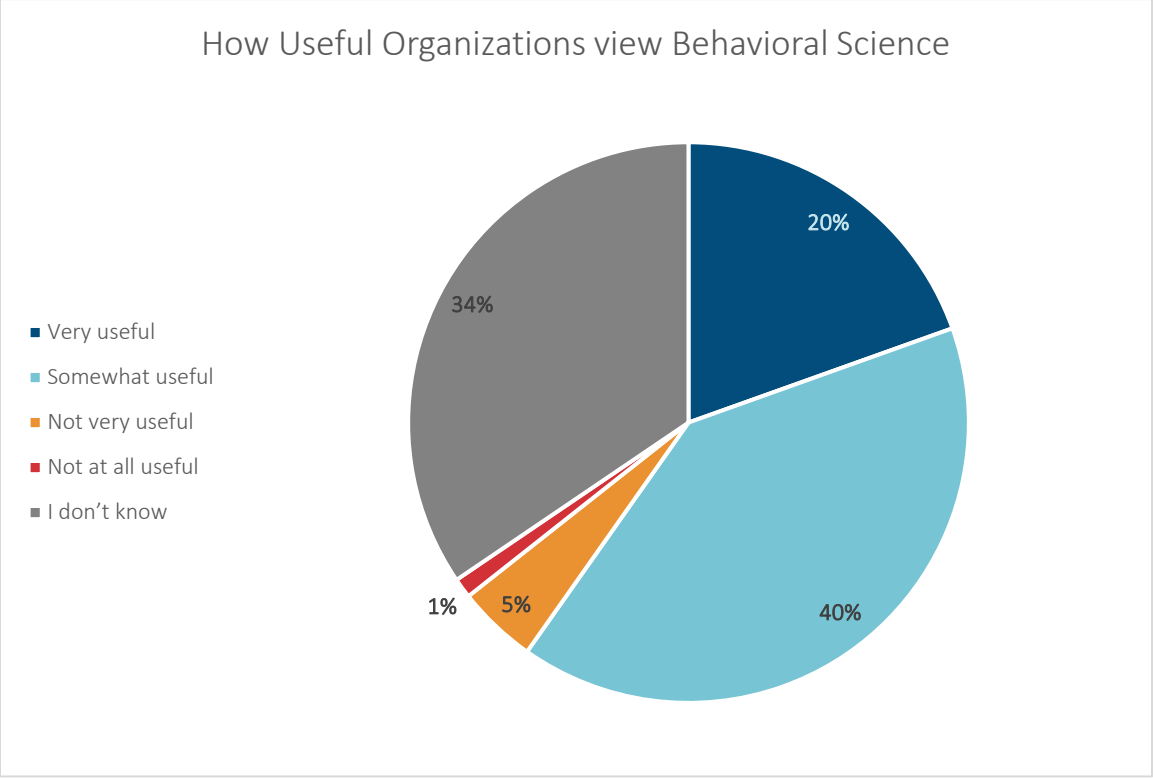
Six respondents stated that their organization plans to start using behavioral science in the next 12 months. Four of those organizations plan to start a specific project, and four plan to educate teams on the topic. It was identified that, out of these organizations, four plan to engage with behavioral science in reviewing wording on underwriting forms.

1.2.4 OPINIONS ON BEHAVIORAL SCIENCE

The survey explored respondents' opinions on behavioral science, looking at how useful organizations view behavioral science to be, if there are any benefits and difficulties in the field, and if they had any ethical concerns on the topic.

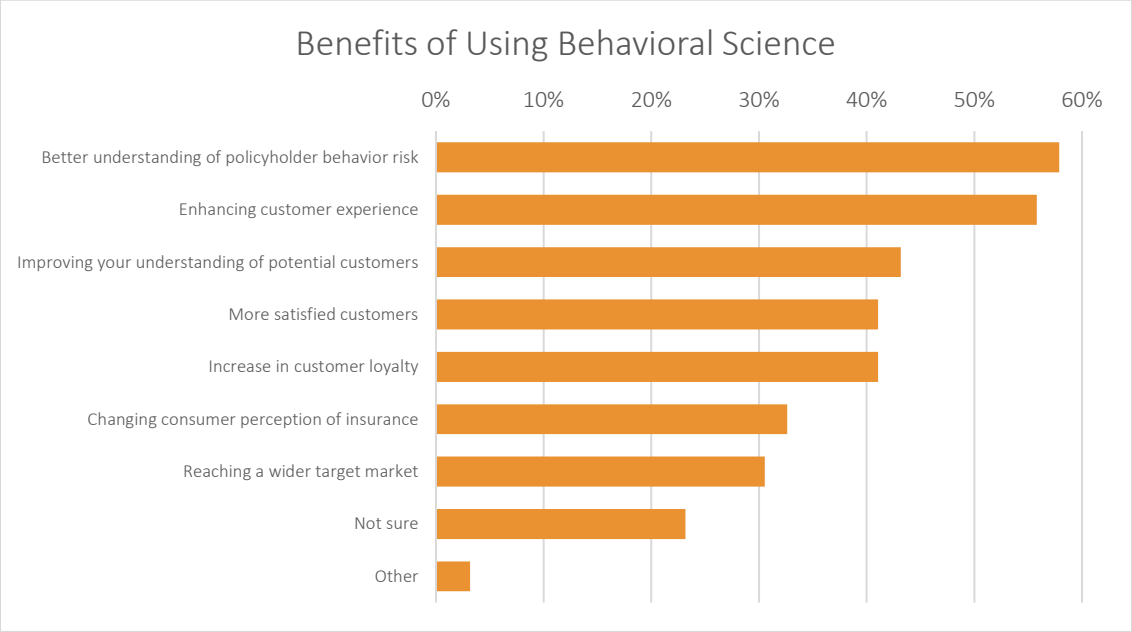
Figure 6

ORGANIZATIONS' VIEWS ON USEFULNESS OF BEHAVIORAL SCIENCE



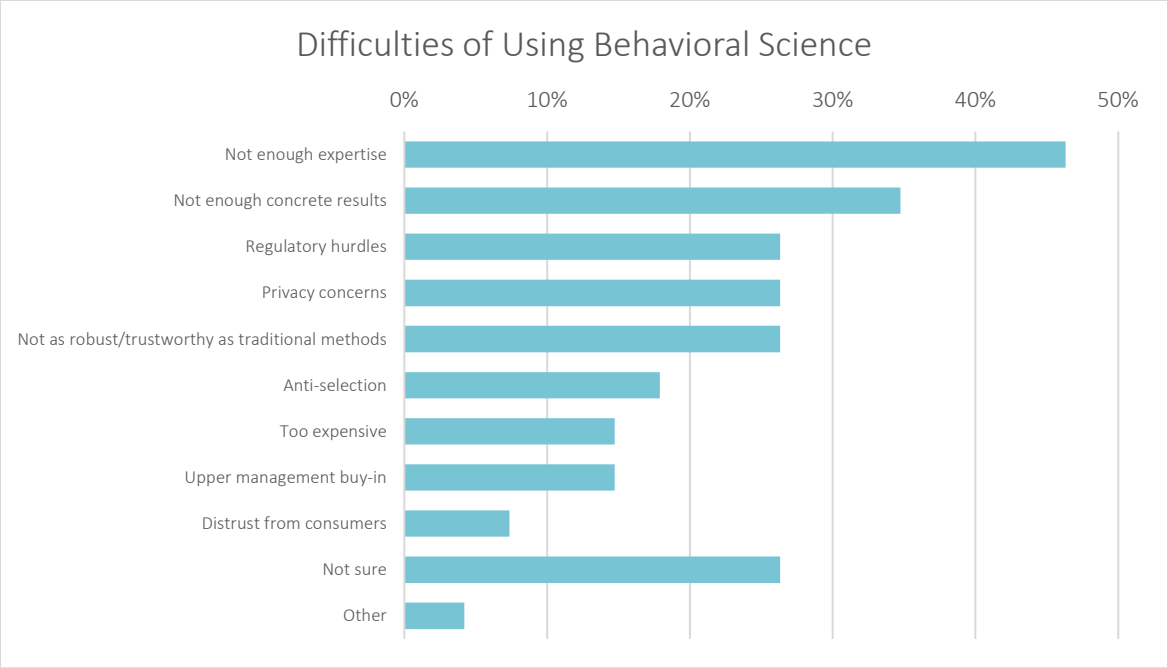
Sixty percent of respondents stated that their organization views behavioral science as at least somewhat useful in improving customer and business outcomes. Organizations in the reinsurance industry find behavioral science most useful, with 43% viewing it as very useful. In life insurance and consulting, more respondents selected 'don't know' based on having less experience with behavioral science to date.

Figure 7
BENEFITS



Individual views of behavioral science are generally positive, with each of the benefits in the list selected by at least 30% of individuals. The most common benefits were better understanding of policyholder behavior risk and enhancing customer experience. Despite not knowing their organization's view on the usefulness of behavioral science, a significant proportion of these respondents believe that enhancing customer experience and more satisfied customers were the greatest benefits.

Figure 8
DIFFICULTIES



When asked about the difficulties in using behavioral science, respondents felt that not enough expertise and not enough concrete results were the biggest challenges. Just over a quarter feel that regulatory hurdles and privacy concerns are some difficulties that will be faced with using behavioral science. A smaller percentage of respondents are concerned about difficulties with cost and upper management buy-in.

Nineteen respondents answered 'Yes' when asked if their organization had ethical concerns about using behavioral science. Most concerns focused on the topic of data privacy. Other important concerns highlighted were the potential for discrimination, risk of bias in the data, and fears of customers feeling manipulated.

1.2.5 SUMMARY OF KEY TAKEAWAYS

The results of this survey are indicative of the global opinion on behavioral science within life insurance products and processes. The majority of survey respondents indicated that their organization found behavioral science useful. Over one-third of organizations in the survey use behavioral science with more planning to use it in the next 12 months, especially in underwriting form redesign.

The biggest difficulties for organizations using behavioral science were not enough expertise and not enough concrete results. From respondents' answers, we can see that there is a high level of interest in learning more about behavioral science. This study aims to better equip organizations to apply behavioral science by providing an overview of how to apply the principles of behavioral science to redesign underwriting questions and by providing concrete results in terms of impact on non-disclosure.

With the highest number of organizations planning on engaging with behavioral science in reviewing wording on underwriting forms in the next 12 months, and over one-third of all respondents interested in learning more about it, this industry survey has highlighted the importance of this investigation's findings for the current global market.

1.3 HOW TO USE THE FINDINGS

Given the findings of the industry survey, which show the importance of the experiment results, it is worth discussing briefly how those results should be used.

1.3.1 LIMITATIONS

In the experiment, two different behavioral science designed questionnaires were tested, allowing the study to investigate two different approaches for each application form question. Neither of the two questionnaires stands out as the best overall approach. Instead, there are specific learnings for each application form question. For some questions, one of the designs clearly achieves the biggest increase in disclosure while, in others, both approaches are valuable in different ways.

The experiment was tested in an online environment, and many of the features of question design, such as reflexive questioning or conditional confirmation questions, work particularly well in an interactive online system. While variations can be designed for paper applications or interviews, the results of this study are most relevant to online application forms completed directly by the insurance customer.

In addition, this study was conducted in an experimental environment. The participants completed a health questionnaire described as being similar to the application form for life insurance. The experiment setting allowed for the testing of multiple approaches and clear scientific evidence of the impact on disclosure. However, it lacked the real-world dynamics that would be at play when an applicant is filing out an insurance application form.

In particular, participants in this study were paid to participate and did not face the same incentives as insurance applicants. Participants did not have an incentive to non-disclose to receive a better premium, as some insurance applicants might. On the other hand, participants in this study also faced no personal consequences for dishonesty – in the real world, applicants might hesitate to be dishonest for fear of their insurance not being valid when they later make a claim. Because of these limitations, further testing (see next section) should explore how behavioral science techniques could impact disclosures in a real-world setting.

Other techniques that were not tested in this study could be impactful, and a truly optimal application form would likely combine many of the techniques used in this experiment. Finally, no attempt is made throughout this report to comment on the possible pricing implications of any of these techniques were they to be implemented in real life, as such calculations would vary widely from company to company.

The recommendations and evidence presented in this report are not intended to be seen as the definitive, absolute, best way to formulate application questions. Some techniques tested work better than others, while some techniques work with specific subgroups only. Specific companies implementing the techniques detailed here might experience different results. This highlights the importance of further testing and monitoring, which can be conducted in two main ways.

1.3.2 POTENTIAL FOR FURTHER TESTING

First, the experimental framework applied in this study and defined in detail in section 2 could be used by companies to test further proposed question designs before implementing in application forms. In the industry survey, we learned that 50% of organizations using behavioral science in life insurance are running experiments. This report intends to make this method accessible to more companies by providing details on how to design an

experiment to test the impact of underwriting question design on disclosure rates. We encourage further experimentation on this topic and suggest specific aims such as:

1. To replicate some or all results of this experiment
2. To test the design of questions not considered in the scope of this study
3. To test a combination of the approaches found to be effective in this study

Second, companies that wish to apply behavioral science designs to their real-world application forms have a unique opportunity to conduct a field experiment (see the explanation of different types of experiments in subsection 2.2.2). In practice, this process would involve:

1. Gathering data about disclosure and other outcome measures under the current application process
2. Redesigning the application form based on the suggestions and evidence in this study
3. Gaining regulatory approval and fully implementing the new form
4. Carefully monitoring outcomes over a period of time to compare the new application form to the old process

To allow for an accurate comparison of disclosure rates achieved with the new application form to disclosure rates of the prior application form, the groups answering the two application forms should be aligned in terms of demographic characteristics, and any changes in distribution approaches should be considered. The exact details of this process will vary, but any company that wishes to implement behavioral science principles in their application process should focus on monitoring outcomes from any changes implemented to quantify the effects of those changes.

Overall, the results in this study provide concrete evidence and expertise, the two things that companies felt were missing the most from behavioral science work. The results combine clear, actionable suggestions for improving the application process with evidence that such changes can produce measurable impacts on important outcomes in the insurance process. This study also offers advice on applying these techniques to application forms and measuring their impact.

This experiment is not the last word on applying behavioral science to insurance, nor is it intended to be. However, it is an important and timely showcase of the power of behavioral science to improve disclosure. It should motivate and inspire others to implement behavioral science techniques in their own application processes.

1.4 CUSTOMER EXPERIENCE FEEDBACK

After the study participants completed the health questionnaire as part of the experiment, an optional qualitative feedback section asked about their experience answering the questions. This provides insights into the attitudes of consumers towards answering these types of questions. A strong majority of participants agreed that the questionnaire was easy to understand (94-96%) and easy to answer (91-94%). Approximately 4-6% disagreed with the statement that there was sufficient clarity on the detail needed.

There was no significant variation in this qualitative feedback among the groups who saw different questionnaires. A poor customer experience in completing the application form is often identified as a pain point in the sales process.

The qualitative feedback highlights that it may not be a poor experience for customers or, at least, it does not need to be a poor customer experience. The types of questions we asked in the application form can be presented in a way that most consumers find easy to understand, easy to answer and clear in terms of details needed.

The qualitative survey also identified that a significant proportion of respondents admitted to omitting answers ranging from 11-14% across groups. Reasons for omitting answers included Sometimes In Order to Answer Quickly (2-6%), Only Where I Wasn't Sure It Was Relevant (4-6%), and Only When The Question Wasn't Clear (3-5%). Despite the improvements in disclosure observed for the behaviorally-designed questions in this experiment, some participants still felt they had left out information and were willing to admit it. The behaviorally-designed questions tested in this study reduce non-disclosure, but do not eliminate it, highlighting that further improvements can be made.

These overall results do not vary among our behaviorally-designed questionnaires and the traditional questionnaire used in the experiment. Some of the qualitative feedback on specific questions does highlight differences in customer experience among the questionnaires, and these are discussed in section 3.

Section 2: Experiment Methodology

2.1 STAGE 1: CONSULTING PREVIOUS RESEARCH

This experiment is built on previous work in the fields of behavioral science and insurance. Our team reviewed the academic literature on improving honesty and disclosure of medical conditions and personal attributes, including height and weight. We also utilized the work of Pe'er et al. (2022) on honesty pledges.

In addition to academic research, our team incorporated insights from underwriters and actuaries to leverage their experience on application form design. We gathered their perspectives on which parts of the application questionnaire were most problematic, what their experience with behavioral science modifications had been (if any), and what they were interested in exploring further. We also discussed informally with the Interstate Insurance Compact about how regulators viewed some of our ideas about possible changes to the application form.

Finally, our team leveraged SCOR's past work in this area, including a previous experiment conducted in collaboration with students from Nanyang Business School in Singapore (Chong & Chua, 2021). In this experiment, two groups of participants were presented with either traditional application questions or questions redesigned with behavioral science principles. Questions covering alcohol, smoking, and diabetes/prediabetes were tested. The rates of disclosure from each group were then compared to determine the effect of the behavioral science redesign. Although the Nanyang experiment was small-scale, with around 200 participants, there were statistically significant results showing increased disclosure for the redesigned smoking question. The study was a valuable resource to inform the design of this experiment.

2.2 STAGE 2: DEVELOPING EXPERIMENTAL QUESTIONS AND BASIC EXPERIMENTAL FRAMEWORK

Upon review of previous work, it became clear that, while there was great interest in the potential of behavioral science, concrete and rigorous research was needed to determine the quantitative effect of various behavioral science techniques on underwriting questions. Our team set out to design an experiment to explore these answers.

2.2.1 EXPERIMENTAL FRAMEWORK

To provide the clearest evidence, we utilized the experimental framework known as a 'randomized control trial' (RCT) or A/B testing. This type of experiment is considered the gold standard in many cases for gathering robust evidence on the effect of the topic being tested. The framework was developed in the medical field, where it is commonly used to test new treatments and drugs. In line with its medical history, in this type of experiment, the techniques, systems, or practices being tested are referred to as the "treatment" (or, sometimes, the "intervention").

The basic experimental procedure is as follows: Participants are randomly assigned to one of two or more groups. They are not informed about which group they are placed in to avoid biasing their answers. One group, referred to as the control, does not receive or experience the treatment. The other groups, referred to as the treatment groups, do receive the intervention being tested. All other possible aspects of the environment are held as consistent as possible between the groups.

The measure of interest for comparing the groups is selected as the "outcome variable." Given the initial randomization of participants into groups and a large enough group size to achieve statistical significance (see next section), the assumption is that, if they encountered the same experience, the two groups should have the same results when the outcome variable is measured. If there is any difference in results observed between the groups,

this change is assumed to be due to the effect of the treatment, since this is assumed to be the only difference between the groups. This method of experimentation allows researchers to eliminate the influence of other factors to isolate the effect of the intervention.

In this experiment, the treatment consisted of the rewording of various application questions in line with behavioral science techniques. The Control group, therefore, was presented with an application questionnaire that did not contain this intervention – it mimicked the traditional application wording currently used in common life insurance application forms and not influenced by behavioral science principles.

The outcome variable of interest was the rate of disclosure of various conditions and lifestyle choices asked about in the experiment. Where the treatment groups had different rates of disclosure than the Control group, the difference was attributed to the effect of the behavioral science redesign.

2.2.2 FORMAT OF EXPERIMENT

After deciding on the basic RCT framework for the experiment, the team considered three possible options in which to administer it.

The first and most complicated option for a social science experiment like this is a field trial. In this type of experiment, the treatment is administered in a real-world setting with actual users. In the insurance context, this would involve replacing the insurance application forms encountered by some life insurance applicants with redesigned forms. One example of a field experiment in life insurance underwriting is provided by Samuell et al. (2023), comparing different recordings heard by applicants before responding to smoking-related questions. Field trials can be incredibly useful as they provide the best external validity of any experiment type. Because they take place in a real-world environment, they closely mimic how the treatment would work when rolled-out for practical use.

However, the logistical and ethical challenges of running a field trial in this context made it unsuitable for our experiment. Changing an application questionnaire is a lengthy, complicated process requiring regulatory approval. In addition to regulatory constraints, there are also administrative and logistical constraints to running two different applications at the same time, either through an online system or with an advisor/agent network. The fair treatment of customers is also an important consideration in a real-life setting, where the outcome would affect their insurance coverage and premium amounts.

A field trial could instead be conducted in a sequential manner – instead of having two different applications being used by the same company at the same time, a new application form could be implemented, and the disclosures from this application could be compared to the disclosures from the previous application form. However, such a setup would make it difficult to ensure that any changes in disclosure were due only to the change in application form, as other factors could change over time that would also affect disclosures (such as a change in the demographics of the applicant population).

The second option for conducting an RCT is to use a lab experiment. This option is most often utilized by academic researchers. It involves bringing a group of participants together in person to some controlled environment, where the treatment is administered and results are gathered. This type of experimental setup is quite helpful for ensuring strict experimental control, since everything in the environment, right down to the room temperature, can be monitored and set by the experimenters.

However, this control comes at a cost – lab experiments are expensive and time-consuming since participants must be paid to show up, and the lab must be staffed and maintained. Because of this, lab experiments are often quite limited in the number of participants who they can involve. Participants are also often limited to those in a particular geographic area near the lab or to those who can be easily recruited and who have time to participate (for instance, working single mothers from remote areas are practically impossible to include). This means that the treatment may end up being tested on a small sample of people who don't resemble the entire target population of the intervention, limiting the ability to transfer the results of the experiment to the real-world.

The final possible experimental framework, and the one utilized for this study, is an online experiment. In this setup, the experiment is conducted digitally on an online platform. Participants access the platform from their home devices and participate in the experiment on their own time, from wherever they are. This option works best when the treatment can be administered in a written format, such as through wording changes or different information, as in this experiment.

Online experiments must sacrifice some amount of experimental control, as it is difficult to monitor participants as they are completing the experiment to be sure that they are all equally engaged or in similar environments. This may result in some participants misunderstanding the instructions of an experiment or giving low-quality answers.

However, the advantage of an online experiment is that it can be completed quickly, for a reasonable cost, and can, therefore, involve large numbers of participants drawn from all geographical areas. In addition, since participating requires minimal equipment, free time and no commute, participants can be drawn from all backgrounds and subgroups of interest. The demographic characteristics of participants can even be monitored to ensure a specific target audience is reached.

Given these advantages, our team conducted the experiment online. To do so, we partnered with Dynata, a company which designs and administers online experiments and surveys around the world. We defined criteria to recruit specific samples of participants to match the profile of U.S. life insurance customers. This required a distribution as follows: 55% male, only 15% from Generation Z, and 90% with at least a college degree.

In addition, the mix of genders, ages, and education levels was balanced and controlled to ensure it was consistent across experiment groups. Other demographic factors, such as income level and employment status, were monitored and later compared to check for broad consistency between groups. Full demographic details for each of the experiment groups are available in the interactive dashboard.

2.2.3 SIZE OF EXPERIMENT

To determine the number of participants needed for this experiment, we conducted a power analysis. This is a statistical test which calculates the minimum number of observations needed to accurately detect an effect of a given size. The more participants, the greater the power of the experiment and the smaller the effect that can be detected.

For insight into the possible size of the effect that could be expected for this experiment, we examined the effects found in previous studies. Some studies (such as SCOR's previous work with students at Nanyang University) found large effect sizes of up to a 22-percentage point increase.

However, some studies found much smaller effects. In particular, Scribani et al. (2014) tested the effect on self-disclosure of BMI by informing participants that they would be weighed shortly after their disclosure. Although this study did not focus on rewording questions, it still attempted to promote honesty/accuracy through changes in the

disclosure process, making it relevant for our research. For women, knowledge of impending weighing increased obesity disclosures by almost six percentage points.

To give ourselves the best chance of detecting a statistically significant effect (while recognizing financial limitations), we set the sample size to be large enough to detect a small effect similar to that found by Scribani et. al. Our power analysis indicated that we needed to recruit 1,050 participants for each experiment group to provide a large enough sample size to detect an effect of this size.

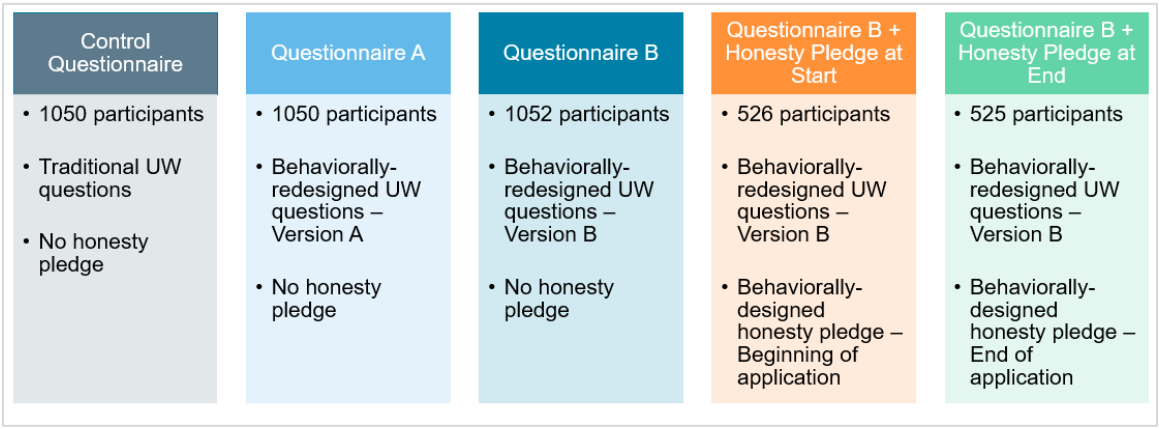
2.2.4 EXPERIMENTAL GROUPS

To comply with the basic RCT framework, the experiment needed at least one control group and one treatment group. However, since we had more than one possible technique to test for each application question, we incorporated two treatment groups, with each one facing a different questionnaire redesigned with behavioral science techniques.

After consultation with the SOA Project Oversight Group (POG) and subject matter experts within SCOR, we implemented two further experimental groups to allow us to test the additional effect of adding an honesty pledge to the redesigned application form, either at the beginning of the questionnaire or at the end. To manage the overall size of the experiment, we recruited a smaller number of participants in each of these two groups (525 participants) so that it was half the size of the control and first two treatment groups. Though this does slightly reduce the power of our analysis for these two groups, we should still be able to detect even small effects – as small as a 10% increase, for instance.[1]

Figure 9

EXPERIMENTAL GROUPS



[1] For those familiar with power analyses, in comparisons between the Control group and the two groups (QA and QB) with 1050 participants, we should be able to detect an effect size of Cohen's *h* = 0.12, which is considered a small effect in social science research. In contrast, in comparisons between the Control group and the two groups (QB + H_Start and QB + H_End) with 525 participants, we should be able to detect an effect size of Cohen's *h* = 0.15.

2.3 STAGE 3: DESIGNING EXPERIMENTAL MATERIALS

2.3.1 CONTROL QUESTIONNAIRE

To determine the content and wording of the Control Questionnaire, we reviewed publicly available life insurance application forms from several U.S. life insurance companies. We focused on common elements and trends in these forms and used them to synthesize a questionnaire for our Control group that mirrored a traditional life insurance form currently in use in the U.S.

We focused on eight key sections of the form:

1. Height/Weight
2. Weight Change
3. Mental Health
4. Medical Conditions
5. Additional Medical Conditions
6. Alcohol
7. Tobacco
8. Substance Use Support

In addition, we asked several qualitative questions after participants had completed their assigned form to gather their opinion about the forms and their experience completing them.

2.3.2 TREATMENT QUESTIONNAIRES

Using the Control Questionnaire as our starting point, we discussed potential behavioral science interventions that could be applied to the application form to improve disclosure. We then matched each technique with the question(s) that we felt would best serve to test its effectiveness. Throughout this process, we prioritized testing a wide range of techniques across the questions while, (as far as possible) only testing one technique at a time per question. In this way, we isolate the effect of each technique while still testing multiple techniques across the application.

2.4 STAGE 4: PROGRAMMING AND TESTING EXPERIMENT

2.4.1 INITIAL PROGRAMMING AND TESTING

Dynata handled the actual programming of the experiment, translating the written questionnaires created by our team into an online survey platform. This process was iterative, a repeating cycle of programming, review, feedback, and further programming. During this process, we sought review and feedback from a wide range of stakeholders, including our team, the SOA POG, and others at SCOR, with the ultimate goal of making the survey as easy to understand and use as possible. We also submitted the survey to numerous stress tests, attempting to “break” it in various ways to anticipate potential problems and correct for them before we began gathering data.

2.4.2 SOFT LAUNCH

When the initial process of programming and testing the survey platform was complete, we piloted the survey, sending it out to approximately 400 participants, about 10% of our target sample size. This “soft launch” allowed us

to perform a final test of the survey with real participants. After analyzing the responses from this pilot sample, we identified a small number of further issues with the survey, which we corrected before sending the survey out to the full sample of participants. These changes were very small and did not affect the overall content of any of the questionnaires, and preliminary analysis indicated that the results from the first and second launches did not differ significantly. Therefore, we combined the data from the soft launch with the data gathered from the full launch in our final analysis.

2.5 STAGE 5: ANALYZING EXPERIMENT RESULTS

Upon achieving the target sample size of 4,200 participants, we carried out a comprehensive set of data checks to ensure that the data was complete, before we began our analysis. We organized the analysis by question, in each case identifying the outcome variable of interest and comparing this variable among the experimental groups, testing for statistically significant differences.

For instance, for the Mental Health question, the outcome variable of interest was the proportion of respondents in each group who disclosed a mental health condition. Once we had calculated this proportion for each group, we used a two-proportion z-test to test for a statistically significant difference between the proportion in the Control group and in the Questionnaire A Group, then between the proportion in the Control group and in the Questionnaire B Group, and so on.

In all our tests, we used p-values to indicate whether the results should be considered significant. P-values are widely used in statistical analysis but require some explanation. When we compared the proportion of respondents in the Control group and Questionnaire A who disclosed a mental health condition, we wanted to know whether Questionnaire A had an effect on disclosures. We could observe that the percentage of respondents disclosing differed between these two groups, but this could be due to random variation.

The p-value of the statistical test shows the likelihood of finding the results that we did if Questionnaire A did not actually have an effect. In other words, it indicates the likelihood that the difference between disclosures in the two groups is only due to random variation. A p-value of 0.05, therefore, means that there is only a 5% chance that the results are due to random variation and Questionnaire A actually had no effect.

The inverse of a p-value is known as the significance level. In this study, we used a significance level of 95%, which is considered standard in most social sciences. This means that only results with a p-value of 0.05 or less are accepted as significant and meaningful.

After completing this basic analysis for each question, we completed a subgroup analysis where possible. In this analysis, we divided the data into subgroups based on various demographic variables and retested the differences among groups to see if the treatment effect differed among demographic groups. For instance, with Mental Health, we divided the responses by gender and tested to see if the proportion of men reporting a mental health condition differed among groups.

This analysis allowed us to determine the effect of each behavioral science technique tested on disclosure rates for each question. The full results by question can be seen in the interactive dashboard. In the following sections of this report, the main findings of the analysis will be summarized and explained.

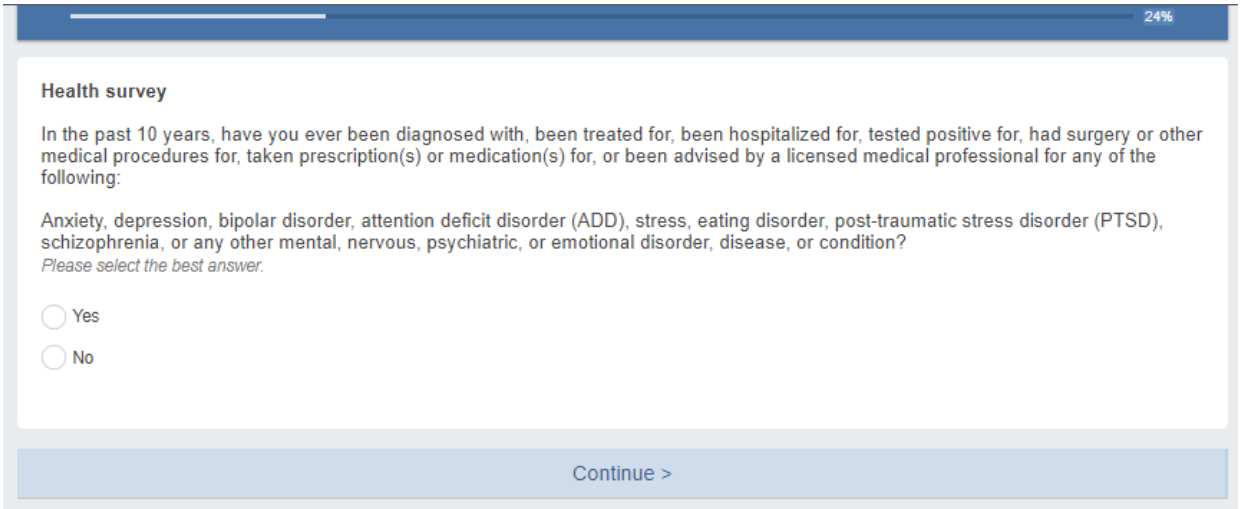
Section 3: Key Findings

3.1 ASK ONE THING AT A TIME

3.1.1 THE CONCEPT

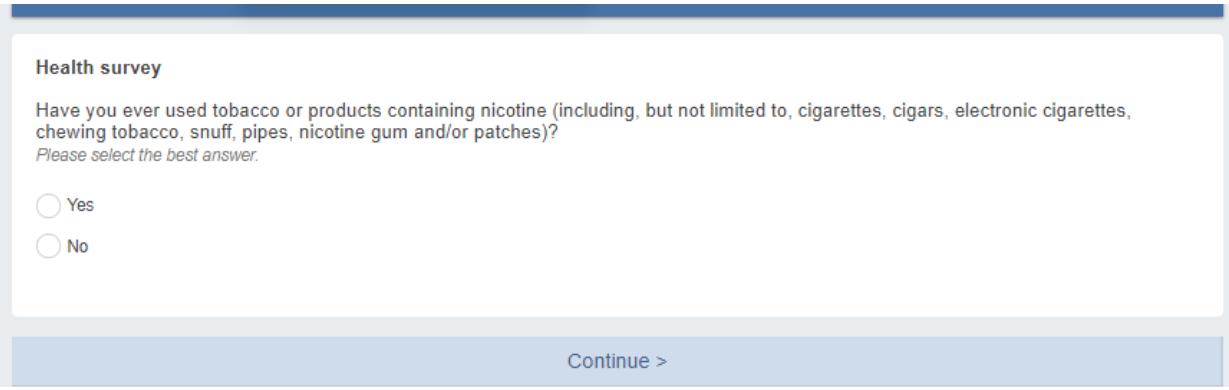
Many questions in a traditional life insurance application form ask about multiple things at the same time. This trend has only increased with recent efforts to shorten the overall number of questions on the application form – one way to do this is to combine several questions into one. For instance, a traditional question about mental health conditions (used in the Control Questionnaire of this experiment) might look like this:

Figure 10
CONTROL MENTAL HEALTH QUESTION



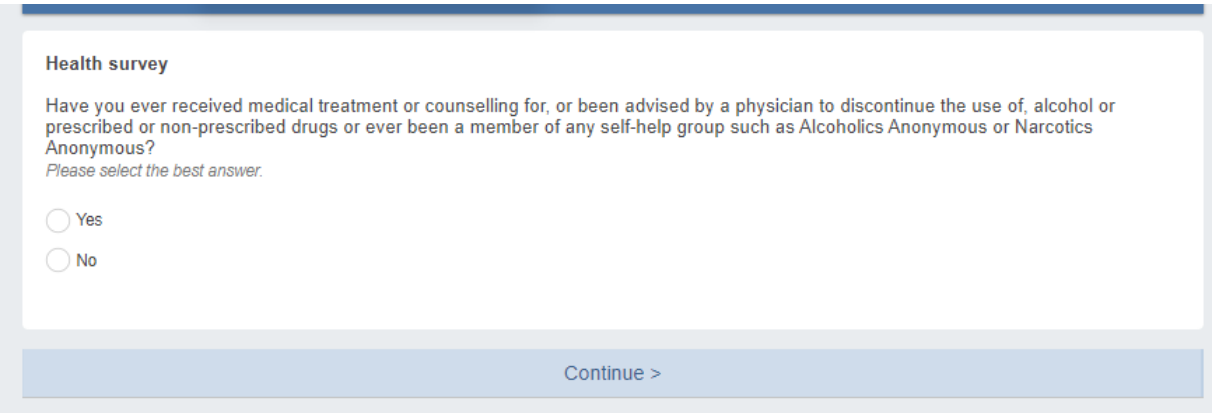
An applicant might encounter a question about tobacco products in a similar format:

Figure 11
CONTROL TOBACCO QUESTION



A final example is the traditional question about substance use support:

Figure 12
CONTROL SUBSTANCE USE SUPPORT QUESTION



This style of question asks the applicant to answer Yes/No to an entire list of conditions, experiences, or products. Reflexive follow-up questions might ask the applicant to clarify which of the list specifically applies to them but, to reach this stage, the applicant much first say "Yes" to the entire list.

Previous research conducted by SCOR's Behavioral Science indicated that questions which group many different conditions/experiences/products together tend to produce more non-disclosure than those which separate long lists into individual questions. One reason for this is social desirability bias – humans are social creatures who prefer to fit in and be in line with the social norms they see around them. They may hesitate to disclose information that they believe would place them outside of those norms or lead to judgement or stigmatization from others.

Mental health is often a stigmatized issue. Furthermore, certain mental health conditions are more stigmatized and viewed more negatively than others. Combining all mental health conditions into one list, therefore, groups conditions with a wide range of societal norms and implications together. Applicants who have a condition like anxiety or depression, which are more "accepted" and less stigmatized, may hesitate to say "Yes" to the overall list, which also contains much more stigmatized conditions such as schizophrenia and bipolar disorder. This can lead applicants to answer "No" to the entire list to avoid being grouped with people who experience those conditions and the societal stigma that goes along with them.

Tobacco use is often seen as a "bad habit," and applicants may worry that the insurance company views it as something wrong or shameful. However, some tobacco products are viewed differently than others. In particular, electronic cigarettes are often seen as a healthier alternative to traditional smoking products such as cigarettes and cigars. Grouping all these products together means that some people who use only electronic cigarettes may hesitate to say "Yes" and group themselves with traditional smokers.

Substance use support is another highly sensitive topic. An applicant who has been advised by a physician to discontinue their use of alcohol (for whatever reason) may hesitate to say "Yes" to a list that also asks about treatment for alcoholism. Again, this can lead to non-disclosure, as this applicant answers "No" to the entire list to avoid admitting to things that do not apply to them.

A possible method to solve this issue is to avoid grouping conditions, experiences, or products of varying severity and/or social stigma together into one question with one Yes/No answer. Instead, applicants should be able to

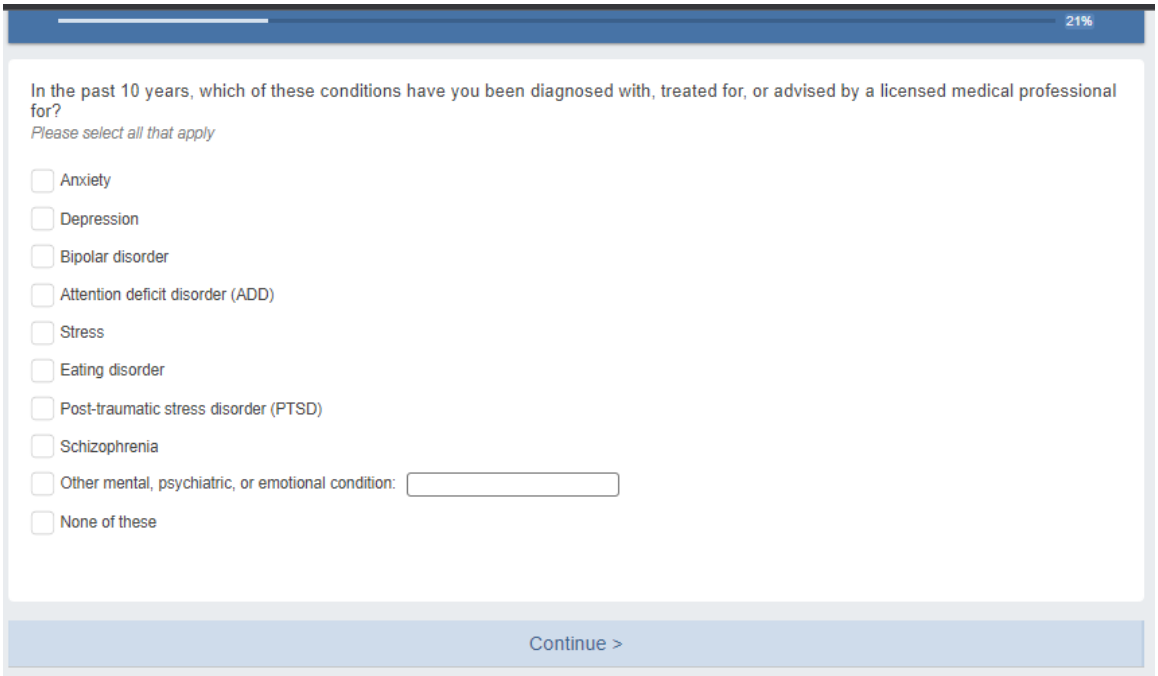
immediately admit to only what applies to them, without having to simultaneously say “Yes” to any other conditions.

3.1.2 HOW IT WAS TESTED

In this experiment, we tested this technique of asking about one thing at a time in two different ways. For the Mental Health question, Questionnaire A skipped the traditional Yes/No question and went directly to a list of mental health conditions with individual checkboxes. This format allowed respondents to only check the boxes of conditions that applied to them, while being able to avoid the other conditions completely.

Figure 13

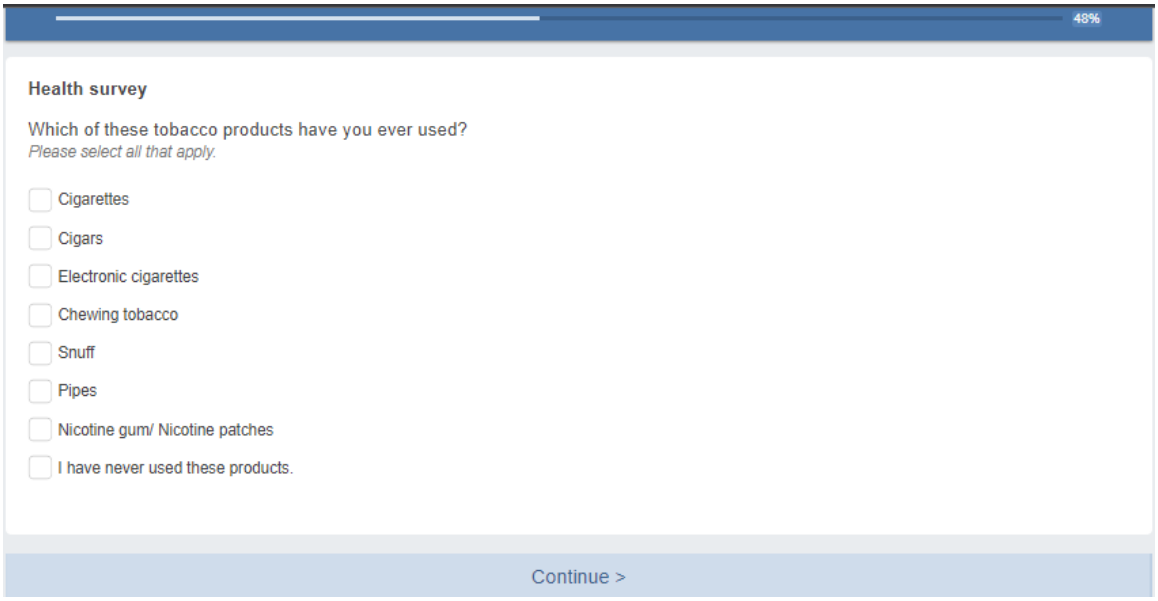
QUESTIONNAIRE A MENTAL HEALTH QUESTION



A similar format was tested with the Tobacco question in Questionnaire A which, again, skipped the initial Yes/No question with the long list of tobacco products and instead presented the following:

Figure 14

QUESTIONNAIRE A TOBACCO QUESTION



In a second variation on the idea of asking about one thing at a time, Questionnaire A divided the traditional Substance Use Support question into six individual Yes/No questions, asking separately about medical treatment, physician advice, and support group attendance for alcohol and narcotics. In this way, respondents could say "Yes" to only the experience that applied to them, while saying "No" to anything else.

3.1.3 FINDINGS

Both versions of the "Ask One Thing at a Time" approach were successful at increasing disclosures compared to the Control Questionnaire. In the Mental Health question, about one-third of respondents in the Control group, which used the traditional, single question format, reported at least one mental health condition. In contrast, when Questionnaire A used individual checkboxes for each condition, over half of the respondents in this group disclosed a mental health condition. This difference was statistically significant (see the explanation of statistical significance in subsection 2.5).

This increase in disclosure persisted when the results were broken down by demographic subgroups, but some subgroups saw an even more dramatic increase. For instance, all age groups showed an increase in disclosure when the question format was changed to ask about one thing at a time. However, the oldest respondents saw the most dramatic increase when presented with Questionnaire A.

This could be because older respondents had more room for improvement, since they had the lowest disclosure rate of mental health conditions in the Control group. It could also be the case that older individuals are particularly sensitive to stigmatization around mental health; though the evidence on this is mixed, the World Health Organization (2023) points to stigma as a primary barrier to older adults seeking help with mental health issues. If older adults do worry about experiencing stigmatization if they admit to mental health conditions, this may make them especially responsive to the change in Questionnaire A, which helps to reduce feelings of social stigma.

In the Tobacco question in the Control group, which asked about use of a long list of tobacco products in a Yes/No format, 28% of respondents reported using some tobacco product in the past 12 months. This increased to 33% for

those who were presented with individual checkboxes for each tobacco product in Questionnaire A. This is a 20% increase in disclosure of current tobacco usage.

In contrast to the results for the Mental Health condition, results for the Tobacco question differed by demographic group. Questionnaire A was most effective for younger respondents, for males, and for those with higher levels of education. For some demographic groups (such as older respondents, females, and those with lower levels of education), there was no statistically significant difference between Questionnaire A and the Control.

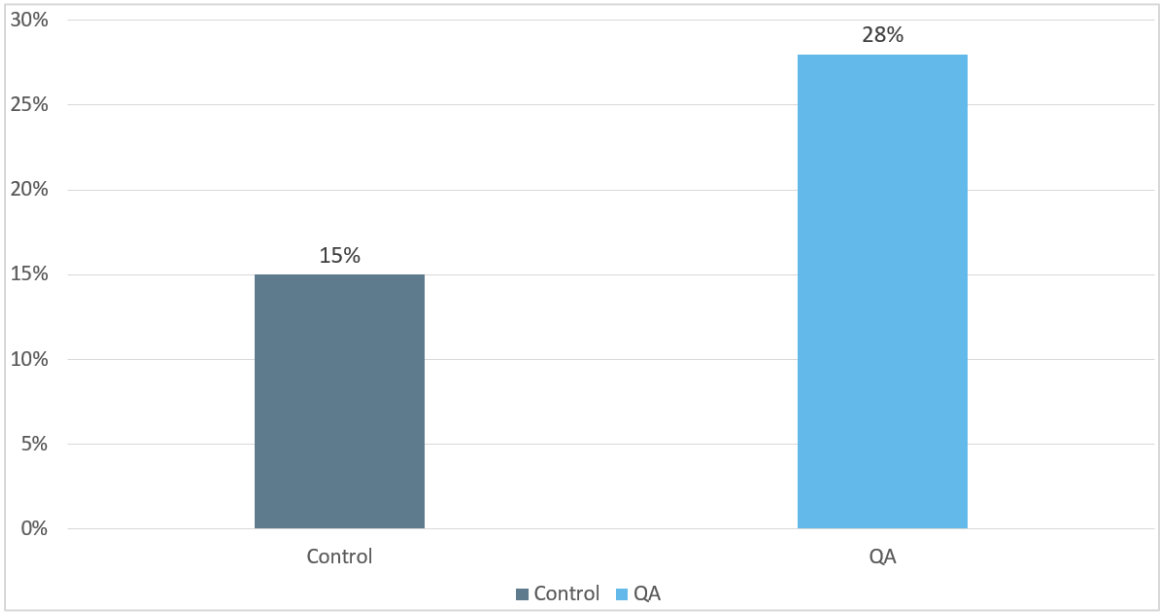
In the Substance Use Support question, slightly more than 1 in 10 respondents in the Control group reported receiving treatment for substance use, being advised to discontinue their substance use, and/or participating in a substance use self-help group. When the traditional single question was divided into six separate questions in Questionnaire A, the disclosure rate more than doubled to over 2 in 10 respondents.

As in the Mental Health question, this increase was present among almost all demographic subgroups. Once again, however, older respondents saw the strongest effect of the behavioral science treatment, with their disclosure rate from Questionnaire A being nearly four times as high as the rate for the oldest respondents in the Control group.

As mentioned above, the theory behind this technique was that non-disclosure in these questions is due (at least in part) to applicants who have less severe/stigmatized conditions or experiences answering “No” to the entire question to avoid being grouped with conditions or experiences that don’t apply to them. This theory is supported by an examination of the differences in which conditions/experiences were disclosed in Questionnaire A compared to the Control Questionnaire.

For instance, in the Control group Mental Health question, 15% of respondents disclosed “Stress,” a relatively common, mild, socially-accepted mental health condition. Those who disclosed “Stress” in the Control group would have also had to say “Yes” to the initial long list of other mental health conditions. By comparison, in Questionnaire A, where they could check only the conditions that applied to them, 28% of respondents disclosed stress, an increase of 87%.

Figure 15
DISCLOSURE RATES – STRESS



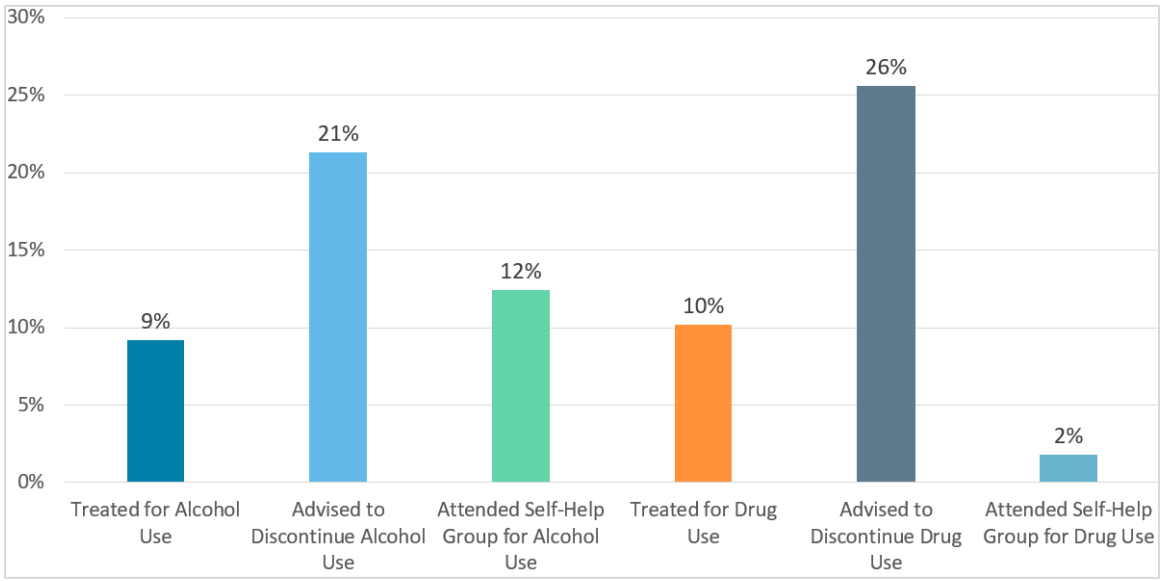
In the Tobacco question, the technique of asking about one thing at a time was particularly effective in targeting disclosure of electronic cigarette use. Around 7% of respondents in the Control group, who faced the traditional Yes/No question with a long list of tobacco products, disclosed using electronic cigarettes. In contrast, around 14% of respondents disclosed e-cigarette use in Questionnaire A. Separating electronic cigarettes from other products doubled the disclosure rate for this product.

A similar pattern can be seen for Substance Use Support. Here, it would be expected that the least severe/stigmatizing experiences to admit would be having been advised by a physician to discontinue use of alcohol or drugs. Physicians may advise this for several reasons, many of which an individual can justify in their mind as having nothing to do with addiction. Therefore, if this is an individual's only experience with substance use support, they may avoid disclosing it when faced with the long list of other addiction-related experiences in the Control Questionnaire. It is this group of individuals that would be expected to be most affected by the splitting of the single question into multiple choices in Questionnaire A.

Indeed, the experiment results show that, of those who disclosed being advised to discontinue their use of drugs or alcohol in Questionnaire A, 21% and 26%, respectively, did not say "Yes" to any other Substance Use Support question. In contrast, if a participant said "Yes" to one of the other Substance Use Support questions ("have you been treated for/attended a self-help group for alcohol/drugs?"), they were much more likely to also say "Yes" to another question.

Figure 16

WHAT PERCENTAGE OF RESPONDENTS WHO SAID "YES" TO EACH SUBSTANCE USE SUPPORT QUESTION SAID "YES" TO THAT QUESTION ONLY?



Many of those who said "Yes" to being advised to discontinue their use of alcohol or drugs in Questionnaire A would likely have said "No" to the longer Control question, since they did not want to admit to any of the other components of this question. The format in Questionnaire A allowed them to instead say "Yes" only to the component(s) that applied to them, increasing overall disclosure from 36% to 52%.

3.2 SOCIAL NORMS

3.2.1 THE CONCEPT

As mentioned above, humans are social creatures and feel most comfortable when they are in line with the social norms around them. They do not want to be seen as extreme or exceptional, especially in a negative way. If applicants think that disclosing certain information would reveal them to be outside the social norm, they may fail to disclose this information to protect themselves psychologically.

However, applicants do not always have clear or correct ideas about the social norms in any given situation. What weight is too extreme to "fit in?" What amount of alcohol consumption is so high as to be shocking? In the absence of good answers to these questions, applicants may instinctively err on the side of underreporting their weight or alcohol consumption, for instance, to be sure they do not violate the social norm.

Insurance application forms can leverage this tendency to pay attention to social norms and, through formatting and wording, suggest what the social norm is to put applicants at ease and encourage honest disclosure of even relatively extreme information. There are two main techniques to be used here – default assumptions and anchoring, both of which frame the question in such a way that the behavior to be reported appears to be within the social norm.

3.2.2 HOW IT WAS TESTED

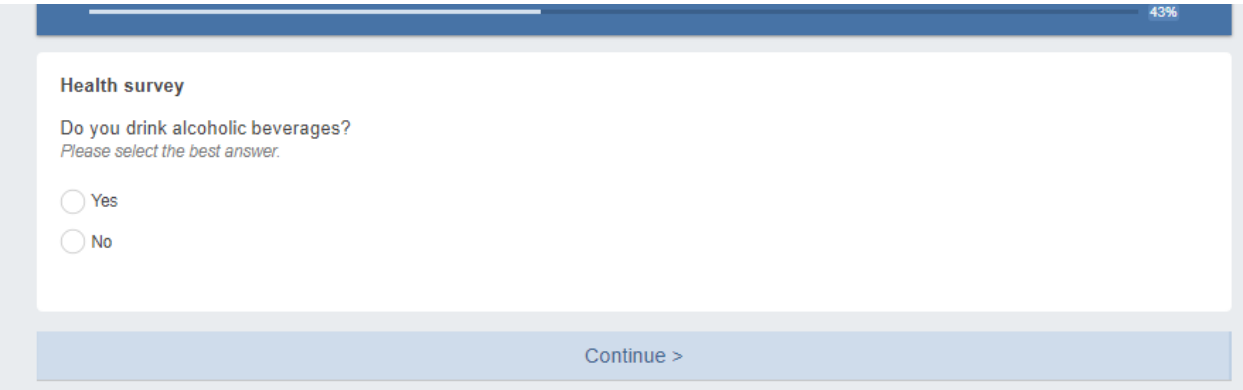
Default Assumptions

First, social norms may be suggested through default assumptions– questions that assume something is true unless the applicant specifies otherwise. Default assumptions may be perceived as indicating the social norm – the "normal" option that the designers of the question expect to apply to most people.

In traditional insurance application forms, many questions are presented without default assumptions, such as in the following Control question on Alcohol use:

Figure 17

CONTROL ALCOHOL QUESTION



On the surface, this question assumes nothing and, therefore, provides no explicit social norm. However, this leaves the applicant to make their own assumption about the social norm. Given the context of the insurance application, which asks similar Yes/No questions about many stigmatized behaviors (such as tobacco use, substance abuse, and driving violations), applicants can infer that the norm expected and desired by the insurance company is to not drink alcohol. If the applicant does drink, they will then feel outside the norm and hesitate to disclose this information.

In this way, a Yes/No question in this format actually can suggest a social norm that engaging in this behavior is wrong. It is as if the application form repeatedly asked applicants: "Do you do any of these bad things?"

A similar format is often used for questions about weight changes, mental health, and tobacco usage, as reflected in the Control questions about these subjects. Full question wording can be viewed in the interactive dashboard.

By introducing a default assumption, the application form can shift the social norm. In this study, several questions were redesigned so that they assumed the behavior or condition in question existed. For instance, the Alcohol questions in Questionnaire A and Questionnaire B assume, as the default, that applicants do drink alcohol:

Figure 18
QUESTIONNAIRE A ALCOHOL QUESTION

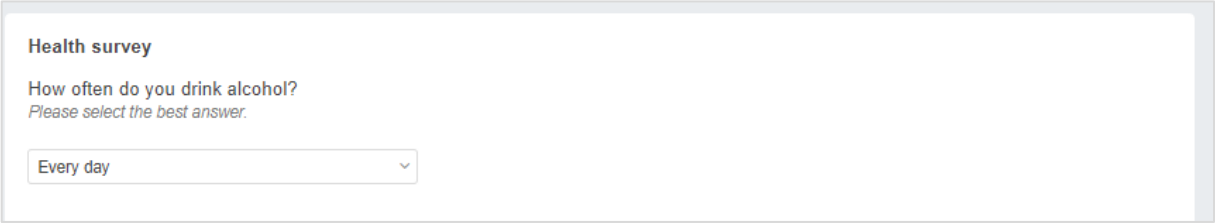
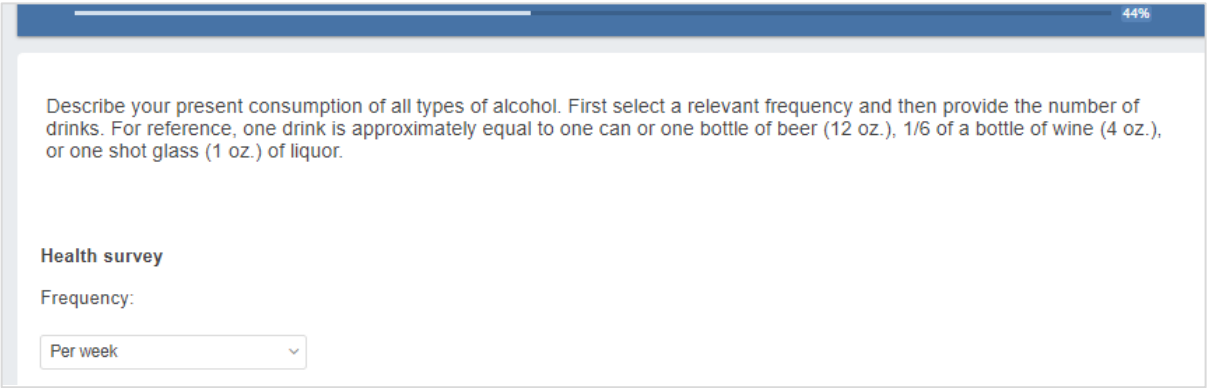


Figure 19
QUESTIONNAIRE B ALCOHOL QUESTION



In both questions, applicants were able to indicate, through their choice of frequency, that they never drank alcohol. However, the default assumption that they do drink indicates implicitly that they are expected to drink – because this is the social norm. This helps applicants feel more comfortable disclosing their alcohol usage.

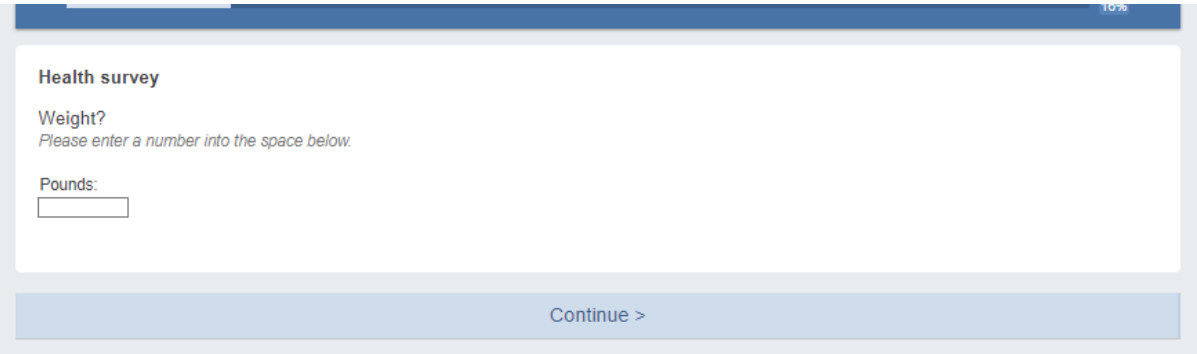
Similar changes were made to the Weight Change questions in Questionnaire A and Questionnaire B and the Mental Health and Tobacco questions in Questionnaire A to indicate that the default assumption in these questions was that applicants had experienced weight changes, did have mental health conditions, and did use tobacco products.

Anchoring

A second technique that was tested to introduce social norms into the insurance application form was anchoring. In this context, anchoring refers to the tendency of humans to be influenced by numerical information presented as part of the question, especially when that information might indicate a social norm.

For instance, in the traditional question about an applicant's Height and Weight, applicants are asked to provide their answers in an open-ended format (note that the Control questions about the amount of weight change, alcohol usage, and tobacco usage follow a similar format):

Figure 20
CONTROL WEIGHT QUESTION



As discussed above, in the absence of any indicated social norms, an applicant may assume that a "normal" weight is fairly low and worry about disclosing his/her own weight if it is higher than this imagined norm.

However, when reference amounts are introduced to this question, applicants can "anchor" on these numbers and use them to understand the social norm, putting them at ease as long as their own weights fall within the range indicated. Using high numbers as anchors will, therefore, reassure overweight applicants and motivate honest disclosures of relatively high weights.

In this study, anchors were introduced through either closed-end options or slider bars. Slider bars were used for the Height/Weight question in Questionnaire A and the Alcohol question in Questionnaire B.

Figure 21
QUESTIONNAIRE A HEIGHT/WEIGHT QUESTION

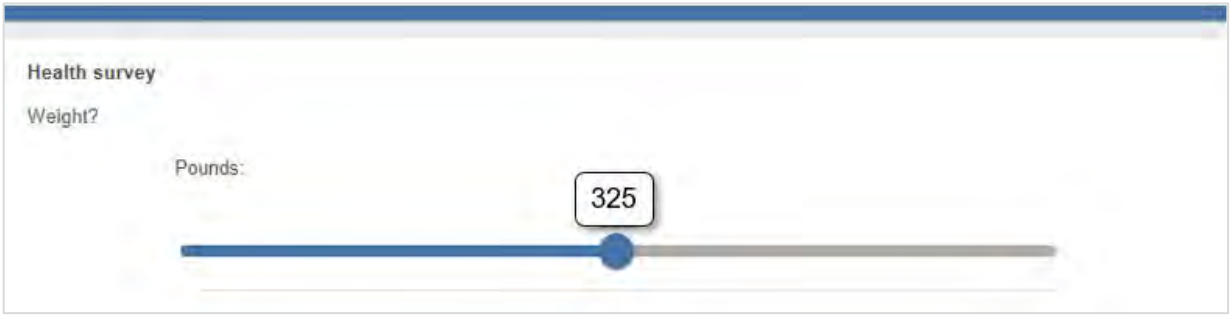
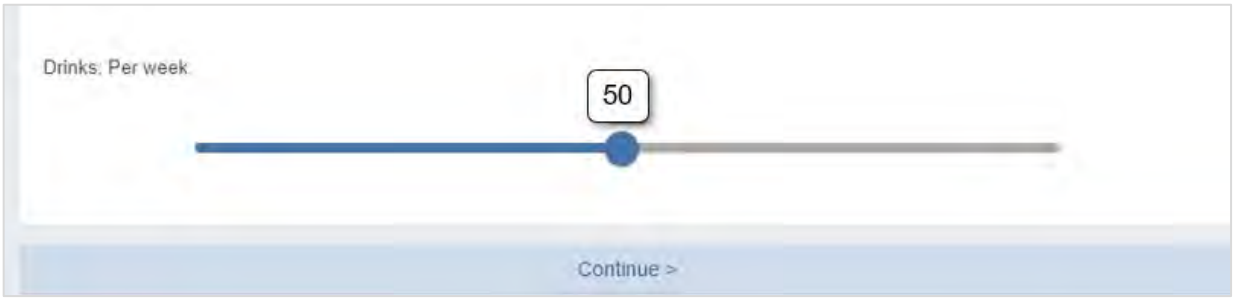


Figure 22
QUESTIONNAIRE B ALCOHOL QUESTION



The starting positions of these slider bars serve as anchors and were purposely set quite high to suggest to applicants that high weights and amounts of alcohol consumption were within the social norm.

Closed-end options were used for the Alcohol and Tobacco questions in Questionnaire A and the Weight Change question in Questionnaire B.

Figure 23

QUESTIONNAIRE A ALCOHOL QUESTION

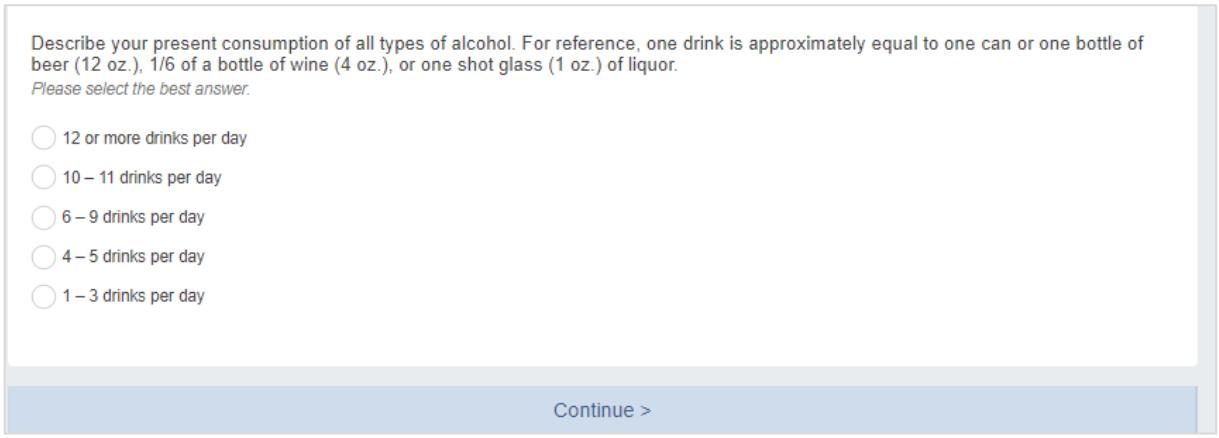


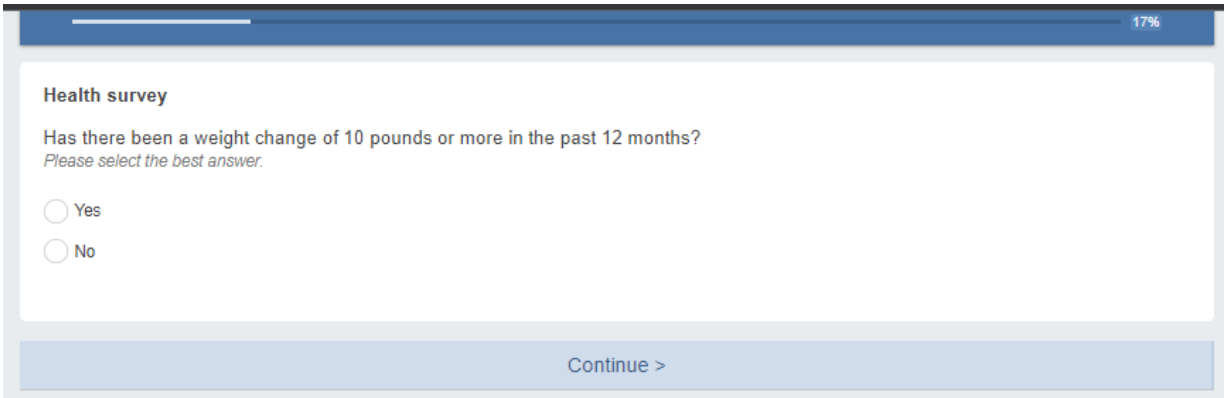
Figure 24

QUESTIONNAIRE B WEIGHT CHANGE QUESTION

Extreme options (such as 12 or more drinks per day) were purposely included in these lists as anchors – indicators to respondents that high amounts of alcohol consumption and weight changes are expected and normal.

It is also important to note that not all anchors are helpful and, at times, removing unhelpful anchors can also help put applicants at ease. The Control question for Weight Change specifies that applicants should report weight changes of 10 pounds or more:

Figure 25
CONTROL WEIGHT CHANGE QUESTION



This cutoff is specified because only weight changes of more than 10 pounds are usually meaningful from an underwriting standpoint. However, from an applicant perspective, this anchor reveals the amount of weight change that is concerning, showing what the "wrong" answer is. An applicant who experienced a weight change of 15 pounds, for instance, may be tempted to round down their weight change amount and convince themselves that it was close enough to 10 pounds that they do not need to disclose it.

Removing this anchor, as the Weight Change questions in Questionnaire A and Questionnaire B do, is intended to prevent this line of thinking and encourage applicants to report any weight changes, not just the ones that cannot be rationalized as unimportant. In a real-world application process, underwriting rules could be designed to accept without investigation weight changes of less than 10 pounds, so that any disclosures of low-risk weight change do not need to be reviewed by an underwriter.

3.2.3 FINDINGS

Default Assumptions

As mentioned above, in both Questionnaire A and Questionnaire B, the Alcohol questions were changed to assume that the applicant did drink alcohol. Introducing this social norm appears to have been successful in improving disclosure of alcohol usage. In the Control group, around two-thirds of respondents reported drinking alcohol. This proportion increased to around three-fourths of respondents in the groups that viewed Questionnaire A and Questionnaire B, an increase of about 15%.

Similar changes were made in Questionnaire A and Questionnaire B to the Weight Change question. Here, the variable of interest is disclosures of weight changes of 10 pounds or more, as this was the cutoff specified in the Control question. Although both behaviorally-redesigned questionnaires showed an increase in reports of weight changes of 10 pounds or more compared to the Control, this increase was not statistically significant for all groups. Only those who saw Questionnaire B with an Honesty Pledge were significantly more likely to disclose weight changes of 10 pounds or more, with an increase of about 15% compared to the Control.

In Questionnaire A, the Mental Health and Tobacco questions combined this technique of default assumptions with the technique of separating conditions and products. As detailed in subsection 3.1 above, both questions produced increased disclosures compared to the Control. Mental health disclosures in Questionnaire A rose by around 45%, while disclosure of current tobacco use increased by about 20%.

These results show that introducing default assumptions can produce significant improvements in the disclosure of information that might otherwise be withheld because of social norms.

Anchoring

Slider bars were used to introduce anchors into the Alcohol question in Questionnaire B. Again, the intention here was to normalize disclosure of higher amounts of alcohol consumption, and this appears to have been successful.

Classification of an individual as a "heavy drinker" is complex and depends on other factors besides their alcohol consumption, such as their sex and age. In this study, two different approximate measures were used to classify respondents as heavy drinkers – those who reported drinking four or more drinks per day and those who reported drinking 16 or more drinks per week. (These criteria were chosen based on the categories provided as answer options in Questionnaire A [see below].)

Based on either of these measures, Questionnaire B saw a large, significant increase in disclosure of heavy drinking compared to the Control. Around 2% of respondents reported drinking four or more drinks per day in the Control group, while this percentage was four times as high at 8% in the groups that saw Questionnaire B. Using the drinks per week cutoff of 16 or more, disclosure rates of heavy drinking tripled in Questionnaire B compared to the Control.

Similar results were found when sliders bars were introduced into the Height/Weight question in Questionnaire A, although these results also reveal a potential for slider bar anchors to backfire. Disclosures of heights and weights indicating high-risk obesity (defined here as BMIs greater than 40) increased by around 70% in Questionnaire A compared to the Control. Using a high anchor on the slider bar seems to have been effective at encouraging disclosure of very high weights that otherwise might have been rounded down by applicants worried about the social norm.

However, underwriters are not only concerned about very high BMIs – very underweight individuals (defined as BMIs less than 17 for this analysis) are also seen as high-risk. Unfortunately, the results from Questionnaire A show that the high anchor on the slider bar may have made underweight individuals feel outside of the social norm, leading them to round their weight up and misreport. Around 2% of individuals in the Control group report high-risk underweight BMIs; this proportion decreases by 80% to 0.4% in Questionnaire A.

These results show that anchors are powerful ways to introduce social norms. However, they must be carefully considered for any potential unintended consequences and optimized according to the desired effect. If an insurance company is very concerned about missing disclosures of high-risk obesity, using a slider bar may be the best practice for them, though they will need to keep in mind its potential effect on high-risk underweight disclosures.

Anchors were introduced into the Alcohol question in Questionnaire A using closed-end options – lists of amounts of alcohol consumption in a specified period of time, such as 1-3 drinks per day or 8-10 drinks per weekend. Extreme ranges were included as options in order to show that high amounts of alcohol consumption were expected and "normal."

This technique appears to have been somewhat effective at increasing disclosure of heavy drinking (although only when this is defined on a per-day basis). The proportion of respondents who reported drinking four or more drinks per day doubled in Questionnaire A compared to the Control – although not as large an increase as that seen for slider bars, this is still a significant effect.

However, when looking at the proportion of respondents drinking 16 or more drinks per day, there is no significant difference between Questionnaire A and the Control. Here, using closed-end options is more effective than using no anchors, but less effective than using slider bars.

A closer look at the results may reveal why this is the case. In most frequency categories (per day, per weekend, etc.), the highest range of amounts in the closed-option list was hardly ever chosen by respondents (i.e., "12 drinks or more per day"). It seems that respondents viewed the highest options as "extreme" and didn't like the idea of being at the top end of the range of options provided (although results from Questionnaire B, which used slider bars instead of closed-end options, showed that the highest options in the list should have been relevant for many respondents).

Most respondents chose one of the middle options in any given list, showing the tendency of humans to want to be in the middle of a range and avoid extremes. Table 1 shows the alcohol consumption disclosure of those who chose to report their consumption in drinks per day, using the same categories as the closed-end options presented in Questionnaire A.

Table 1

OF THOSE WHO REPORTED DRINKS PER DAY, HOW MANY SELECTED EACH FREQUENCY CATEGORY?

Group	Total Reporting Drinks/Day	1-3 Drinks/Day	4-5 Drinks/Day	6-9 Drinks/Day	10-11 Drinks/Day	12 or more Drinks/Day
QA	62	39	16	5	1	1
QB	107	61	18	8	1	19

Very few Questionnaire A group participants selected the highest or second highest amounts option presented, but a significant number of respondents disclosed drinking 12 or more drinks per day when answering using the slider bar in Questionnaire B.

A similar pattern can be seen in the results for the Weight Change question in Questionnaire B, which also added closed-end options for the amount of weight gain or loss experienced ("I gained 5-9 pounds"). The highest option offered was "I gained/lost 30 or more pounds." Respondents appeared very hesitant to choose this option – disclosures of weight changes of 30 pounds or more actually decreased slightly in Questionnaire B compared to the Control. The majority of respondents in Questionnaire B chose the middle option listed - "I gained/lost 15-19 pounds."

As with slider bar anchors, closed-end options are a potentially effective tool that must be carefully implemented to avoid unintended consequences. In particular, the highest (or even two highest) options offered should be set to be so extreme that no one is expected to fall into those ranges – because it is likely that very few applicants will select these options even if they do fall into the ranges specified.

Finally, the Weight Change questions in Questionnaire A and Questionnaire B tested the technique of removing an unhelpful anchor - the instruction to only report weight changes of 10 pounds or more. Again, to make an accurate comparison between this and the Control group, analysis focused on weight changes of 10 pounds or more. While the proportion of weight changes of 10 pounds or more did increase slightly in Questionnaire A and Questionnaire B, this difference was not statistically significant based on our sample size. There is no evidence that removing this anchor is effective at increasing disclosure of high-risk weight changes.

Overall, the various approaches of implementing social norms tested in this study show promise for improving disclosure. However, they require fine-tuning and a careful design to maximize their potential.

3.3 METHODOLOGICAL BIAS AND DISCLOSURES

3.3.1 THE CONCEPT

Methodological bias refers to limitations in question design that constrain the scope of information collected. Application forms are designed by insurance professionals who have a specific perspective and a focus on risk assessment, as well as legal and regulatory requirements. Designers know very clearly the criteria for risk assessment and design questions to collect this information. However, those filing out the application form have a different level of knowledge on insurance risk and a different perspective on the topics covered in the questionnaire. If we think about these questions from a consumer perspective, we can design questions that resonate better with applicants and draw out more information as a result.

For example, questions about weight change typically ask about weight gain or weight loss, but what about people whose weight tends to fluctuate? It is difficult to answer a question that does not appear to have a relevant checkbox for the individual applicant's situation, and this may lead to incomplete answers or missing out on the information if the applicant feels there is not an applicable option. This also gives the applicant the opportunity to justify to him/herself (either consciously or subconsciously) that the disclosure is not relevant or does not matter and that it is okay to leave off certain information.

Another example of methodological bias would be alcohol use questions, which are too narrow in scope and fail to consider different types of alcohol consumption. A traditional question on alcohol use may ask only about weekly consumption. A question on weekly consumption is difficult to answer for those who drink alcohol daily as the number of drinks per week may seem very high when multiplying daily consumption by seven days. Further, weekly consumption is difficult to answer for those who do not drink alcohol every week. It may be tempting for applicants to say none if that is true of some weeks, and this ambiguity in the question may lead to missing out on key information for those who drink large amounts of alcohol every other week or a few times per month.

We also identify a methodological bias if a question appears burdensome to answer. If the question design makes it obvious that answering 'yes' will lead to further detail being requested, it may be tempting for applicants to rationalize that their information is not relevant enough to justify entering all the required details. Again, this requires the question designer to consider the applicant's perspective and how the question appearance may impact the applicant's attitude toward answering.

Similarly, tobacco use has different products and frequencies. Here again, there is a need to capture those who used to smoke but ceased, as well as those who are current smokers. Well-designed questions allow for all possibilities so that each applicant can find a response relevant to his/her situation, while not appearing burdensome to answer.

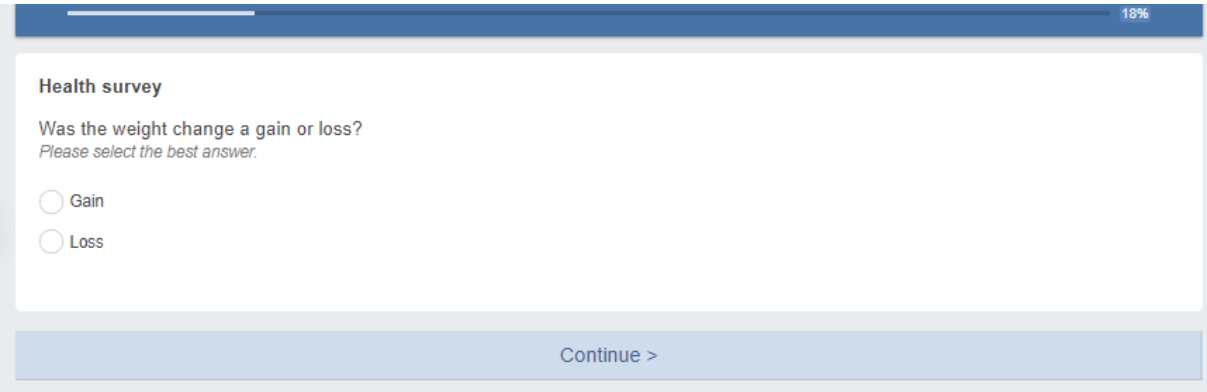
3.3.2 HOW IT WAS TESTED

To test the impact of methodological bias, we compare disclosures from behaviorally-designed questions that include additional options to disclosures from traditional question design in the Control Questionnaire.

For example, a typical weight change question design implemented in the Control Questionnaire asks only about weight gain or weight loss:

Figure 26

CONTROL QUESTION: WEIGHT CHANGE



In the behaviorally-designed question on weight change in Questionnaire B, there is an additional option aimed at those whose weight fluctuates.

Figure 27

QUESTIONNAIRE B QUESTION: WEIGHT CHANGE

Participants who selected 'My weight went up and down' were then asked to provide their maximum and minimum weight over the past 12 months. This information can be used to calculate the relevant weight change for risk assessment purposes, and rules can be designed to accept or refer the case depending on the result.

In the Alcohol question design, the treatment questionnaires asked about frequency of alcohol consumption, while the Control Questionnaire asked only about weekly consumption. Both behaviorally-designed questions allow participants to choose their relevant frequency of alcohol consumption.

Table 2
FREQUENCY OPTIONS FOR BEHAVIORALLY-DESIGNED ALCOHOL QUESTIONS

Control	Questionnaire A	Questionnaire B
Weekly	Every day	Per day
	Every weekend	Per weekend
	1 to 3 times per week	Per week
	1 to 3 times per month	Per month
	On special occasions (Less than 5 times per year)	Per year

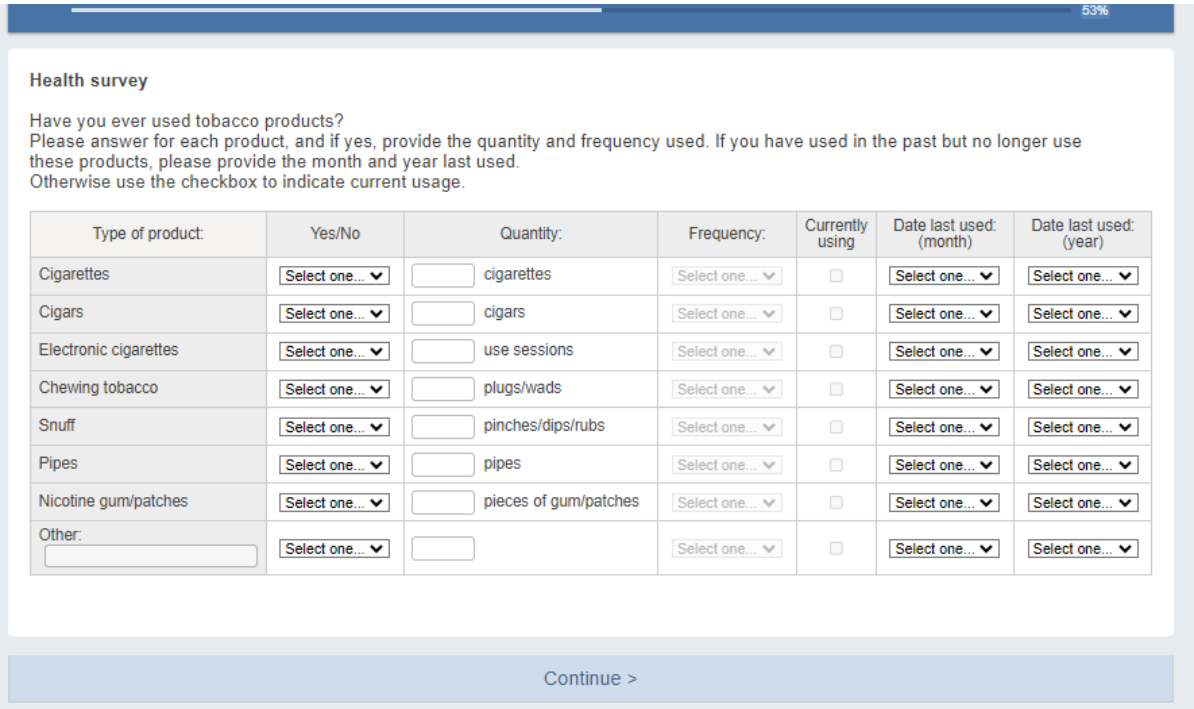
The Tobacco question in Questionnaire A also tests the impact of an additional frequency option to better reflect consumer perspectives. The frequency options in the Tobacco question in Questionnaire A are daily, weekly, or socially, which is defined as less than once per week, compared to the traditional options of daily, weekly, monthly, or yearly included in the Control and Questionnaire B questions.

Another consumer perspective on tobacco use is related to current and former smokers. The Control question reflects the approach often used in application forms whereby those who answer 'yes' to the Tobacco question are presented with a table to complete, which requests frequency, quantity and date last used. Date last used is difficult to complete for a current smoker (today? one hour ago? today's date?) and is, of course, not relevant for underwriting. To avoid confusion about the need to provide this information, both behaviorally-designed questionnaires ask only former tobacco users for the date last used.

In Questionnaire B, the table includes a checkbox for 'currently using' and, if this is selected, the 'date last used' section is grayed out to indicate it is not required. The question intro wording also provides an explanation.

Figure 28

QUESTIONNAIRE B QUESTION: TOBACCO



Health survey

Have you ever used tobacco products?
Please answer for each product, and if yes, provide the quantity and frequency used. If you have used in the past but no longer use these products, please provide the month and year last used.
Otherwise use the checkbox to indicate current usage.

Type of product:	Yes/No	Quantity:	Frequency:	Currently using	Date last used: (month)	Date last used: (year)
Cigarettes	Select one...	cigarettes	Select one...		Select one...	Select one...
Cigars	Select one...	cigars	Select one...		Select one...	Select one...
Electronic cigarettes	Select one...	use sessions	Select one...		Select one...	Select one...
Chewing tobacco	Select one...	plugs/wads	Select one...		Select one...	Select one...
Snuff	Select one...	pinches/dips/rubs	Select one...		Select one...	Select one...
Pipes	Select one...	pipes	Select one...		Select one...	Select one...
Nicotine gum/patches	Select one...	pieces of gum/patches	Select one...		Select one...	Select one...
Other:	Select one...		Select one...		Select one...	Select one...

Continue >

Questionnaire A is fully reflexive and only those who indicate they have ceased smoking will be asked about date last used.

In the results, we can identify the proportion of participants who selected the additional options and consider the overall impact on disclosure of having more relevant options available in each of these questions.

3.3.3 FINDINGS

Weight Change

In each of the three Questionnaire B groups, 20%-24% of respondents selected 'my weight went up and down,' highlighting that this option resonates with a significant proportion of respondents. The proportion of respondents disclosing weight gain or weight loss in Questionnaire B (53%) is broadly in line with the Control group (49%), which suggests that those who report a weight that goes up and down would not have disclosed a weight change in response to the Control question. These results are based on any weight change.

We observe smaller differences between the groups when focusing on a weight change of 10 pounds or more. In the Control group, 44% disclosed a weight change of 10 pounds or more and this increased to 48% for Questionnaire B. This increase is not statistically significant based on our sample sizes. However, Questionnaire B with an Honesty Statement at the Start further increased disclosure of a weight change of 10 pounds or more to 50%, and this is a statistically significant increase.

Including an option for 'my weight goes up and down' significantly increased disclosure of any weight change. The amount of weight change varied depending on the anchors provided in the questions, as discussed in subsection 3.2

on social norms. Incorporating a combination of the techniques based on these concepts could be considered to further improve disclosure.

Given that this experiment identified a significant proportion of respondents (at least 1 in 5) who consider their weight as fluctuating over the year, the concept could also be applied to the weight question (as it was only tested for the Weight Change question in this experiment). Future research could consider giving participants the option of entering a maximum and minimum weight over the past year rather than only requesting one number that assumes a static weight.

Alcohol Use

Overall, alcohol consumption was disclosed by 66% of respondents who received the Control Questionnaire, and this increased to 76% in each of the behaviorally-designed questionnaire groups. All Control group participants provided weekly consumption, but only 23%-24% of the treatment group participants selected weekly given other options. This highlights that other frequencies are more relevant to most respondents.

Table 3
DISCLOSURE OF DAILY USE AND HIGH-RISK ALCOHOL CONSUMTPION

	Control	QA	QB	QB_H_Start	QB_H_End
Daily Drinking	NA	6%	10%	12%	10%
4 or more drinks per day	1.7%	3.9%	8.3%	8.2%	8.0%

In the Questionnaire B groups, 10%-12% of respondents selected 'Per Day' as their relevant alcohol consumption frequency. In Questionnaire A, 6% of respondents selected 'Every Day' in response to the question, 'How often do you drink?' This suggests that more respondents were comfortable disclosing daily drinking in response to a question to select a frequency than the question that asked, 'how often?'.

While the classification of high-risk alcohol use varies among companies within the industry, one possible approach is to define high-risk alcohol use as four or more drinks per day. This definition includes those who disclosed daily drinking and those who disclosed less frequent drinking, but sufficiently high amounts that would be equivalent to four or more drinks per day. For example, those who drink 28 or more drinks per week would be considered high-risk.

Both behaviorally-designed question styles increase the disclosure of high-risk alcohol use, highlighting that identifying more daily drinkers is a key step towards identifying high-risk alcohol consumption. Questionnaire B is more successful at eliciting disclosure of high-risk alcohol use, which may be attributed to the format of the frequency question, and also the format of the amount question as discussed on the topic of anchoring in subsection 3.2 on social norms.

In Questionnaire A, the frequency options also included 'on special occasions' that was defined as 'less than 5 times per year.' This option was selected by 20% of the respondents. This was materially higher than the 7% of Questionnaire B respondents who selected 'per year' as their frequency. In addition, the inclusion of the option 'on special occasions' appears to have reduced the disclosure of other frequencies. For example, Questionnaire A had a lower proportion than Questionnaire B selecting daily use (6% vs 10%) and weekend use (10% vs 16%). It is possible

that the ‘on special occasions’ option is too informal and gives respondents the opportunity to describe how they would like to portray their alcohol consumption rather than their realistic frequency of consumption.

Overall, the inclusion of additional frequency options improved the disclosure of alcohol use and high-risk use, particularly with the approach of Questionnaire B.

Tobacco

Current tobacco use is defined as those currently using tobacco and those who have given it up less than one year ago. Questionnaire A that asked, 'which tobacco products?' had an increased disclosure of current tobacco use (33%) compared to the Control (28%), and Questionnaire B (26%) styles that asked, 'have you ever used?'. This is discussed on the topic of default assumptions in subsection 3.2 on social norms.

Questionnaire A also differed from the Control and Questionnaire B in the frequency options offered. All questionnaires had the options 'daily' and weekly,' but the Control and Questionnaire B also included 'monthly' and 'yearly' as options, while Questionnaire A simply included the option 'socially,' which was defined as less than once per week. Questionnaire A had a slightly higher disclosure of daily tobacco use (20%) compared to the Control (18.5%), but the main difference in disclosure came from those disclosing less frequent than daily use, which was 13% in Questionnaire A compared to 9% in the Control group. The inclusion of the option 'socially' in combination with the question format in Questionnaire A led to an increased disclosure of current tobacco use.

Questionnaire B did not increase disclosure of tobacco use compared to the Control. Questionnaire B applies the approach of asking separately about each tobacco product and requires participants to make an active choice for each product by selecting Yes or No. This is much more onerous than the 'select all that apply' approach used in Questionnaire A to ask separately about each product.

The question style in Questionnaire B also displays all the required information in one large table. This makes it clear to participants that each time they select 'yes' for a tobacco product, they will be required to enter additional information about frequency, amount and dates if they have stopped using tobacco. Since the burden of answering is highlighted, participants identify this question as arduous to complete and try to justify selecting 'no' to some tobacco products to simplify their effort to answer. This is highlighted by looking at the results for disclosure by those who have given up tobacco use more than one year ago. In Questionnaire B, this category includes 9% of respondents compared to 18% of respondents in the Control group. Disclosure of tobacco use by those who have given up more than one year ago is reduced by half due to the style of the Questionnaire B Tobacco question, and disclosure of current use is not increased compared to the Control.

Questionnaire A is an example of a question style that would work much better in an online application form where the functionality is available to allow the participant to select the tobacco products relevant to them and then ask subsequent questions only on the relevant products. Questionnaire B is a format that could be implemented in paper form, which allows asking separately about each product, but highlights the burden of answering by displaying all the additional information that is needed if 'yes' is selected for any product. Online application forms provide more opportunity to implement the behavioral principles identified as most effective in this study.

In the qualitative feedback section, participants were asked which questions they found most difficult to answer. The Tobacco question was ranked as the most difficult to answer by 10% of the Control Questionnaire respondents. Significantly fewer Questionnaire A respondents selected Tobacco as the most difficult to answer question at 3%. This highlights that the reflexive-style Tobacco question in Questionnaire A was a better customer experience that far fewer participants found difficult to answer. This format improved customer experience, as well as increasing disclosure.

3.4 ASK AGAIN

3.4.1 THE CONCEPT

As part of life insurance underwriting, we collect information about a long list of medical conditions. Through a behavioral science lens, the cognitive load associated with a long list of medical conditions is high, making the typical medical conditions question style[2] onerous to understand and answer. Due to the cognitive load, applicants are likely to take mental shortcuts. This can result in skimming and moving quickly through the lists of conditions, automatically clicking 'No' to each list.

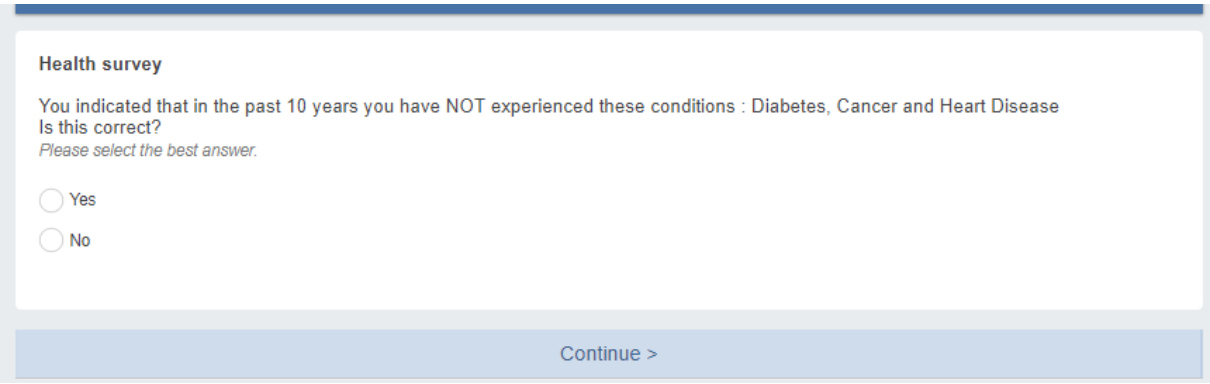
To address this, Questionnaire B includes a confirmatory question that asks a second time about three high-risk and relatively common impairments. The aim is to slow down applicants, to move from automatic thought processes to more deliberate thinking. In addition, presenting a shortened list offers something that is easier to understand and answer than the full list of medical conditions. Finally, the approach to ask again makes it difficult for applicants to justify leaving out the information.

The approach may be viewed as asking more than once about the same impairment, which is contradictory to a desire to ask as few questions as possible and be conscious of the length of the underwriting questionnaire. Therefore, testing in an experimental setting is key to identifying whether the additional disclosure warrants the addition to the questionnaire.

3.4.2 HOW IT WAS TESTED

In Questionnaire B[3], the participants first answer the same question as the Control group for all medical conditions. Then, participants who did not disclose diabetes, heart disease, or cancer are asked to confirm that they have not experienced these conditions:

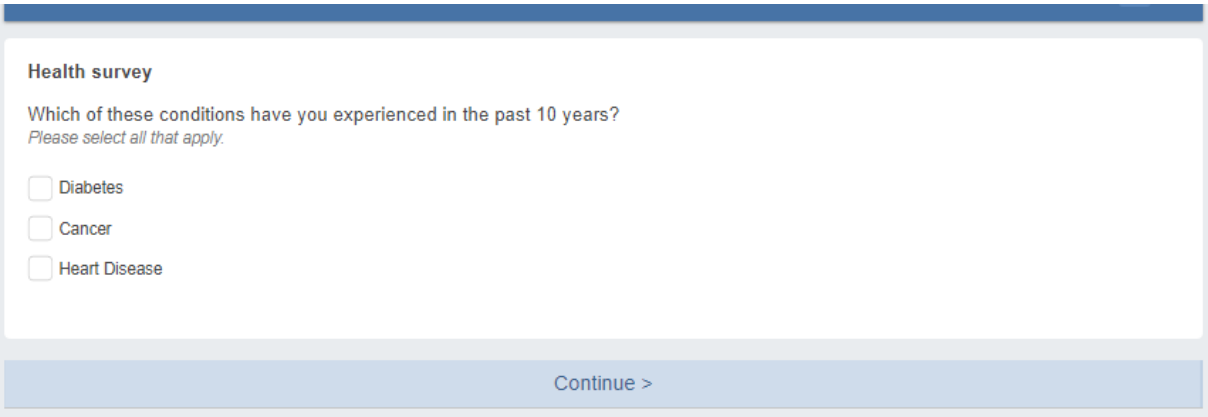
Figure 29
QUESTIONNAIRE B RE-ASK QUESTION



[2] All questions asked can be found in the Tableau dashboard.
[3] Without honesty statement, honesty statement at the beginning and honesty statement at the end.

If they answer No:

Figure 30
QUESTIONNAIRE B RE-ASK FOLLOW-UP QUESTION



With this confirmatory question, this questionnaire enables participants to correct errors, ensuring the accuracy of the information provided.

From an underwriter's perspective, identifying these three impairments is key as they are the leading causes of death worldwide. For instance, in the U.S. in 2021 according to the Centers for Disease Control and Prevention, heart disease and cancer were the two leading causes of death.

Note that this question was shown only when participants had not disclosed the condition. For example, if a participant already disclosed diabetes, they were asked only to confirm that they did not have cancer or heart disease. An online platform allows this flexibility to adapt to applicants' responses to avoid repetition for those who have already disclosed.

3.4.3 FINDINGS

Overall, asking a second time provided a higher disclosure rate for each condition (Table 4). The treatment provided 45% more disclosures of diabetes, 100% more disclosures of heart disease, and 25%-100% more disclosures of cancer.

Table 4

DISCLOSURE COMPARISON ON DIABETES, HEART DISEASE AND CANCER FOR FIRST AND SECOND ASK

Impairment	Control Group	First Ask	Second Ask
Diabetes	QB	11%	16%
	QB Honesty S.	11%	15%
	QB Honesty E.	11%	17%
	QB (All)	**11%**	**16%**
Heart Disease	QB	3%	6%
	QB Honesty S.	3%	7%
	QB Honesty E.	3%	7%
	QB (All)	**3%**	**6%**
Cancer	QB	3%	6%
	QB Honesty S.	8%	10%
	QB Honesty E.	4%	6%
	QB (All)	**5%**	**7%**

Additionally, compared to the Control group and Questionnaire A, asking a second time gathered similar or even higher disclosures for the three impairments considered.

Questionnaire B was asked with an honesty statement at the start for one group and with an honesty statement at the end for another group. For diabetes and heart disease, the results were similar across all Questionnaire B groups, each showing an increased disclosure following the second ask.

The cancer disclosure rate on the first ask was higher for Questionnaire B with the Honesty Statement at the Start than the two other Questionnaire B groups. There was a further increase still in disclosure based on the second ask for Questionnaire B with the Honesty Statement at the Start. This highlights that the treatments are complimentary – an honesty pledge can increase cancer disclosure, but the confirmatory second ask still incrementally increases disclosure.

The impact of asking again seemed to be more pronounced for women of younger ages. For instance, in Questionnaire B, participants who disclosed diabetes on the first ask were, on average, 50 years old, compared to 42 years old for those who disclosed on the second ask. A similar pattern was observed for heart disease and cancer disclosures. This result might motivate insurers to apply this treatment for specific populations.

Looking at participants' qualitative feedback on the survey experience, we found differences between individuals who disclosed medical conditions on the second ask and those who disclosed in response to the first question on medical conditions. Participants who disclosed on the second ask only were less likely to consider the questionnaire "easy to answer." Also, more of these participants admitted to having omitted details during the questionnaire. Of

those who disclosed diabetes on the second ask, 14% admitted that they had omitted details to answer quickly, compared to 7% of those who disclosed diabetes on the first ask[4]. This feedback supports the experiment result showing that asking again provides more accurate responses, controlling non-disclosure behavior coming from individuals trying to answer questions quickly.

3.5 EXPERIENCES

3.5.1 THE CONCEPT

The traditional approach to asking about medical conditions is to provide a long list of medical conditions organized by body system. Applicants are required to select 'Yes' or 'No' to an extensive list of medical conditions. The medical terminology may be seen as jargon from an applicant perspective, and the application form may not always use the same terminology with which the applicant is familiar. Reading the large number of conditions listed creates cognitive load for the applicant, which can lead to skimming as applicants try to move quickly past the difficult section. This was highlighted in the previous section where a shortened confirmation question was shown to elicit additional disclosure.

An alternative to asking about medical conditions is the concept of asking about experiences. The theory is that individuals more readily recall experiences, so questions about taking medication, staying in the hospital, or having surgery are more relatable and easier to recall than exact condition terminology.

The aim is to reduce the long list of medical conditions that need to be asked by instead eliciting disclosures through experience questions. Application forms often include some of these types of questions after the medical condition lists questions. For example, applications often ask about any medications 'apart from what you have already told us.' Rather than using these as catch-all questions for anything missed in the medical conditions lists, the principle tested is to ask the experiences questions first.

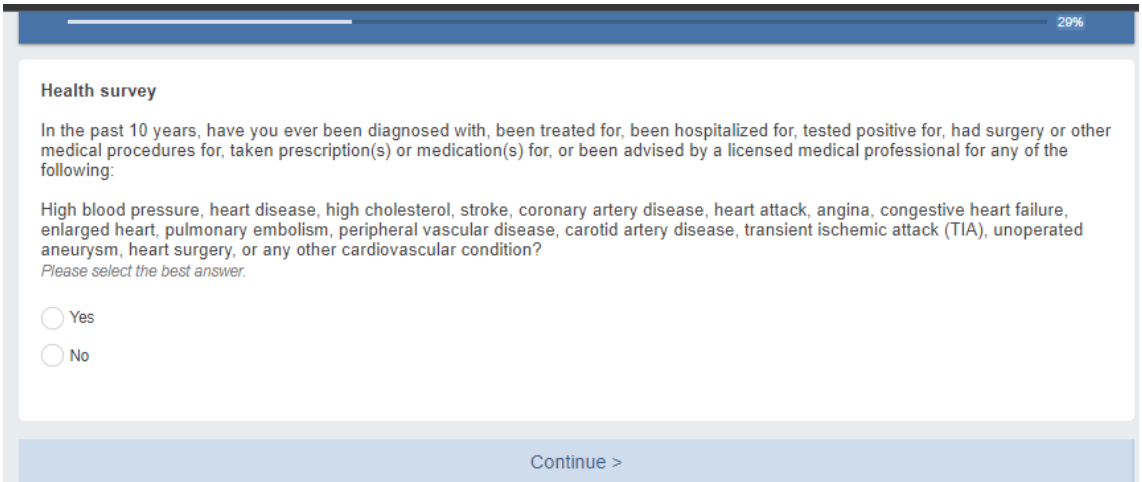
3.5.2 HOW IT WAS TESTED

The Control question uses the traditional approach of asking about medical conditions by providing a long list of conditions organized by body system.

See an example for the cardiovascular conditions:

[4] Note that we are looking here only at Questionnaire B

Figure 31
CONTROL MEDICAL CONDITIONS QUESTION CARDIOVASCULAR



A similar question style was asked for four other body systems covering endocrine, cancer related, respiratory and neurological conditions.

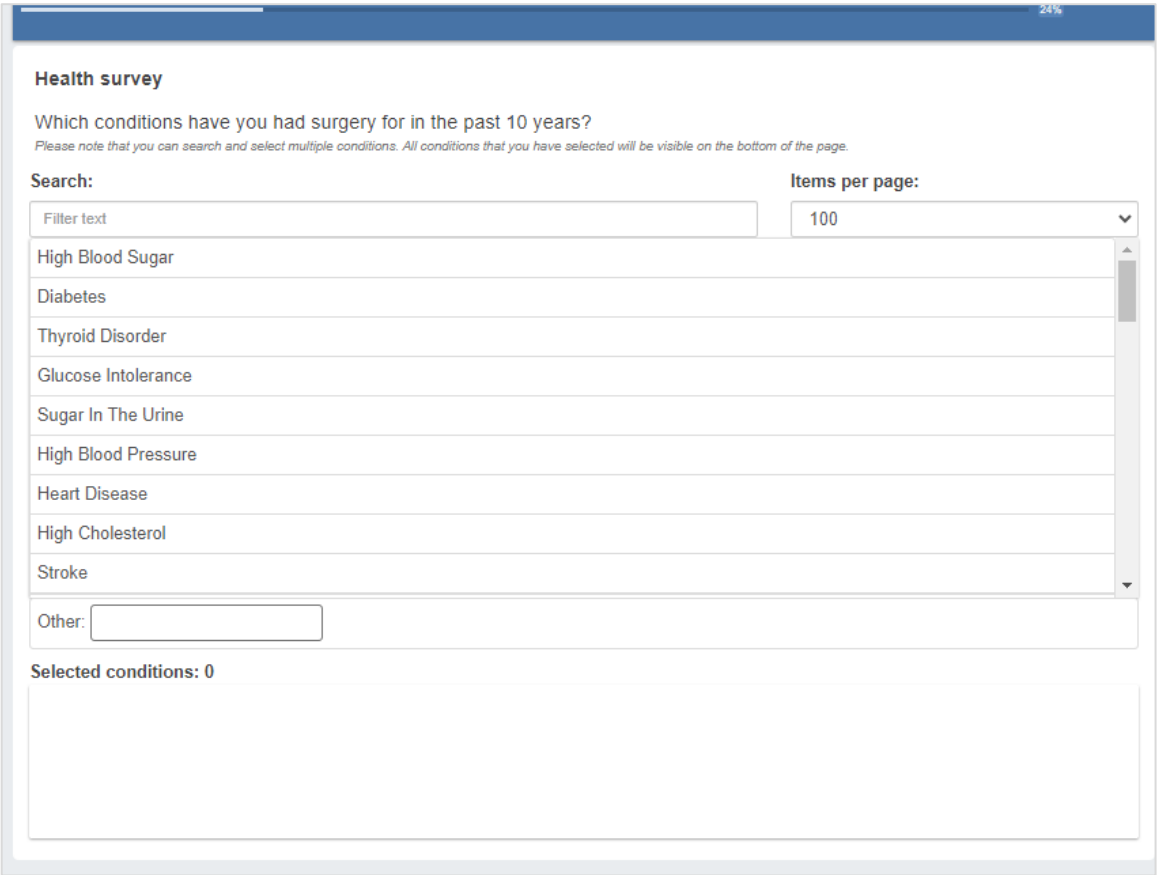
Questionnaire B uses the same question as the Control, but follows up with a confirmation question on three key conditions (as discussed in subsection 3.4). Questionnaire A does not ask about any medical conditions specifically. The medical section includes only three questions asking about medication, hospitalization, and surgery. Respondents are provided with a catalog to select conditions for which they have had any of the three treatments, and the catalog includes all the medical conditions listed in the other questionnaires. In all questionnaires, respondents also have the option to describe their condition using a free text response box.

See an example for the question on surgery:

Figure 32
QUESTIONNAIRE A SURGERY QUESTION

If answered Yes:

Figure 33
SURGERY FOLLOW-UP QUESTION



A similar question was asked for hospitalization and medication.

3.5.3 FINDINGS

More respondents reported at least one condition in response to the questions about medication, hospitalization, and surgery than those presented with lists of medical conditions. Questionnaire A has the highest disclosure of any medical condition with 80% of respondents disclosing at least one medical condition. The increase in disclosure is statistically significant and represents a 51% increase compared to the Control, which had 53% of respondents disclose at least one medical condition. More participants engaged with these questions about experiences compared to questions about medical conditions. Participants more readily recalled experiences of hospitalization, surgery, or taking medication and were willing to share the information.

Many of the additional disclosures for Questionnaire A were for minor conditions that would not impact underwriting. While Questionnaire A showed an increase in disclosure of any condition, there is no statistical difference in disclosure between the groups when considering only the listed medical conditions. The catalog that was available to search in Questionnaire A included a list of medical questions consistent with the conditions listed in the other questionnaires. The increased disclosure was due to respondents entering free text responses in

Questionnaire A. While some of the conditions disclosed in free text would be relevant for underwriting, much of the additional disclosure covered low-risk conditions.

When considering only disclosures of interest including listed conditions and free-text responses deemed relevant based on underwriter review, the level of disclosure for Questionnaire A was similar compared to the Control group, highlighting that disclosure of medical conditions can be elicited using experience-style questions. This varied by condition with some conditions captured through disclosures of surgery, hospitalization, and medication (e.g., cardiovascular conditions including heart disease), while others need to be specifically asked about (e.g., cancer).

Questionnaire A shows a statistically significant decrease in disclosure of cancer-related conditions compared to the Control. Less disclosure of cancer was generated by the experience-style question compared to the medical condition lists. This highlights that cancer is a condition that should be asked specifically as it may not be disclosed in medication, hospitalization, or surgery questions. Questionnaire A shows a similar level of disclosure compared to the Control for cardiovascular conditions. This highlights that similar levels of disclosure of cardiovascular conditions were generated by the experience-style questions as the medical condition lists.

In this study, Questionnaire A tested the concept of fully replacing the long list of medical conditions with experience questions. By going into more detail on the results by specific condition, we identify that a shortened list of medical conditions would need to be asked alongside the experience questions to maximize disclosure overall.

In the qualitative feedback section, participants were asked which questions they found most difficult to answer. The Medical Conditions question was ranked in the top two most difficult to answer for all treatment groups. Significantly more Questionnaire A respondents selected Medical Conditions as the most difficult to answer question at 13% compared to 9% in the Control. Though more participants engaged with the experience-style questions in Questionnaire A, the Medical Conditions question was rated as difficult to answer by a higher proportion of Questionnaire A participants.

A searchable catalog is essential to automating the management of responses to experience-style questions. In this study, the catalog was designed to align with the list of medical conditions provided in other questionnaires. In practice, a more comprehensive list would be valuable from both the customer and the analyst perspectives. Our analysis of the free text responses highlights some common minor conditions that are disclosed that could be included in the catalog to allow them to be automatically accepted as low-risk. In addition, there is an opportunity to include alternative terminology that may be used so that an applicant can find the condition in the format that they are familiar with. Some examples are discussed in subsection 3.7.

While the style of this question shows promising results in terms of engagement with the question, it could be refined by combining with a shortened list of medical conditions and designing with an extensive searchable catalog to handle minor conditions and a range of terminology.

The experience questions are more relatable for respondents than long lists of medical conditions. To achieve the most complete disclosure of medical conditions, a combination of approaches is needed - asking about experiences, asking specifically about a shortened list of medical conditions, including a confirmation question, and having an honesty statement at the start of the application.

3.6 FRAMING HONESTY

3.6.1 THE CONCEPT

An honesty pledge is a statement that informs applicants of their duty to answer honestly and requires them to confirm that they understand and agree. It is typical for insurance applications to include some form of honesty pledge. Most commonly an honesty pledge is found at the end of the application as part of the terms and conditions.

Within behavioral science research, honesty pledges have been studied in other contexts (e.g., cheating on exams). The approach is to test the hypothesis that an honesty pledge seen before completing a task, that requests the participant complete the task honestly, would improve the honesty of responses.

Research into the effectiveness of different forms of honesty pledges highlights the importance of involvement (Drs. Eyal Pe'er, Nina Mazar, Yuval Feldman, and Dan Ariely, 2022). The greater the involvement, defined in terms of the attention and effort required of the participant, the greater the impact of the honesty pledge. Placing the honesty pledge at the start of the policy application and separating it from other terms and conditions is one approach to increase attention. Asking each participant to insert his/her name in the honesty pledge increases effort and also adds identification, a feature Pe'er and coauthors found to improve honesty when used in combination with high involvement.

There has been some controversy surrounding behavioral science studies of honesty pledges (Simonson et al., 2021) with a 2012 study on dishonesty under investigation for potential data fabrication. This creates a gap with respect to research on this topic, which motivated us to test the effectiveness of an honesty pledge in this study.

The wording of the honesty statement may also leverage behavioral science principles to improve effectiveness. Loss aversion, the principle that humans tend to fear losses more than they seek gains, can be invoked by focusing on the loss that applicants could face if they are dishonest. In insurance applications, the honesty pledge can highlight the risk of losing out on benefits at the time the beneficiary (usually the applicant's family) needs the financial support of the insurance product.

While there is an issue with non-disclosure in insurance applications, most applicants are honest, and highlighting this fact can improve honesty, following the principle that most humans want to conform to social norms. Wording could also focus on morality, calling on an applicant to think about others and their personal values when considering being dishonest. In the context of insurance applications, accurate information is needed to ensure that the cost of insurance is fair for all.

There is a full spectrum of non-disclosure ranging from unintentional to fraudulent non-disclosure. The behavioral science principles used elsewhere in this experiment will mainly address non-disclosure that is unintentional, or intentional but when the applicant can justify to themselves that the information was not requested clearly or is not relevant. The inclusion of the honesty pledge at the start of the application attempts to address intentional non-disclosure.

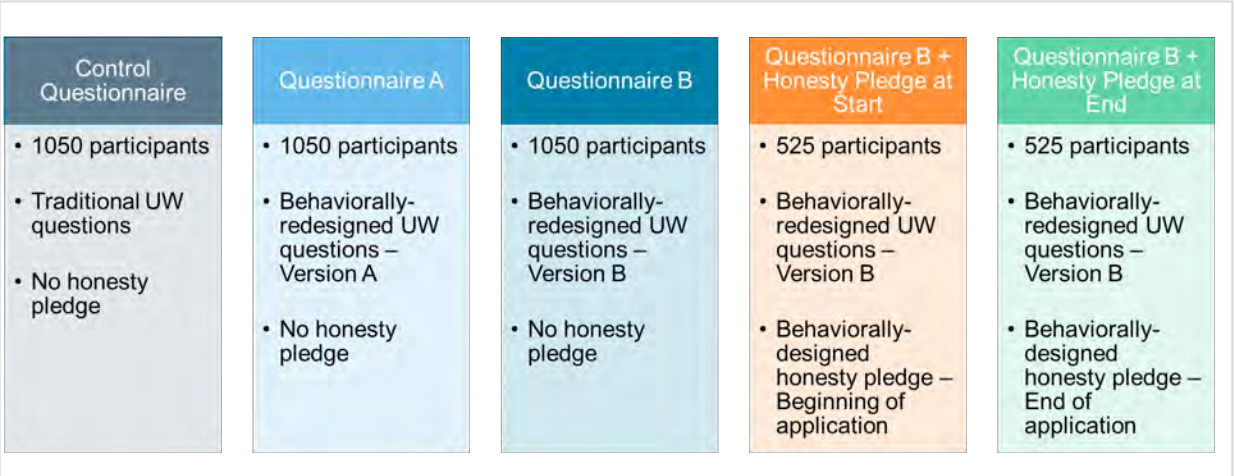
3.6.2 HOW IT WAS TESTED

In this study, we have tested the inclusion of an honesty pledge compared to not having an honesty pledge alongside a behaviorally-designed application questionnaire. Further, we have tested the placement of the honesty pledge by testing two alternatives. In one treatment questionnaire, the honesty pledge is placed at the start with

the aim to improve disclosures in line with the behavioral science research. In another treatment, the same honesty pledge is placed at the end of the questionnaire and this approach is designed to be reflective of current practice in the insurance industry.

Figure 34

EXPERIMENT DESIGN INCLUDING HONESTY PLEDGE GROUPS



For the testing of the honesty pledge, three of these groups are relevant – Questionnaire B without an honesty pledge, Questionnaire B with the honesty pledge at the start and Questionnaire B with the honesty pledge at the end. Note that the treatment groups with an honesty pledge have a smaller number of participants than the other groups. Given practical limitations that prevented adding two additional groups of 1050 participants each, we chose to have two smaller groups, which allowed us to test the placement of the honesty pledge and maximize the insights we can gain for a given number of participants, with half of the participants in each group.

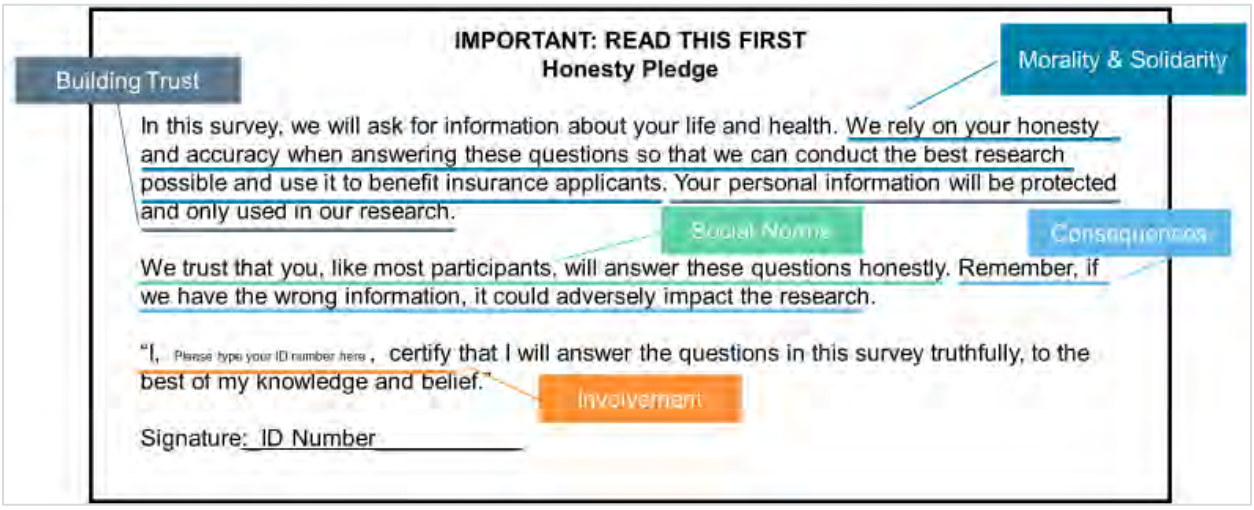
Only Questionnaire B was tested with the honesty pledge. Therefore, we cannot make any comment about Questionnaire A and the potential impact of including an honesty pledge with this questionnaire.

Both honesty pledges include the same wording, leveraging principles from behavioral science. The honesty pledge is worded to be relevant to the study participants rather than insurance applicants. Therefore, the principles rather than the specific wording of this honesty pledge will be useful to those designing an honesty pledge for an insurance application.

It was not possible to test the principle of loss aversion as the participants in this study do not face the same potential loss as an insurance applicant. The principle of consequences is applied to the context of the survey highlighting that inaccurate answers could adversely impact the research. Similarly, we invoke solidarity but, in the survey context, highlight that honest answers are required to ensure the research will benefit insurance applicants.

Building trust by addressing data privacy concerns is applied in a manner that could easily be made relevant for an honesty pledge for insurance applications. Similarly, the social norms message highlighting that most participants answer honestly could be applied to an insurance application honesty pledge.

Figure 35
HONESTY PLEDGE WORDING – EXAMPLE SHOWN AT START



Both honesty pledges ask the participant to enter the first name in the pledge and then sign to acknowledge agreement. Due to differences in placement, participants agreed to slightly different statements. When the honesty pledge appeared at the start, applicants acknowledged that they would answer honestly. When the honesty pledge appeared at the end, applicants acknowledged that they had answered honestly. If, upon seeing the honesty pledge at the end, an applicant wished to improve the honesty of his responses, he would need to go back and change his answers.

3.6.3 FINDINGS

The results of this study show that the inclusion of an honesty statement at the start of the application can improve disclosure. The impact of the honesty pledge is seen for some questions only and varies depending on the topic. There is no evidence of an adverse impact by including an honesty pledge at the start. Across all questions, Questionnaire B with an Honesty Pledge at the Start had similar or higher disclosure compared to Questionnaire B (the same questionnaire without an honesty pledge).

Regarding the mental health question, the honesty pledge included at the start of the application had a statistically significant impact on disclosures. The behaviorally-redesigned mental health question in Questionnaire B applied the concept of building trust, including a message about the confidential treatment of sensitive personal data. The mental health disclosures did not materially increase for this group compared to the Control group. However, the group who saw Questionnaire B with an honesty pledge at the start had a statistically significant increase in mental health disclosures compared to the Control group.

Similar behavioral concepts were tested for the substance use support question, where Questionnaire B included a message to build trust around the confidential treatment of sensitive personal data. This did not increase disclosure relative to the Control and, for this question, the group who answered Questionnaire B with an honesty pledge at the start also did not have an increased disclosure rate compared to the Control. So, while the disclosure of mental health conditions was higher when an honesty statement was shown at the start, the disclosure of substance use support was not increased.

The honesty pledge at the start increased disclosure for some, but not all, medical conditions. When looking at diabetes, cancer and heart disease, there was an increase in disclosures of cancer in the group who saw the honesty statement at the start, but no impact for the other two conditions. The increase in disclosure for cancer was significant, with 10% of the group who viewed Questionnaire B with the honesty pledge at the start disclosing cancer compared to 6% in each of the other Questionnaire B groups, including those who did not see an honesty pledge and those who saw the honesty pledge at the end.

The honesty pledge also increased disclosure of extreme BMI values in response to the Weight question. In analyzing the results for the Weight question, differences between the experiment groups were only statistically significant when looking at the high-risk BMI categories. For the analysis, high-risk obesity was defined as BMIs of greater than or equal to 40. Questionnaire B with no honesty statement and Questionnaire B with the honesty pledge at the end each had 4.8% disclosure of high-risk obesity, while Questionnaire B with the honesty statement at the start had 7.6% disclosure of BMIs of 40 or more. This is a 58% increase in disclosure and a statistically significant increase compared to the Control, highlighting that the honesty pledge increased disclosure of high-risk obesity.

The inclusion of the honesty pledge at the end of the application form had no impact on disclosure. Throughout the analysis, we saw very similar results between the group who saw an honesty pledge at the end of the form and those who saw no honesty pledge. This treatment group was included in the study to reflect that most insurance application forms include a statement at the end, often as part of a long list of terms and conditions, where the applicant must certify that they have provided accurate and honest information.

In all questionnaires, participants were given the opportunity to go back to review and edit their answers. Across all groups, very few participants took this option – ranging from 0.8% -1.2%. The group who saw the honesty statement at the end did not have a higher take-up of this option compared to other groups. Placing an honesty statement at the end has limited impact because participants are not likely to go back to review answers upon reaching the end of the application form and viewing an honesty statement at this stage does not increase the likelihood.

The evidence suggests it is optimal to include the honesty pledge at the start of the application form and that the inclusion of a behaviorally-designed honesty pledge increases disclosure for specific elements of the application form that are subject to more intentional non-disclosure.

3.7 THE ANALYST PERSPECTIVE

Throughout this study, the goal was to design an application process with the applicant perspective in mind – a form that would be easy and straightforward for an applicant to complete.

However, this focus does not mean forgetting about those on the other side of the process – underwriters and others who must analyze the responses gathered in the application form and use them to come to a decision about insurance coverage. Underwriters want an application form that produces clear, structured, usable information. Many insurance companies also emphasize improving straight-through processing rates, so that more applications can be processed without the need for an underwriter to manually review the form or seek additional information.

Fortunately, the two perspectives need not be in conflict. There are ways to design the application form to produce a good experience for both the applicant and the analyst. The following suggestions provide a starting point for this design.

#1 – Minimize the Use of Open-Response Boxes

Open-response boxes – answer boxes that allow an applicant to enter any information they choose, in their own words – are frustrating for both applicants and analysts. For an applicant, an open-response box can be daunting, leaving them with little direction about how much information they should provide.

For an analyst, open-response boxes make straight-through processing difficult since advanced text-mining techniques are needed to feed free-text responses through an automated process. Even if insurance companies are comfortable with the fact that open-response boxes will require extra review, that review can be far from simple. Open-response boxes can lead to answers containing misspellings, alternate or incorrect terms, or confusing or missing information. The applicant's response can also be subjective and difficult to place with established underwriting guidelines.

For instance, in the soft launch of this study, applicants in the Control group were asked to describe their frequency of tobacco usage in an open-response box. This question was ultimately changed in the final version of the study as the responses from the soft launch were so confusing and inconsistent as to be almost impossible to analyze. For instance, one respondent simply wrote "2." Unfortunately, many application forms currently in use contain many instances of open-response boxes, reducing the effectiveness of the form on multiple levels.

To improve the application for both applicants and analysts, open-response boxes should be avoided as much as possible. As an alternative, depending on the question, designers of the application may wish to use one of the other answer designs tested in this study.

Slider bars or closed-end options can be effective answer formats for questions asking about frequencies or amounts (such as Weight Change, Alcohol, or Tobacco questions). As well as introducing social norms (see subsection 3.5 above), these answer formats offer structure and guidance to the applicant and allow for automatic processing of the responses. However, it is important to carefully consider any closed-end options offered to ensure they cover all possible responses and do not unduly influence applicants' answers.

Another good option, specifically for Medical Conditions questions, is to replace open-response boxes with a searchable catalog of medical conditions. As explained in subsection 3.4 above, respondents in this study were comfortable engaging with a searchable catalog.

However, the lists of medical conditions in such a catalog must be considered carefully to ensure that they include as many relevant conditions as possible. For instance, the catalog used in this study did not include many common, mild medical conditions such as broken bones, joint replacement surgeries, and everyday illnesses – and this led many respondents to write their answers into the open-response box instead of using the catalog. This will be the outcome whenever medical conditions are missing from a catalog, undermining the effectiveness of this format at improving STP rates.

Catalogs should also be as "smart" as possible, accounting for differences in spelling and common alternative terms. The easier the catalog is for the applicant to use, the more likely it is that applicants will use it – and use it correctly – preventing the need for a manual review of their answers by the underwriter.

#2 – Consider Well-Designed Error Messages

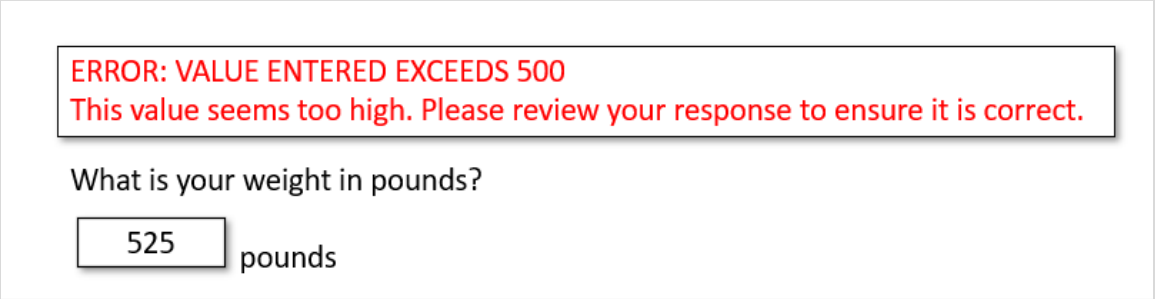
At several points in an application form, applicants must provide numerical answers – their height/weight, their amount of alcohol or tobacco consumption, and so on. In these cases, there is the possibility that, due to confusion or a lack of attention, applicants will enter answers that cannot be true – BMIs that are impossibly high or low, for instance.

However, the line between truly impossible answers and merely very extreme ones is blurry and hard to define. Each insurance company will likely define it differently, and values which are near the line will probably always require some amount of extra review to determine if they make sense in the context of the entire application.

To minimize the need for this review, error messages can be introduced into the application form to prompt applicants who enter suspicious values to review their own responses. Behavioral science principles can be applied to optimize the design of these error messages.

For instance, if an insurance company feels that weight values of over 500 pounds are suspicious, they might wish to build into their application a message that asks anyone who enters a value of more than 500 on the Weight question to review their response:

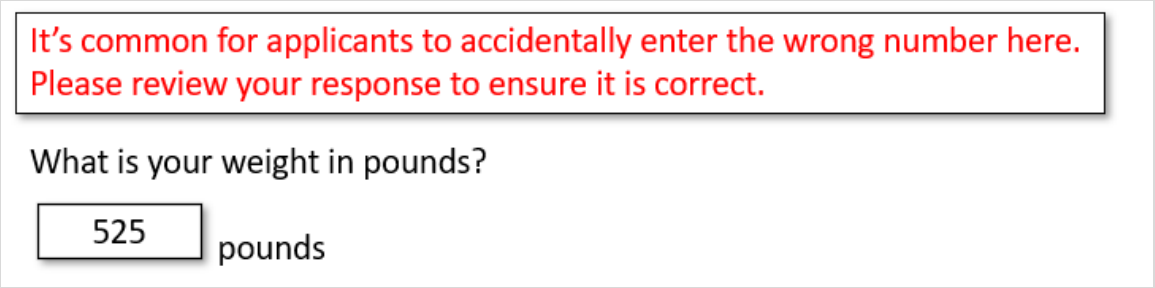
Figure 36
NEGATIVE ERROR MESSAGE



However, the wording of this error messages may seem threatening or accusatory and could make the applicant defensive and uneasy. It also suggests directly that the value entered is too high. This could lead the applicant to feel self-conscious and change their answer, even if it was originally correct. The reference to the value (500) that triggered the message highlights for the applicant what the "wrong" answer is and could likewise lead them to change their answer to a new value below the reference.

An alternative design for an error message would be the following:

Figure 37
BEHAVIORALLY-DESIGNED ERROR MESSAGE



This message is friendly and doesn't single the applicant out as causing a problem, making the review seem like part of a routine check requested of everyone. There is no indication that the value entered is too high or any mention of the limit that triggered the message.

A well-designed error message like this can help minimize impossible answers, while still allowing for the collection of real, extreme values.

#3 – Map Options to Rules

As described in subsection 3.5, closed-end options can be used in questions asking about amounts (such as Weight Change, Alcohol, or Tobacco questions) to leverage social norms and put applicants at ease. However, to make the process easy for analysts as well, the content of these closed-end options needs to be considered carefully in the context of any rules or criteria about risk levels that the insurance company wishes to use.

We discovered this insight during the analysis of this study, when the closed-end answer options we had used for alcohol consumption in one of the questionnaires caused difficulties with efforts to define a measure of heavy drinkers. The options were designed to align with the customer perspective and ranges were selected to be appropriate to how individuals consider alcohol consumption for different frequencies. Aggregating across different frequencies to define a single measure of heavy drinking was challenging because we had not considered the risk criteria in the design of the options. The learnings from this experience highlight the importance of considering both customer and analyst perspectives in the design of closed-end answer options.

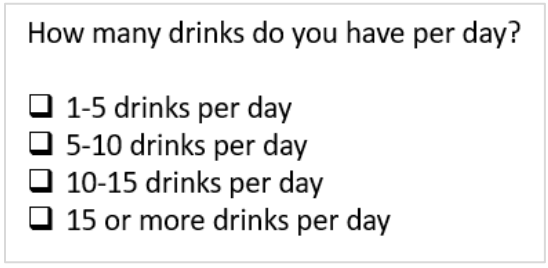
An extended example will be helpful here. Suppose that an insurance company would like to collect information about alcohol consumption using closed-end options. Applicants will be presented with a series of ranges of alcohol amounts (i.e., "x to y drinks per day"). To determine exactly what these ranges should be, the insurance company should consider how they want to use the information they gather on the form.

Suppose the insurance company plans to use the following rules in underwriting applications:

1. Any applicant who drinks less than two drinks per day will be considered low-risk.
2. Any applicant who drinks two to four drinks per day will be considered normal-risk.
3. Any applicant who drinks more than four drinks per day will be considered high-risk.

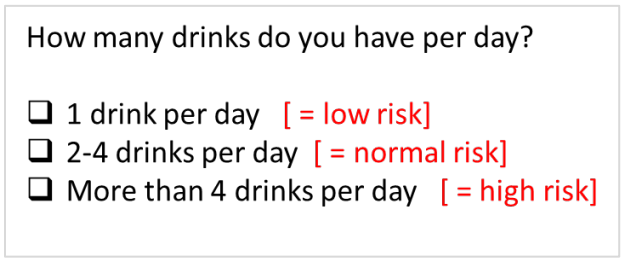
In this case, it is important that the ranges provided in the closed-end answer options correspond to these rules and allow applicants to be easily categorized as low, normal, or high-risk, whatever they answer. For instance, the following ranges for answer options would not make sense:

Figure 38
STARTING DESIGN FOR "PER DAY" ANSWER OPTIONS



With these answer options, an applicant who chose "1-5 drinks per day" might be low, normal, or high-risk. The answer options are not specific enough to map applicants accurately to the risk criteria. An improvement on these answer options would be the following:

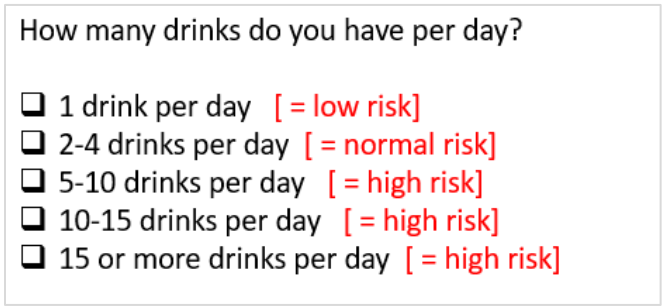
Figure 39
ANALYST PERSPECTIVE ONLY DESIGN FOR "PER DAY" ANSWER OPTIONS



These answer options allow applicants to be easily placed into risk categories, as shown in the red text. However, this is still not an optimal set of answer choices. As described in subsection 3.2, due to a desire to avoid choosing extreme options, applicants are less likely to choose the highest or even second-highest answer options provided. Although the above categories would be sufficient from an analyst perspective, they would likely lead to non-disclosure of high-risk alcohol consumption as applicants avoid choosing the highest range.

Combining the applicant and the analyst perspective leads to the optimal design for these answer options:

Figure 40
AN OPTIMAL DESIGN FOR "PER DAY" ANSWER OPTIONS



Although the final two answer options lead to the same risk categorization as the third and are technically unnecessary from an analyst perspective, they serve to help applicants feel comfortable disclosing high levels of alcohol consumption. Since all these answer options can be tied to an automated rules engine, there is no cost to the analyst of adding the extra options.

#4 – Combine the Applicant and Analyst Perspectives

These examples of specific design features share a common theme: the best application form designs consider both the applicant and the analyst perspectives. The desires and requirements of applicants and analysts need not be in conflict – indeed, they are often completely compatible. It is possible to design an application form that is both understandable and straightforward for applicants and useful and clear for analysts.

Behavioral science principles can be used to incorporate and harmonize these perspectives, improving the application process for everyone involved.

Section 4: Interactive Dashboard Presentation

4.1 MAIN FEATURES

This report is supported by an interactive dashboard that presents the detailed results of the experiment. The dashboard can be accessed at this link:

The dashboard provides the results of the analysis for each question:

1. Height/Weight
2. Weight Change
3. Mental Health
4. Medical Conditions
5. Additional Medical Conditions
6. Alcohol
7. Tobacco
8. Substance Use Support

For each question, a key variable of interest is identified, and the disclosure rates are compared across the five experiment groups. Using the interactive features of the dashboard, it is possible to review these results by demographic characteristics including:

- Age
- Gender
- Employment status
- Education level
- Income group

Further, the analysis investigates other variables of interest to measure disclosures that would be most meaningful to the underwriting decision.

For each question, it is possible to view the exact wording used in each of the questionnaires.

Section 5: Conclusion

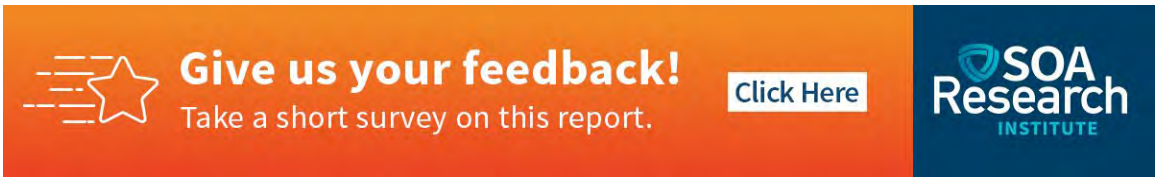
Traditionally, life insurance applications have been designed from an underwriting and regulatory perspective. Behavioral science techniques provide ways to rethink the design from the applicant perspective. Taking the applicant perspective into account, we identify the opportunity to include different options and use reflexive-style questioning to enhance customer experience and capture additional information as a result. Considering both the applicant and the analyst perspectives strives to achieve the optimal design that increases disclosure, improves customer experience, and maintains or improves processing ease for the insurance company.

Insights discussed in this report will educate on the behavioral science techniques that can be applied to question design. Results of the experiment offer clear evidence of increases in disclosure achieved by specific question designs. The experiment provides evidence that reducing social stigma by separating out conditions or behaviors of different severity improves disclosure. Implementing social norms through the removal of Yes/No questions that highlight negative behaviors and framing the questions to assume the behavior exists was shown to improve disclosure. The study highlights the power of anchoring and the potential for carefully designed slider bars and closed-option response style questions to improve disclosure of high-risk behaviors. The experiment demonstrates that long lists of medical conditions are often subject to skimming and that a confirmation question improves disclosure of important medical conditions. Participants engaged with questions about experiences such as surgery, medication and hospital stays suggesting that these could be asked first to reduce the number of medical conditions that need to be specified in lists. A behaviorally-designed honesty pledge placed at the start of the questionnaire improves disclosure on topics that are most impacted by intentional non-disclosure.

The experiment framework used in this study is explained in detail and offers a new tool to companies wishing to test further the impact of behaviorally-designed questions. As a next step, further testing is encouraged both in experimental setting and in real-world implementation. This study was carried out in an experimental setting, which allowed testing of several different behavioral science techniques on different questions. The findings highlight the techniques that were most effective at increasing disclosure, as well as generate new ideas for potential future testing.

The disadvantage of the experimental setting is that the insurance applicant perspective is not fully captured. Since the experiment participants were not applying for life insurance, their incentives were not aligned to those of life insurance applicants. Life insurance applicants have an incentive to non-disclose with the aim of keeping their insurance premium low, but also an incentive to disclose due to the risk of invalidating the insurance contract if information is later discovered. This unique perspective would be best captured through field experimentation – implementing the new question design on life insurance application forms and monitoring the impact on disclosure.

In the industry study, we learned that underwriting is a key area where companies are applying behavioral science or plan to in the coming year. We also found that the lack of expertise and concrete evidence were the biggest difficulties for companies when it comes to applying behavioral science. This study addresses both concerns, by providing a framework for testing the impact of question re-design on disclosure rates and demonstrating the potential increase in disclosure through the strong results.



Section 6: Acknowledgement

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group and others for their diligent work overseeing questionnaire development, analyzing and discussing respondent answers, and reviewing and editing this report for accuracy and relevance.

Project Oversight Group members:

Shaikh Mujtaba Ali, MBA

Greg Fann, FSA, MAAA, FCA

Anthony Ferraro, FSA, MAAA

Bruce Friedland, FSA, MAAA

Syed Muhammad Razi Hasnain, ASA, ACIA

Winston He, FSA, MAAA

Blake Hill, FSA, FCIA

Ri Hao Kang, FSA, FCIA

Jahanzeb Meghani, FSA, CERA, FCIA

Zack Moore, FSA, MAAA, CERA

Kevin Pledge, FSA, FIA

Renee West, FSA, MAAA

Tina Yang, FSA, MAAA, CERA

Yifan Zhang, FSA

At the Society of Actuaries Research Institute:

Korrel Crawford, Senior Research Administrator

Donovan Dotson, Market Research Analyst

Dale Hall, FSA, MAAA, CERA, Managing Director of Research

Ritesh Patel, DevOps Specialist

Emily Rafferty, Director of Business Insights

David Schraub, FSA, MAAA, CERA, Senior Practice Research Actuary

Patrick Wiese, ASA, Modeling Researcher

The researchers would also like to express their appreciation for their colleagues at SCOR who contributed to the success of this study:

Andrew Hallsworth, FSA, MAAA

Bruce V. Miller, MBA, PMP, A-CSM, CSPO, CAL

Stacy Stauffer, FALU, FLMI, FFSI

Alexandra Stultz, PhD

Niamh Uí Cheallacháin, FSAI

Appendix A: Methodological and Statistical Considerations

A.1 PROCEDURE FOR DATA CORRECTIONS

In all cases, as much as possible, the baseline analysis was done on the data as received from Dynata. This is in recognition of the fact that any data correction involves some number of assumptions about which responses are "mistakes" and what the corrected response should be. In case these assumptions are incorrect, it's usually best to minimize the amount of data manipulation that's conducted and to rely on the responses received as is. However, data corrections can be used for sensitivity analysis to determine if making specific corrections changes the results found.

There are three main types of data corrections that we made or considered making:

1. Recoding Participants with Contradictory Answers
2. Removing Outliers
3. Eliminating Participants with Questionable Response Patterns

A.1.1 RECODING PARTICIPANTS WITH CONTRADICTORY ANSWERS

This type of data correction was indicated for questions such as the Mental Health question or the other Medical Conditions questions. In these questions, a certain number of participants answered "Yes" to the overall "Yes/No" question ("Do you have any of these mental health conditions?). Then, when presented with the list of conditions included in that section, these participants selected the "Other" option and wrote "None" in the open response box. It seems possible that these participants made a mistake in clicking "Yes" on the overall "Yes/No" question, then attempted to "correct" their mistake by reporting "None" on the list of specific conditions.

Another data issue in this category is from the Weight Change question, in which a small number of participants reported "My weight went up and down," then reported Maximum and Minimum weights that showed no weight change.

Finally, this data issue shows up in the Tobacco question, where some participants reported that they had used a certain tobacco product, then entered their quantity used as "0."

The counts of how many participants displayed this type of data issue are listed below. Note that some differences between groups are due to differences in question formats (for instance, QA did not have a separate question about Endocrine conditions, so there was no opportunity for this data issue in QA).

Table 5

COUNTS OF PARTICIPANTS WITH CONTRADICTORY ANSWERS

	Control	QA	QB	QB + H_Start	QB + H_End	Total
Weight Change	0	0	2	0	1	3
Mental Health	1	0	9	6	7	23
Endocrine	3	0	9	2	1	15
Cardiovascular	1	0	3	2	0	6
Respiratory	0	0	1	0	0	1
Nervous	1	0	4	1	0	6
Surgery	0	1	0	0	0	1
Hospitalization	0	1	0	0	0	1
Medication	0	6	0	0	0	6
Other Medical Conditions	2	28	22	18	11	81
Quantity of Cigarettes	1	0	0	0	0	1
Quantity of Chewing Tobacco	1	0	0	0	0	1
Quantity of Snuff	1	0	0	0	0	1
Quantity of Pipes	1	0	0	0	0	1
Quantity of Nicotine Gum	1	0	0	0	0	1
Quantity of Nicotine Patches	1	0	0	0	0	1
Type of Other Tobacco Product Used	2	0	0	0	0	2
Total # of Participants with These Data Issues	**12**	**35**	**48**	**29**	**21**	**145***

* This number is not the sum of the totals above it as it does not double-count participants who have this data issue for more than one question.

The procedure for correcting these data issues is simply to recode the responses of these participants to match what we assume to be their intended answers. For instance, if a participant said “Yes” to Overall Mental Health, then wrote “None,” their response would be recoded as “No” to the Overall Mental Health question. The analysis in question can then be rerun to check if the recoding has changed any results.

After carrying out this procedure, we found that the recoding did not affect any of our main results. Therefore, we continued to use the raw data received from Dynata in most of our main analysis. As discussed above, this was because we believed it was important to avoid making excessive assumptions about the “true” answers intended by the respondents and to instead use their responses as given.

The exception to this was in the Other Medical Conditions question, where those who did not select “No other conditions,” but then wrote “None,” were counted as reporting no other medical conditions.

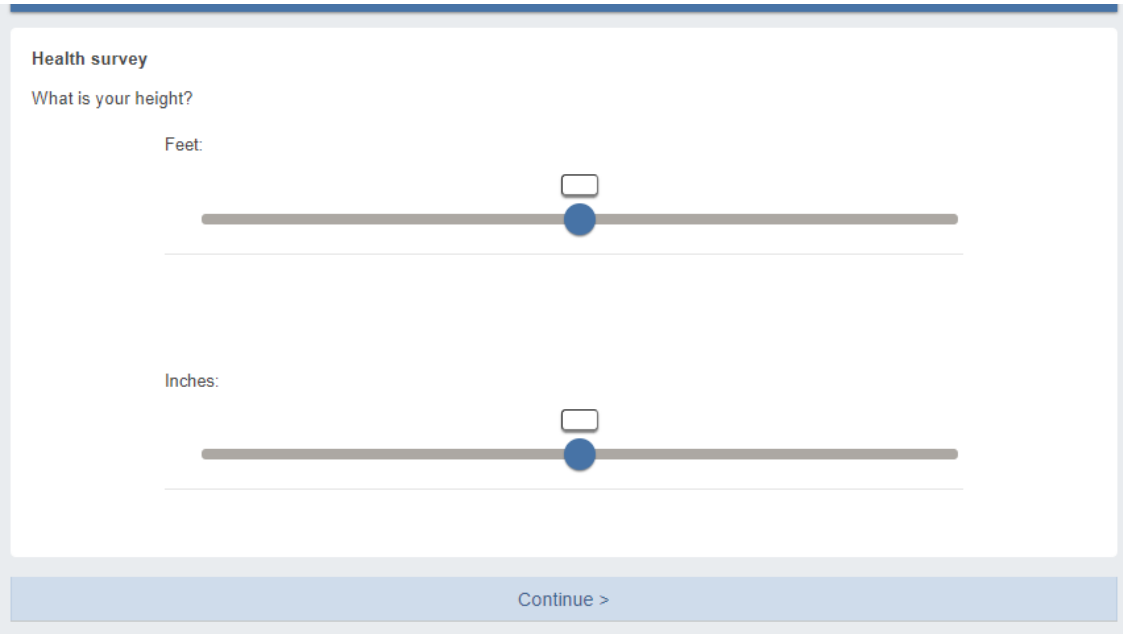
A.1.2 REMOVING OUTLIERS

This data issue primarily relates to the Height/Weight question. Some participants entered heights and/or weights that seemed unbelievably high or low. These outliers could have potentially skewed our results, so we needed some criterion of identifying them and checking to see if excluding them changed our results.

We approached this issue in two stages; first, we identified and excluded responses that we believed to almost certainly be errors, then we excluded the top and bottom extremes of the BMI distribution to remove outliers.

Questionnaire A incorporated slider bars that respondents used to enter their height and weight.

Figure 41
QUESTIONNAIRE A SLIDER BAR EXAMPLE HEIGHT QUESTION



However, the slider bar would occasionally "jump" to the extreme end of the height/weight range if clicked accidentally. To account for this issue, we decided to exclude responses where the height or weight entered was equal to the maximum height or weight that was allowed to be entered (for instance, 8ft or 575lbs). We believe that these responses likely represent errors or mistakes caused by glitches in the survey itself.

The second step in handling this data issue was to exclude responses with BMIs that were unreasonably high or low. We set these limits of reasonable BMIs intentionally wide to avoid accidentally excluding genuine, extreme results. We decided to exclude responses with a BMI of less than 12 or greater than 100, following the example of the Centers for Disease Control and Prevention (CDC) in its analysis of its Behavioral Risk Factor Surveillance System (2021).

The counts of how many respondents are excluded under each of these two steps are listed below:

Table 6
COUNTS OF PARTICIPANTS EXCLUDED UNDER OUTLIER CRITERIA

Exclusion Criteria	Control	QA	QB	QB + Honesty_Start	QB + Honesty_End	Total

Min/Max	11	16	4	8	3	42
<12/>100	23	16	11	10	5	65

A.1.3 ELIMINATING PARTICIPANTS WITH QUESTIONABLE RESPONSE PATTERNS

This type of data correction is arguably the most subjective and difficult to implement. It involves looking at participants' holistically and identifying any who have an overall pattern of responses that raises suspicion that they may have been answering randomly or providing nonsense information.

For instance, one participant reported they had used every possible tobacco product (including "Other"), but entered 0 as the quantity for all of them. Another participant reported a weight under 100lbs, answered "Yes" to every category of medical conditions, and reported having used every possible tobacco product. In isolation, each of these data issues is worrying or slightly suspicious but, together, they call into question the quality of the entire set of responses for these participants.

To correct for this type of data issue, a list of "problem respondents" was compiled to identify respondents with unusual or suspicious patterns of responses. We then excluded these respondents from the data set to determine if doing so changed any results. No results were materially impacted by this exclusion. Therefore, in the final analysis, data from all respondents is included (except where otherwise noted).

Appendix B: Summary Statistics of Participants

See the Demographics tab in the Study Overview page in the interactive dashboard.

References

Centers for Disease Control and Prevention. (2024, January 17). *Leading causes of death.* https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Centers for Disease Control and Prevention. (2021, December 7). *Definitions*. Definitions | Data, Trends and Maps | CDC

Chong, K., & Chua Y. J. (2021). Dissecting dishonesty using behavioural economics. *SAwatch Newsletter*, 1-9. https://www.scor.com/sites/default/files/2021-12/Dissecting_Dishonesty_Using_Behavioural_Economics__Final_.pdf.

Christodoulou, D., Samuell, D., Slonim, R. & Tausch, F. (2022). Counteracting dishonesty strategies: A field experiment in life insurance underwriting. *Journal of Behavioral Decision Making*, 36(2). https://doi.org/10.1002/bdm.2302

Lyons, C. (2022, September). Misrepresentation: Is the truth out there? *SCORacle*, 2-9.

Pe'er, E., Mazar, N., Feldman, Y., & Ariely, D. (2022, September 8). Honesty pledges to reduce dishonesty: Involvement and identification [Conference session]. 1st International Behavioural Public Policy Conference, London School of Economics, London, United Kingdom.

Scribani, M., Shelton, J., Chapel, D., Krupa, N., Wyckoff, L., & Jenkins, P. (2014). Comparison of bias resulting from two methods of self-reporting height and weight: A validation study. *Journal of the Royal Society of Medicine*, 5(6). https://doi.org/10.1177/2042533313514048.

Simonson, U., Nelson, L., & Simmons, J. (2021, August 17). Evidence of fraud in an influential field experiment about dishonesty. Data Colada. https://datacolada.org/98.

World Health Organization. (2023, October 20). *Mental health of older adults*. https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults.

These third-party links are being provided for informational purposes only. The content in these third-party links do not necessarily reflect the opinions of Society of Actuaries Research Institute, Society of Actuaries, or the Education & Research Section. Neither the Society of Actuaries Research Institute, Society of Actuaries, nor the Education & Research Section are responsible for the reliability, accuracy or content of the third-party site(s). If you have questions about the content on such sites, please contact the site directly.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
8770 W Bryn Mawr Ave, Suite 1000
Chicago, Illinois 60631
www.SOA.org