Incentives, Distributions, and the Accuracy of Subjective Health Risks

Ben Mosier IV*

February 18, 2025

Abstract

Beliefs about personal health risks shape health decisions, and researchers have long studied these beliefs through surveys. However, none have tested whether monetary incentives increase thoughtful or truthful responses. This paper tests whether such incentives improve the accuracy of elicited beliefs about personal health risks. I also evaluate the role of confidence in health expectations by comparing two response modes - point estimates, and complete belief distributions - in a 2×2 treatment design. In an online sample of U.S. adults, I elicit subjects' beliefs about their risk of common chronic health conditions, which I compare to highly personalized statistical estimates. Monetary incentives reduce error size by 4.8 percentage points when subjects report complete belief distributions, but have no effect on point estimates. Importantly, beliefs more strongly predict preventive health behaviors when they exhibit high confidence, suggesting that eliciting both risk perceptions and confidence may improve our understanding of health decisions.

Keywords: Incentives, Belief Distributions, Belief Elicitation, Health Expectations

JEL: C9, D8, I12

*Georgia State University, Atlanta, GA Email: rmosier2@gsu.edu

1 Introduction

To make optimal health decisions, individuals require an understanding of their own health risks. Biased beliefs about disease risk can lead to inefficient health investments, causing individuals to incur unnecessary costs or underinvest in preventive care. Despite a high national death toll from preventable chronic disease (Garcia, 2019), fewer than 10% of American adults receive all high-priority preventive services recommended to them (Borsky et al., 2018). Understanding discrepancies like this one requires studying not only objective health risks but also how individuals assess their risks themselves. Research has shown that the measuring of subjective health expectations provides unique insights beyond what can be learned from objective risk calculations alone (Sloan, 2024).

For over 30 years, researchers have measured beliefs about health risks through surveys, showing these beliefs help explain important economic decisions like retirement, savings, and insurance choices. However, existing data exhibit recurring patterns - bunching of reports at 0, 50, and 100%, large overestimation of rare events, and other irregularities suggesting that current methods lead to measurement error (Hudomiet et al., 2023). While monetary incentives are widely used in experimental economics to address such issues and motivate truth-telling, they have not yet been applied to improving the accuracy of health belief measurements.

This paper makes three main contributions. First, it provides the first experimental test of whether incentive-compatible monetary rewards improve the accuracy of elicited beliefs about personal health risks. Second, it evaluates whether eliciting complete belief distributions, rather than point estimates, allows for better measurement by capturing both bias and confidence. Third, it reveals that confidence moderates the power of beliefs to predict preventive health behaviors - a finding with important implications for both research methodology and health policy.

Using an online sample of 490 American adults, I implement a 2×2 experimental design varying both monetary incentives and response formats. After collecting detailed health profiles, I estimate each participant's health risks for a set of chronic diseases using machine learning and clinical risk calculators. I then elicit their subjective beliefs about these same health risks, comparing these to the personalized objective estimates. Subjects in incentivized treatments are scored for accuracy using a binarized quadratic scoring rule adapted from Harrison et al. (2017), where subjects report their subjective beliefs. Half of subjects report full belief distributions, communicating their confidence by allocating 100 tokens across the probability support, while the remainder allocate only one token, reporting a point estimate in a manner similar to much of the preexisting work eliciting probabilistic health expectations.

The results reveal several key patterns. First, monetary incentives improve the accuracy of reported beliefs, but only when subjects can communicate complete distributions - reducing mean absolute error by 4.8 percentage points. When limited to point estimates, incentives show no significant effect on accuracy. Second, subjects systematically overestimate their risk of low-probability health outcomes, particularly for conditions like stroke and skin cancer where objective risks are small. Third, I find beliefs more strongly predict preventive health behaviors when they exhibit high confidence, suggesting that the ability to measure confidence through distributional responses provides valuable information for understanding health decisions.

The paper proceeds as follows. Section 2 reviews related research on the accuracy of health expectations and introduces the framework for incentivized belief elicitation. Section 3 details the experimental design, including methodological solutions for three challenges: estimating personalized health risks, implementing incentive-compatible scoring rules, and detecting online searching. Section 4 presents the main results on accuracy and effort across treatments, and examines the relationship between beliefs, confidence, and preventive behaviors. Section 5 concludes and discusses implications for survey design and health policy.

2 Background

2.1 Eliciting Health Expectations

A substantial body of work explores the relationship between subjective beliefs and health behaviors. Since the 1960's, researchers in public health and health psychology have elicited mostly qualitative assessments of subjective risk, aligning with various psychological models of behavior (Rosenstock, 1974; Rogers, 1975). While these measures have been shown to correlate with health behaviors (Carpenter, 2010), qualitative responses prohibit the evaluation of beliefs for numerical accuracy and present substantial challenges with regard to interpersonal comparability. Numerical subjective probabilities can also be implemented directly into life cycle models, allowing economists to evaluate their influence on decisions like retirement, savings, and life insurance enrollment.

Health economists began studying probabilistic measures of health beliefs primarily following the introduction of mortality expectations in the Health and Retirement Study (HRS). Mortality expectations of this kind have been shown to play a role in many major economic decisions. Smith et al. (2001) found these beliefs to be predictive of mortality outcomes, even when objective health risks are controlled for, and that beliefs shift appropriately as respondents' health status changes over time. Researchers have also evaluated the accuracy of

HRS mortality expectations, documenting a consistent "flatness bias," in which respondents tend to underestimate their probability of living to younger ages, but overestimate their chances of survival to ages beyond 80 (Perozek, 2008; Elder, 2013).

Research on the accuracy of health expectations has since expanded to other population surveys and health topics, including chronic disease. Khwaja et al. (2009) compare subjective probabilities of developing lung disease, heart disease, and stroke by age 75 to objective estimates derived from HRS data, finding that individuals tend to overestimate these risks. Carman and Kooreman (2014) find that women overestimate their probability of breast cancer in the next five years, where the mean subjective probability is 19.1%, over 20 times higher than the objective rate of 0.9%. Similarly for heart disease, their subjects report a mean probability of 16.4%, while the true probability of new diagnosis is 6.4%.

This quantitative economic research on health expectations has established a number of empirical regularities. Individuals appear to be largely insufficiently informed about population-level probabilities, with substantial dispersion of aggregate beliefs (Sloan, 2024). Despite this aggregate uncertainty, there appears to be no systematic under- or over-estimation of beliefs, except in the case of an observed center-biased tendency for subjects to report probabilities closer to 0.5 when the true answer lies particularly close to 0 or 1 (Hudomiet et al., 2018; Carman and Kooreman, 2014; Giustinelli et al., 2022). Subjective beliefs about health have also importantly been shown to contain private information that is typically uncaptured by public health surveys. This information enables health expectations to provide meaningful power for improving predictions of future health outcomes (Hurd and McGarry, 2002).

In addition to patterns that may be explained by underlying beliefs, several features are present in health expectations data that prevailing methods lead subjects to misreport, report distorted probabilities, or employ limited effort. In the HRS, 30 percent of subjects report an equal chance of living to ages 75 and 85, about one-third of which reported their probability of survival as precisely 1 (Hurd and McGarry, 1995). Many studies also document substantial bunching of reports at 0, 50, and 100 percent, as well as a considerable degree of rounding (Manski and Molinari, 2010; Bruine De Bruin and Carman, 2012). Overestimation of low probability outcomes has also been shown to occur to implausible degrees in some cases (Fischhoff et al., 2000). Hudomiet et al. (2023) also decompose the flatness bias in mortality expectations from the HRS, arguing that the bias can be explained by a general center-bias in elicited probabilities, rather than a true property of beliefs about survival.

These results beg the question as to whether the measurement of health expectations can be improved via alternative belief elicitation methods. Experimental economists have developed a host of such methods with the intention of improving data quality by aligning

incentives with truth-telling and ensuring sufficient subject effort. Work in this area has focused on constructing theoretically-sound and behaviorally incentive-compatible mechanisms that subjects can understand. Experimenters have also developed response modes and payoff mechanisms to elicit full belief distributions, allowing for the evaluation of both accuracy and confidence of beliefs. This paper seeks to introduce these established experimental methods to research in health expectations and to evaluate their efficacy in this domain.

2.2 Incentives

Experimental economists have long argued that monetary incentives are essential for aligning subjects' behavior with experimental objectives and ensuring adequate effort. Smith (1982) establishes key precepts for maintaining experimental control, emphasizing that motivation for desired behaviors (in this case, truth-telling) must be salient, and subjects must be non-satiated in rewards. This requires monetary incentives be large enough that rewards for accurate reporting meaningfully influence subjects' utility and dominate competing motives to misreport or exert minimal effort. Without this feature, other motives may outcompete the induced preferences an experimenter hopes to test.

Although monetary incentives are standard practice in experimental economics, they have not yet been applied to elicit beliefs about personal health risks, despite clear potential for competing motives. Subjects may choose not to apply effort to cognitively demanding tasks in an uncertain environment like health outcomes. High risk subjects may also be inclined to report lower probabilities to justify past health behaviors or avoid negative emotions (Festinger, 1962). Respondents may even misrepresent their beliefs out of social desirability concerns (Zizzo, 2010). To overcome these competing motives, sufficiently-scaled incentive-compatible monetary rewards can create a salient and dominant driver for thoughtful and truthful reporting.

Tests for the effects of incentives on belief elicitation outside of health have yielded mixed results. Some studies find incentive-compatible mechanisms improve belief performance in games (Gächter and Renner, 2010; Wang, 2011). Harrison (2014) documents a variety of hypothetical biases related to beliefs about economic indicators and national-level health outcomes. A lack of incentives has also been shown to amplify the frequency of bunched responses (Burfurd and Wilkening, 2022). Trautmann and van de Kuilen (2015) find that unincentivized belief measurements in a two-player game yield similar accuracy compared to several incentivized revealed preference mechanisms. However, they also find that incentivized methods are better predictors of participants' own behavior compared to introspection. Danz et al. (2022) find that providing detailed information to subjects on the binarized quadratic

scoring rule (BSR) increases center-biased deviations from the truth, and that subjects perform better when they are simply told that truth-telling is in their best interest.[1]

Incentivizing accurate reports of personal health beliefs presents distinct methodological challenges. Unlike laboratory experiments where outcomes can be immediately verified, health outcomes unfold over long time horizons. This leaves researchers with two options: either follow subjects longitudinally to observe and reward based on realized outcomes, or develop reliable estimates of personalized health risks that account for subjects' private information. Previous work implementing incentivized belief elicitation in health has largely focused on full population-level statistics that could be immediately verified (Di Girolamo et al., 2015; Harrison et al., 2022). This paper advances the literature by implementing a model that generates highly personalized risk estimates conditioned on detailed individual health profiles, enabling incentive-compatible elicitation of beliefs about personal health risks.

2.3 Subjective Belief Distributions

Eliciting complete belief distributions may also provide substantial value for health researchers compared to capturing only point estimates. From a data quality perspective, Engelberg et al. (2009) show that when a point forecast and belief distribution are elicited from the same macroeconomic forecasters, more than 20% of point forecasts show no correspondence with the mean, median, or mode from the distributional forecast for the same outcome. Bruine De Bruin and Carman (2012) also show that bunching of reports at 50% is explained in part by subjects with a lack of confidence in their beliefs. By eliciting complete distributions, this circumvents any inconsistent approaches that participants may use to reduce their beliefs to a singular summary statistic, allowing the researcher to do so in a consistent manner.

Differentiating between beliefs of high and low confidence is also of considerable normative importance for understanding health behaviors. If subjects with biased beliefs report wider distributions (i.e. have lower confidence), this is normatively very different than if they are biased but submit degenerate reports (maximum confidence). High confidence and high bias could lead to suboptimal health decisions, while low confidence and high bias may simply

[1] Danz et al. (2022) provide results only for the elicitation of point estimates. As Danz et al. (2024) also state, while their results "make clear that something in the binarized and quadratic-scoring rule is malfunctioning, it is not clear what." In the case of eliciting full distributions - a key experimental feature of this paper, as detailed in Section 2.3 - a simple request for truth-telling has more ambiguous implications for how subjects should best respond. Without a consistent means to weigh the reporting of low or high confidence, subjects may interpret what constitutes a "true" belief distribution in a myriad of ways, limiting interpersonal comparability. Further, theoretical results from Harrison et al. (2017) find that distortionary effects of risk aversion should manifest as "flattening" of reported distributions, rather than shifts to center as highlighted in Danz et al. (2022). As such, I default to clearly informing subjects of their potential rewards in incentivized belief elicitation tasks to ensure greater consistency in the interpretation of incentive mechanisms across subjects.

result in individuals seeking more information. In doing so, they may reduce their bias as well.

Differentiating between beliefs of high and low confidence is of considerable normative importance for understanding health behaviors. If subjects with biased beliefs report wider distributions (i.e., have lower confidence), this is normatively very different than if they are biased but submit degenerate reports (maximum confidence). High confidence and high bias could lead to suboptimal health decisions, while low confidence and high bias may simply result in individuals seeking more information. Despite these clear benefits, relatively few studies capture measures of confidence or imprecision in health beliefs. Several have examined methods for eliciting probability ranges, where participants convey their confidence by providing a point estimate along with a probability interval denoting minimum and maximum values (Delavande et al., 2024). Others implement a response format where subjects first report a point estimate and then subsequent probing questions allow for the optional report of an interval (Giustinelli et al., 2022). Results from these elicitation methods have demonstrated that many subjects do hold imprecise subjective beliefs, and importantly, that this imprecision matters. For instance, Kerwin and Pandey (2023) shows that individuals with more imprecise beliefs about HIV transmission have a greater propensity to update their beliefs in response to new information. However, such approaches present methodological challenges when attempting to generate summary statistics of individual or aggregate beliefs, or to incentivize subjects for accuracy. Elicitation of a complete distribution, as in this paper, allows for consistent incentivization and analysis of belief accuracy.

2.4 Proper Scoring Rules

In order to properly motivate truth-telling in belief elicitation, monetary incentives require the implementation of a scoring rule. Beyond concerns of competing motives like cognitive complexity and effort level, some scoring rules incentivize subjects to distort their reports on account of risk aversion. Researchers thus place a focus on "proper" scoring rules, where truth-telling is incentive-compatible. Among the most popular of these mechanisms is the Quadratic Scoring Rule (QSR), where subjects are penalized according to the squared error of their answer relative to a realized outcome (Brier, 1950; Matheson and Winkler, 1976). As outlined in Harrison et al. (2017), the QSR can be modified to elicit belief distributions over a continuous support. By partitioning the domain over which subjects can report into K intervals and denoting r_k the report of the likelihood that the event falls in interval

$k = 1, 2, ..., K$, if the true value of lies in interval k, then the QSR payoff is defined as:

$$S = \alpha + \beta \left[(2 \times r_k) - \sum_{i=1}^{K} (r_i)^2 \right], \tag{1}$$

where α and β are parameters that can be set to scale payoffs as desired.

If subjects are risk neutral, this formulation of the QSR is a truth-revealing mechanism. However, as proven by Harrison et al. (2017), risk-averse agents will maximize their subjective expected utility by reporting a "flattened" belief, such that a maximally risk-averse agent would report a uniform distribution across the support of their latent belief distribution. Hossain and Okui (2013) show that a binary lottery procedure paired with the QSR results in a theoretically incentive compatible scoring rule independent of risk-preference, assuming subjects maximize subjective expected utility. This procedure requires that subjects are paid in probabilities rather than dollars, where probabilities correspond to the chances of earning the larger of two monetary prizes. Several studies have also shown that the inclusion of a binary lottery procedure improves performance of the QSR in a lab setting (Hossain and Okui, 2013; Harrison et al., 2013; Erkal et al., 2020). I operationalize these findings in my experiment by binarizing QSR payoffs in incentivized treatments.

3 Experimental Design

I construct a 2×2 treatment design, varying the use of distributional reporting and monetary incentives for accuracy between subjects. This treatment structure allows me to identify whether incentives and distributional reporting are complements that enhance each other's effectiveness, or whether one feature alone is sufficient to effectively measure health expectations. The treatment naming structure is shown below in Table 1.

	Point Estimates (1 Token)	**Distributions** (100 Tokens)
No Incentives	N1	N100
Incentives	I1	I100

Table 1: 2×2 Treatment Structure

3.1 Subject Pool and Recruitment

I recruit subjects from Prolific, limiting the sample to American adults aged 35 to 74. Table 2 displays demographic characteristics of the sample. Of note, subjects were predomi-

nantly non-Hispanic white (73 percent) and well-educated (52 percent hold a bachelor's degree or higher). 82 percent of subjects report having at least one chronic disease, predominantly high blood pressure, high cholesterol, arthritis, and diabetes. These subjects were experienced online study participants; the median number of approved Prolific study submissions is 703. Using Prolific's device restriction settings, subjects were also allowed to participate only from a computer, and could not access the experiment from a phone or tablet.

In initial experimental sessions, 46 subjects were each paid $12 for their participation, independent of monetary incentives for accuracy in incentivized treatments. The median completion time for the study was lower than anticipated at 18 minutes, resulting in a participation payment rate of $40/hr. To align with the upper bound of Prolific's recommended payment rate of $8 to $16/hr and to increase remaining sample size, the participation payment for the remaining 444 subjects was reduced to $8, while incentives for accuracy were unchanged. Subjects did not learn their treatment condition in advance; only the participation payment was shown to them before they agreed to participate. Zhong et al. (2024) report that variation in offered participation payments of similar magnitude had little impact on participation rates.

3.2 Incentives

Payoffs were determined according to a Binarized Quadratic Scoring Rule for discrete belief distributions (detailed in Section 2.4) adapted from Harrison et al. (2017). In simple terms, subjects reported their beliefs by allocating tokens to potential outcomes. A subject's token allocation and the correct answer (their personalized risk estimate) together determined the number of points they earn for each task. One task was selected at random for payment, and the points earned in that task dictated the percentage chance the subject earned the larger of two monetary prizes.

To report beliefs, subjects allocated "tokens" across 10 "bins." Bins correspond to a discrete set of equal-sized intervals spanning the probability support from 0 to 1. By allowing subjects to report belief distributions, subjects maximize subjective expected utility by reporting their true latent belief. The choice of ten bins balanced two competing objectives: allowing subjects to report distributions with sufficient precision, while keeping the interface simple enough for intuitive allocation of tokens. In addition to the well-documented rounding to the nearest multiple of 5 or 10 typically observed in elicitation of probabilistic expectations, Delavande et al. (2011) also present results that probabilistic expectations are robust to variations in elicitation design, including the choice of support intervals.

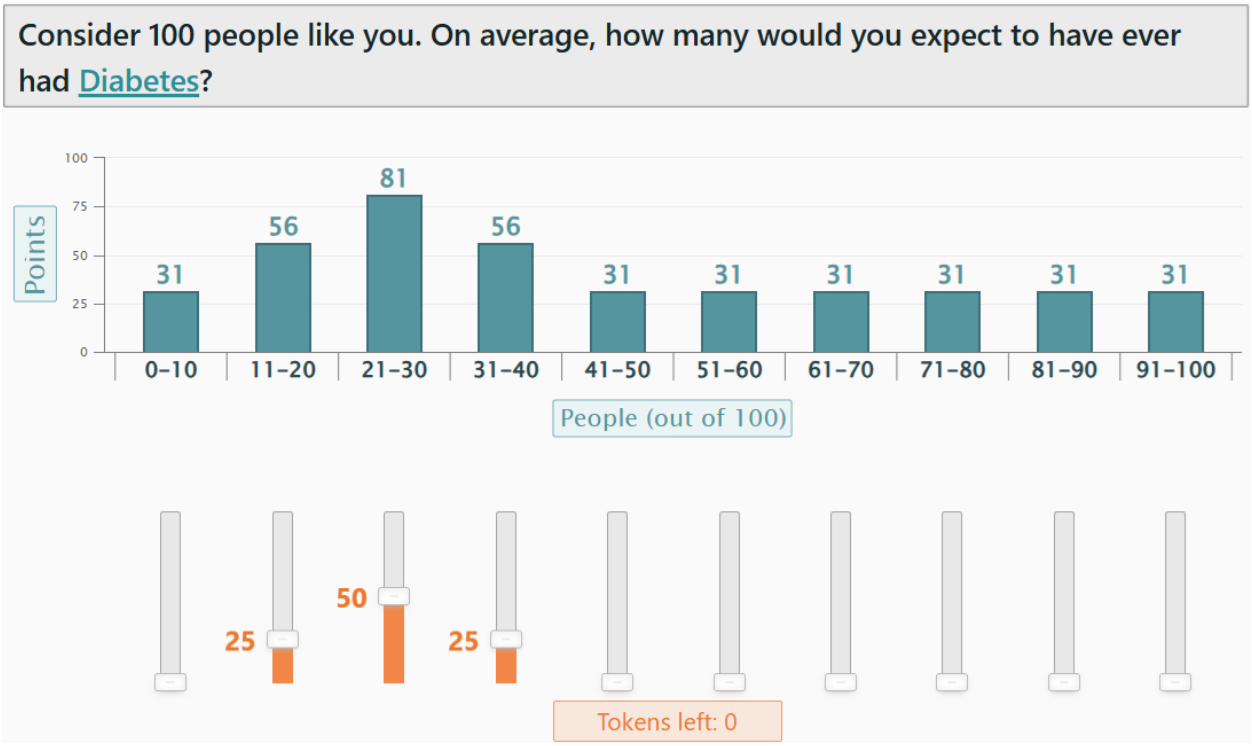


Figure 1: Task Interface with Example Token Allocation

As payoffs are binarized, subjects can earn two potential monetary prizes - \$5 or \$25. Instead of converting QSR scores to dollars directly, subjects earn points, p. By setting both α and β to 50, points can range from 0 to 100. For payment, a random number, q, is drawn from 0 to 100 on a uniform distribution. If $p <= q$, subjects earn \$25, or \$5 if $p > q$. By doing so, a subject's QSR score is equal to their percentage chance of earning the larger prize. If a subject is certain in their answer, they can allocate all 100 tokens to one bin. This will ensure they earn the larger prize if the correct answer falls in this bin, and the smaller prize otherwise. If a subject is totally uncertain, they can allocate 10 tokens to each bin and receive a guaranteed number of points, independent of the correct answer. After all tasks were completed, the score from one task was selected at random for payment.

3.3 Response Modes

The key distinction between response modes was the number of tokens available to subjects. In distributional treatments, subjects received 100 tokens to allocate across bins, while subjects in point estimate treatments received only one token to place in a single bin. Figure 1 shows the task interface for a distributional treatment, which included a bar chart displaying potential points. This chart also dynamically updated as subjects adjusted their allocations. In point estimate treatments, the implemented formulation of the QSR

is inherently simplified, such that a subject receives the larger prize for allocating their token to the correct bin, and the smaller prize otherwise. The scoring chart was thus not displayed for point estimate treatments as it did not provide informative value to subjects. After completing all tasks and the post-task questionnaire, subjects were shown their token allocation, the correct bin, and their final score for each task.

3.4 Pre-Task Questionnaire

After consenting to participate, subjects completed a pre-task survey to provide a detailed health and demographic profile used for personalized risk estimation. First, subjects read the definition of each health condition of interest, and reported their overall level of familiarity with the causes, risk factors, and symptoms of the condition. Subjects then answered a set of health and demographic questions that match those used in the National Health and Nutrition Examination Survey (NHANES) administered by the US Centers for Disease Control. Predictor variables include information relating to: age, sex, height, weight, gender, diet, alcohol use, smoking, physical activity, race, education, income, marital status, insurance coverage, healthcare utilization, condition diagnoses, and family health history. Later, subjects were also informed that when a question referred to someone "like them," this meant that the correct answer was personalized to them based on their own survey responses, and relied on statistical estimates. Figure 2 shows a screenshot of this page in the subject instructions.

3.5 Task Types

Subjects completed 10 health belief elicitation tasks, following two formats. Tasks either elicited a subject's belief about the *prevalence* of a condition, or about the chance of developing a condition *in the next 10 years*. All quantities were presented in natural frequency terms (e.g., "X out of 100 people"), which have been shown to improve subject comprehension compared to probability formats (Gigerenzer and Hoffrage, 1995). Health condition definitions, as specified for subjects, are listed in Appendix Section A.[2]

[2]Beliefs in each task type can be interpreted in slightly different ways. Ten-year risk assessments relate directly to the probability of a subject's future health outcomes, making them more relevant for preventive health decisions. At the same time, future risk assessments may lead subjects to consider eventual changes in health status, income, medical advancements, and other influencing factors, as argued by Perozek (2008). Prevalence tasks, by contrast, focus on assessing the current health status of a population matching the subject's profile. This approach avoids uncertainties tied to future projections but may be less applicable to forward-looking health decisions. Additionally, because prevalence estimates are based on cross-sectional data, where outcomes and predictors are measured simultaneously, they do not account for variations in past health behaviors that might shape current disease status. In both cases, with incentives tied to statistical estimates rather than realized outcomes, these tasks primarily assess health literacy through a personalized

3.5.1 Prevalence

Tasks eliciting beliefs about the prevalence of a condition provided the following prompt:

> "Consider 100 people like you. On average, how many would you expect to have ever had _____?"

Each subject reported their belief about the prevalence of 8 chronic conditions - heart disease, diabetes, stroke, high blood pressure, high cholesterol, arthritis, skin cancer, and all other cancers (excluding skin cancer). Prevalence in this setting is defined as the proportion of a given subpopulation to have ever been diagnosed with one of these conditions. This format allows for consistency with the question formats in the population survey data used to estimate risks (detailed in Section 3.5.1 and Section B), where survey respondents report if a doctor ever diagnosed them with a given condition.

Prevalence was estimated using a machine learning ensemble prediction algorithm, adapting the approach of Einav et al. (2018). The model was trained on data from the NHANES years 1999 to 2018, where survey items mirror the question format from the experimental pre-task questionnaire. The NHANES dataset was chosen for its rich set of health predictor variables, especially relating to family history of diabetes and heart disease. Family history is particularly important to include as a predictor, as medical studies have established family history as a critical risk factor, and other work has demonstrated subjects have greater distrust in tailored health risk estimates if calculators fail to include sufficient personal and family history (Scherer et al., 2013).

The complete algorithm is an ensembling of three models: random forest, gradient boosting, and LASSO. Combining multiple models enhances flexibility and robustness, making it well-suited for application to multiple health outcomes, as in this study. This approach captures nonlinearities, interactions, and other unique data features that a standard linear probability model may overlook. To train each model and avoid overfitting, I subset the data into a "train" sample, used to develop the algorithm, and a "test" sample, withheld from training, to validate model performance. Appendix Section B details the training steps employed for the algorithm, and shows that the training and ensembling process produces reliability and performance comparable to other work using machine learning to estimate health risks. This model also performs well on out-of-sample prediction when used with other common health datasets and, importantly, among the recruited subject pool.

framework rather than a strict individual risk analysis. While the interpretive nuances vary, I show that risk estimation methods perform well and treatment effects are consistent across task types.

3.5.2 10-Year Risk

10-year risk tasks elicited beliefs about the probability of developing a condition in the future. This prompt read:

> "Consider 100 people like you who have not had _____. On average, how many would you expect to develop _____ in the next 10 years?"

These beliefs were elicited for the probability of developing diabetes and breast cancer. Any subjects with a prior diagnosis of diabetes or breast cancer were excluded from 10-year risk tasks, and males were excluded from the breast cancer task entirely.

10-year risks were estimated using standard clinical risk tools. For breast cancer risk, I employ the National Cancer Institute's Breast Cancer Risk Assessment Tool (BCRAT), also known as the Gail Model (Gail et al., 1989; Banegas et al., 2017; Zhang, 2020). The tool takes women's medical, reproductive, and first-degree family health history as inputs to estimate the probability of developing breast cancer over a chosen interval. This tool is used by approximately 40% of physicians practicing in family medicine, internal medicine, or gynecology (Corbelli et al., 2014; Pruitt et al., 2024). 10-year diabetes risks were calculated using the Finnish Diabetes Risk Score (FINDRISC) estimator (Lindström and Tuomilehto, 2003). This tool is also commonly used by physicians and has been validated in international populations (Zhang et al., 2014; Makrilakis et al., 2011). Both the BCRAT and FINDRISC tools incorporate family history, and rely on health and demographic information that subjects should readily know about themselves, as opposed to laboratory test results.

3.6 Post-Task Questionnaire

After the completion of all decision tasks and before receiving information about their payoffs, subjects completed a questionnaire about their preventive health behaviors. These questions related to the time since each subject last received screenings for high cholesterol, high blood pressure, diabetes, and breast cancer, as well as how frequently they use sunscreen. Choices for preventive screenings were categorical; either less than 1 year, 1 to 2 years, 2 to 3 years, 3 years or more, or Never. This information enables me to identify whether subjective beliefs are predictive of preventive health behaviors, and the moderating effects of incentives and response modes in this relationship.

3.7 Controlling for Online Searching

Monetary rewards for accuracy create an incentive for subjects to seek outside information, and online experiments limit an experimenter's ability to observe or prevent such behavior.

While subjects were explicitly instructed to base their decisions only on their own knowledge and experience, this request was unenforceable. Prior research confirms that online searching can be a concern when beliefs are elicited for questions with readily searchable answers. For instance, Grewenig et al. (2022) find that when a belief task involves a searchable fact, incentives improve belief accuracy and increase time spent on the task - closely mirroring behavior in an unincentivized treatment where subjects were explicitly encouraged to use search engines. However, they find no such accuracy improvements when the answer is not easily accessible online.

To identify whether subjects engaged in online searching during my experiment, I implemented a control task designed to be highly searchable. Subjects were asked: "How many grams of sugar are in one regular 12 fl oz can of Coca-Cola?" They then allocated their allotted tokens across 10 bins ranging from 0 to 100 grams, following the same structure as the risk tasks. The correct answer can be found online within seconds, making this an effective benchmark for detecting search behavior.

Not all tasks in the experiment were equally searchable. In the 10-year risk tasks, online tools exist that allow subjects to generate personalized risk scores, but these require time and effort. For instance, various FINDRISC diabetes risk score calculators provide 10-year estimates using the same health information collected in the pre-task questionnaire. Similarly, for breast cancer risk, BCRAT online calculators exist but only offer 5-year and lifetime estimates rather than 10-year projections.

For other health conditions, such as heart disease or stroke, some related online risk calculators are available, but I am unaware of tools that allow for personalized prevalence estimates based on multiple characteristics. If subjects attempted to refine their beliefs about the prevalence of these conditions among similar subpopulations, they would likely find only broad demographic trends (e.g., by sex, age, or race) rather than highly tailored estimates. As a result, my analysis focuses primarily on responses from the prevalence tasks, where online searchability is most limited.

To further assess potential search behavior and effort, I track two measures: total time spent on the task, and "focused" time , the duration during which subjects exclusively viewed the browser tab displaying the experimental interface. If a subject switched to another tab or application, this counted toward total time but not focused time. The difference between these two measures provides an indication of subjects who spent time "unfocused," potentially using external resources to update their beliefs. While this measure cannot detect if subjects searched using another device, it has been validated in prior research. Graham (2024) employ a similar method to detect cheating in online surveys by identifying instances where the

survey window was obscured, finding that this approach captures 70% to 85% of search attempts.

4 Results

4.1 Elicited Beliefs

To compare accuracy across health conditions and subjects with varying objective risk levels, I normalized the reported beliefs. Normalization was done by indexing the bin containing the subject's predicted risk at zero and shifting other bins accordingly. This approach provides a consistent framework for evaluating how subjective beliefs align with objective probabilities, making it easier to identify systematic patterns of error. In the experimental design, beliefs are measured in discrete units of "bins", where each bin represents a probability range with width 0.1. Figure 3 presents the pooled distributions of normalized beliefs, aggregated by task and response mode.

Subjects systematically overestimate their risk of low-probability health outcomes, while results for other health conditions are mixed. This trend is observable by comparing correspondence between the mean of pooled distributions relative to predicted outcomes. For less-common conditions such as stroke and skin cancer, the mean of each cumulative belief distribution is 1-2 bins above the predicted risk, equivalent to errors in probability of 10-20%. Evaluating these errors in terms of relative magnitudes is more striking. For 10-year breast cancer risk, where the average predicted probability among eligible participants is only 3.42%, allocating one's tokens 2 bins higher than the correct bin implies one has subjective probabilities six to nine times higher than the objective estimate. This overestimation result is consistent with unincentivized work examining subjective beliefs about the risk of breast cancer Carman and Kooreman (2014).

Individual and pooled belief distributions are also quite diffuse. Interquartile ranges for pooled belief distributions fall between 2 and 4 bins (equivalent to probability ranges of 0.2 to 0.4), indicating substantial imprecision in the subject pool's cumulative ability to predict their own health risks. At the individual level, subjects in distributional treatments also demonstrated a substantial degree of uncertainty in their answers. In both treatments I100 and N100, subjects allocated tokens to a median of 3 bins per task, averaging similar levels of confidence measured via standard deviation as well. Figure 4 displays a random sample of responses from distributional treatments. Less than 10% of responses were degenerate beliefs, where subjects allocated all 100 tokens to the same bin. This clearly shows that subjects are ready to communicate a richness of information about their beliefs beyond what is typically

captured by a point estimate, even in the absence of incentives. Researchers who choose to elicit point estimates may then miss out on informative data.

4.2 Incentive Effects

The implemented formulation of the Quadratic Scoring Rule penalizes subjects equally for allocating tokens to incorrect bins, regardless of how far those bins are from the correct one. To better assess the accuracy of participants' reported beliefs, I calculate the mean absolute error (MAE) of each elicited belief. The MAE is defined as:

$$MAE_j = \sum_{i=1}^{10} p_i \cdot |i - b_j| \,, \tag{2}$$

where i indexes each bin (from 1 to 10), p_i denotes the proportion of tokens allocated to bin i, and b_j is the index of the bin corresponding to the correct answer for elicited belief j. This measure captures the sum of absolute deviations between the participant's reported distribution and the objective estimate, weighted by the probability mass allocated to each bin. In 100 token treatments, each token corresponds to a probability mass of 0.01, and the singular token in 1 token treatments is equal to a probability mass of 1. This statistic more heavily penalizes answers that are farther from the truth, unlike the formulation of the QSR implemented for incentive-compatible belief elicitation in the experimental design itself. An MAE measure may be interpreted as the average distance between a subject's token allocation and the true belief, measured in units of bins. MAE is always positive, and lower values correspond to higher accuracy.

Monetary incentives increase accuracy in distributional treatments, but have no effect on elicited point estimates. Figure 7 displays incentive effects on accuracy for all tasks. In point estimate treatments, estimated effects have mixed signs and are insignificant at the 5% confidence level. For distributional treatments, all estimated effect sizes are negative and are statistically significant for 6 of the 10 health condition tasks. Averaging across prevalence tasks, incentives reduce mean absolute error of distributional belief reports by 0.48. This is equivalent to a shifting of beliefs 4.8 percentage points closer to the truth. Figure 8 also shows these results are consistent when I limit the sample to non-searchers. This pattern may suggest that under a simpler one-token reporting format, subjects do not have sufficient means to increase the precision of their answers. However, when subjects are asked to report complete distributions and account for their subjective confidence level in a more granular manner, incentives do drive them to provide more accurate self-assessments of their health risks.

4.3 Response Mode Effects

To compare the effects of response mode on accuracy, I collapse elicited belief distributions to their mean. Beliefs from point estimate treatments are closer to predicted risks than the collapsed belief distributions. Across all treatments, distributional means have an average absolute error of 2.23 bins, while point estimates perform better at an MAE of 1.78. This is equivalent to a reduction in probability error of 4.5 percentage points. This effect size is also similar in magnitude to the previously highlighted effect of incentives under distributional response modes.

I find that point estimates are more accurate than distributional beliefs when subjects are unincentivized, but these differences largely disappear when subjects are incentivized for accuracy. Figure 9 displays differences in MAE, separating effects by incentive treatment conditions. Across prevalence tasks, errors in treatment N100 are 7.0 percentage points larger than in treatment N1, but this this number falls to 2.6 across treatments I100 and I1. This suggests that incentives may drive improved truthful revelation only when reporting beliefs is a more cognitively demanding task. It may require only marginal increases in effort for subjects to tell the truth when reporting a point estimate relative to making a purely random selection, but as subjects spent three times the duration to report belief distributions, incentives may help to ensure that they are sufficiently motivated to allocate all tokens in a precise manner.

4.4 Effort and Online Searching

I find that distributional response modes lead to large increases in time spent per task, while incentives have comparatively smaller but still significant effects. Figure 5 displays focused time per task, organized by treatment and condition. In point estimate treatments, subjects spend an average of 12.9 seconds per task when unincentivized and 17.1 seconds under incentives. By comparison, subjects in distributional treatments spend more than twice as much focused time per task, averaging 36.3 seconds without incentives and 40.8 seconds when incentivized. Pairwise t-tests confirm that average task times differ significantly across all treatments at the 95% confidence level. The substantial difference between response modes can be partially attributed to the additional physical effort required to allocate more tokens. However, since time spent also varies by incentive condition while holding response mode constant, this suggests that incentives lead to greater cognitive effort, prompting subjects to reflect longer before submitting their responses.

Incentives also increase online searching during the control task by 8.1 percentage points, a 62% increase. To identify likely use of external sources, I compare focused and total task

time, classifying any subject who spent more than 5% of their control task time unfocused as a "searcher." Overall, 17% of participants meet this criterion, with the highest proportion in Treatment I100 (25%) and the lowest in Treatment N1 (12%). Figure 6 illustrates these differences, showing the proportion of searchers across all treatments and tasks. While fewer than 10% of subjects spend appreciable time unfocused on tasks other than the control, I define "searchers" based solely on control-task behavior, where online lookup is most straightforward. As expected, searchers perform significantly better on the control question, with QSR scores 33.0 points higher than non-searchers in distributional treatments and 53.2 points higher in point estimate treatments. In contrast, I find no significant differences in scoring between searchers and non-searchers in non-control tasks. My findings on the effect of incentives on the frequency and efficacy of online searching align with Grewenig et al. (2022), who similarly show that incentives encourage greater online searching and have a stronger impact on belief accuracy when answers are easily searchable. Confirmation of this result is an important note for future work implementing incentivized health belief elicitation online, as careful attention must be paid to the searchability of answers to a given task.

4.5 Beliefs and Health Behaviors

I examine the relationship between subjective beliefs and self-reported preventive health behaviors. In the post-task questionnaire, participants indicate how long it has been since they were last screened for a set of health conditions, as well as how often they use sunscreen. I use responses to generate a binary variable indicating if the subject reported a screening within the past three years. In the case of beliefs about skin cancer, I create an indicator variable for whether subjects report using sunscreen sometimes, often, or always. I then regress each indicator variable on the participant's belief about their personal risk of the condition, where beliefs from distributional treatments were collapsed to their mean.

I find a positive relationship between health risk perceptions and related preventive health behaviors. Figure 10 displays regression results, where each coefficient represents the change in probability of engaging in the preventive behavior associated with a 10% increase in subjective risk. Coefficient estimates for breast cancer screening, sunscreen use, and cholesterol tests are all approximately 3.5%. For someone who believes their risk of a disease is 50%, this means they would be 17.5% more likely to have received a screening for this condition than someone who believes they have little to no risk. I find smaller estimates for beliefs about the risk of high blood pressure and diabetes.

Confidence also appears to play a role in the predictive power of subjective beliefs. I calculate the standard deviation of beliefs from distributional treatments and split subjects

into groups of high and low confidence at the median standard deviation. Results from subsample regressions are shown in Figure 11. Beliefs for individuals in the high confidence group are more predictive of preventive health behaviors than those from the low confidence group. In the case of sunscreen use and beliefs about skin cancer, I find a strong divergence. For individuals with high confidence in their skin cancer risk, I estimate a coefficient of 0.087, while for low confidence, this coefficient is 0.021. I find similar trends across all tasks, where coefficient estimates from the high confidence group are greater than that in the low confidence group. This implies that if two individuals hold the same mean belief about their risk of disease, the one who feels more certain in their belief is more likely to have undertaken preventive action against the condition.

5 Summary and Concluding Remarks

This paper addressed three key questions about measuring subjective health beliefs: whether monetary incentives improve accuracy, whether distributional responses capture valuable information beyond point estimates, and how confidence moderates the relationship between beliefs and preventive behaviors. Using a 2×2 experimental design varying both incentives and response modes in an online sample of U.S. adults, I found that monetary incentives improve the accuracy of reported beliefs, but only when subjects can communicate complete distributions - reducing mean absolute error by 4.8 percentage points. When limited to point estimates, incentives showed no significant effect on accuracy. The results also revealed systematic overestimation of low-probability health outcomes and demonstrated that beliefs more strongly predict preventive behaviors when they exhibit high confidence.

The findings have important implications for both survey methodology and health policy. For survey designers measuring health expectations, the choice of elicitation method should align with specific research objectives. When the goal is simply to measure average beliefs across a population, traditional point estimates perform adequately and incentives provide limited value at potentially substantial cost. However, researchers interested in understanding health behaviors may consider eliciting complete belief distributions with monetary incentives, as this approach helps overcome the cognitive difficulty reporting more detailed risk perceptions while improving predictive power. The systematic overestimation of low-probability health risks replicates previous findings and indicates potential gaps in public health literacy that could be addressed through targeted education initiatives. Additionally, the finding that confidence moderates the belief-behavior relationship suggests that public health campaigns may be more effective when they not only inform people about health risks but also help them develop well-calibrated confidence in their risk assessments.

Several limitations of this study warrant consideration. First, the online experimental environment created challenges in controlling for subjects' use of external information sources. While I implemented methods to detect potential online searching and found evidence that incentives do increase search attempts, the main results are robust to the exclusion of likely searchers. Future work may benefit from controlled laboratory settings. Second, the recruited sample was predominantly well-educated and experienced with online studies. (d'Uva et al., 2020) find that errors in mortality expectations are largest for subjects with low educational attainment and cognition, potentially altering the generalizability of my findings to population surveys. Third, the relationship between beliefs and preventive behaviors was measured using cross-sectional data, making it difficult to establish the causal direction of these associations.

Looking forward, this research opens several promising avenues for future investigation. The experimental framework developed here could be extended to study belief updating in response to new health information, potentially informing the design of public health information campaigns. The treatment design could also be applied to other domains where public uncertainty meets strong objective evidence, such as environmental risks or financial planning. Further research might explore whether similar patterns emerge for relative rather than absolute risk perceptions, as people often evaluate their health risks in comparison to others. Finally, the role of confidence in moderating health behaviors warrants deeper investigation, particularly regarding how different types of health information influence both risk perceptions and confidence levels.

Figures and Tables

Characteristic	Subgroup	Count	Percentage
Treatment	I100	130	26.53%
	I1	118	24.08%
	N100	121	24.69%
	N1	121	24.69%
Gender	Male	261	53.27%
	Female	229	46.73%
Age Group	35-44	120	24.49%
	45-54	119	24.29%
	55-64	121	24.69%
	65-74	130	26.53%
Race	White, non-Hispanic	355	72.45%
	Black, non-Hispanic	86	17.55%
	Hispanic	22	4.49%
	Other	27	5.51%
Education	Some High School	4	0.82%
	High School Degree	64	13.06%
	Some College/Associate's Degree	169	34.49%
	Bachelor's Degree or More	253	51.63%
Income	0 to 24,999	84	17.14%
	25,000 to 49,999	110	22.45%
	50,000 to 74,999	131	26.73%
	$75,000 or more	165	33.67%
Healthcare Visits	No Visits	54	11.02%
	1 Visit	73	14.90%
	2 to 3 Visits	189	38.57%
	4 to 9 Visits	130	26.53%
	10 or more visits	44	8.98%
Insurance Status	Insured	455	92.86%
	Uninsured	35	7.14%

Table 2: Sample Descriptive Statistics

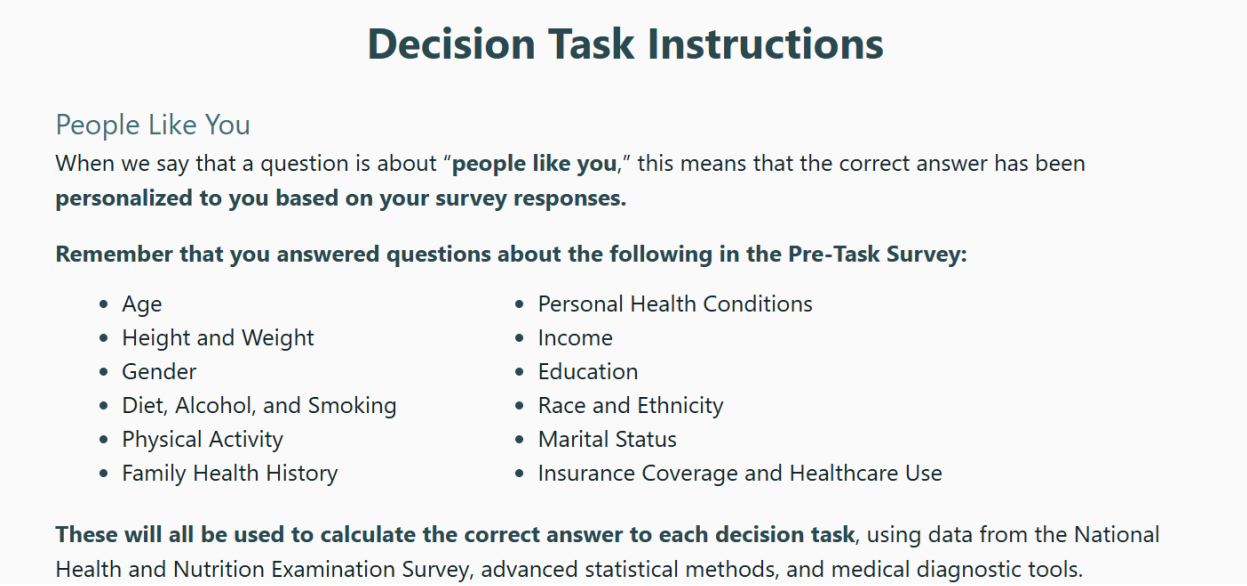


Figure 2: Subject Instructions Screenshot: Explanation of Personalized Risk Estimates

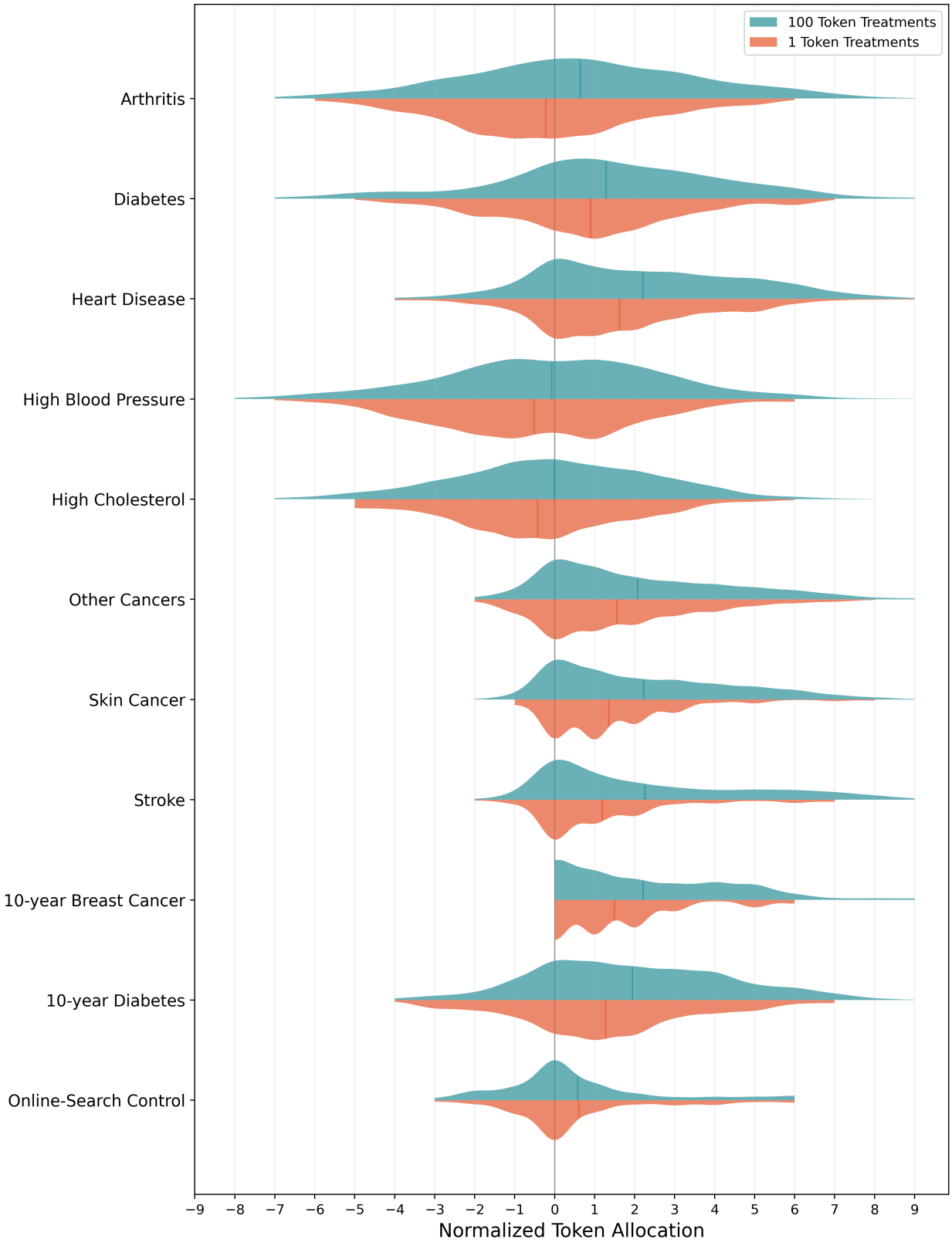


Figure 3: Pooled Normalized Beliefs

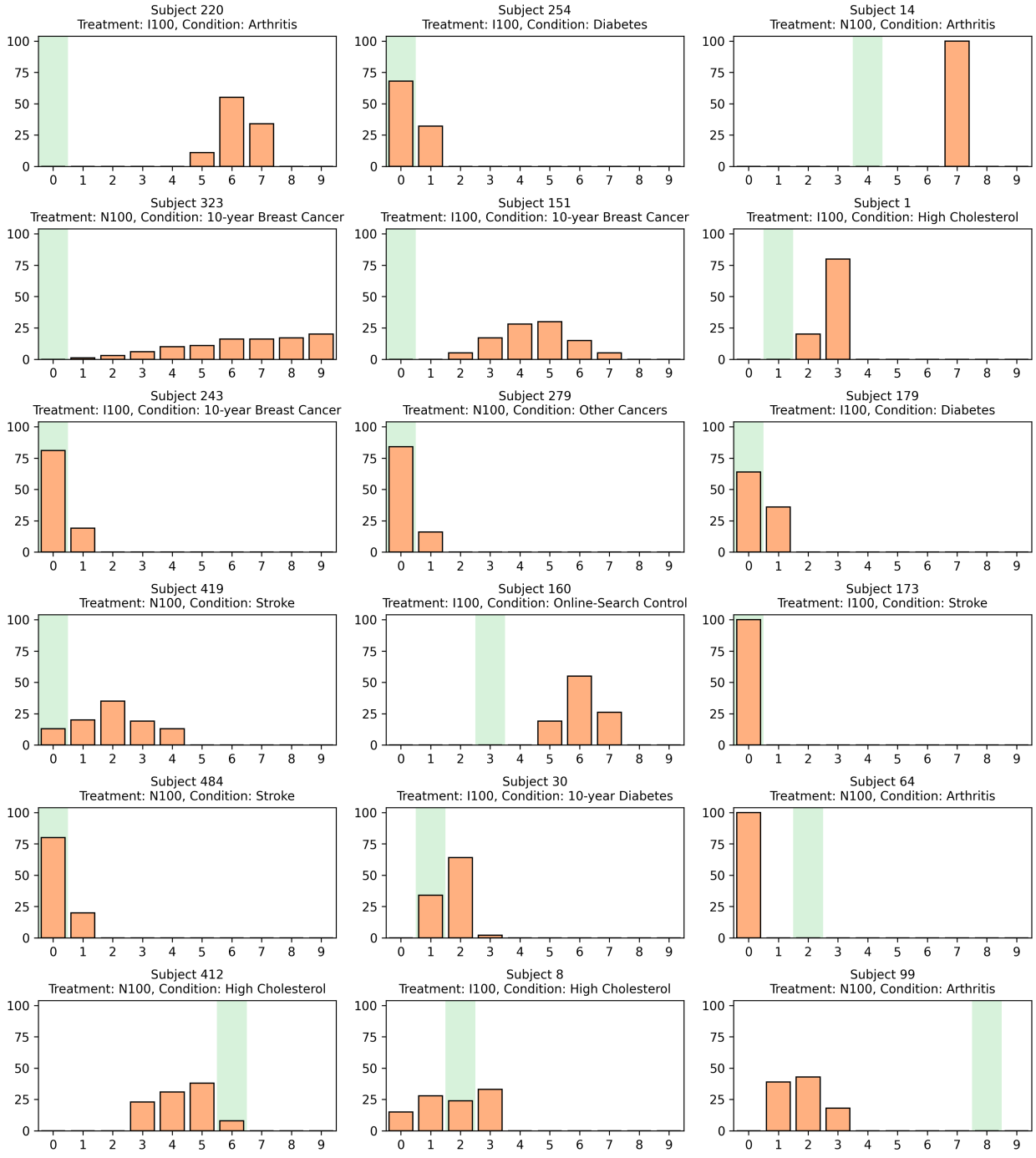


Figure 4: Sample of Distributional Responses (Correct Bins Highlighted in Green)

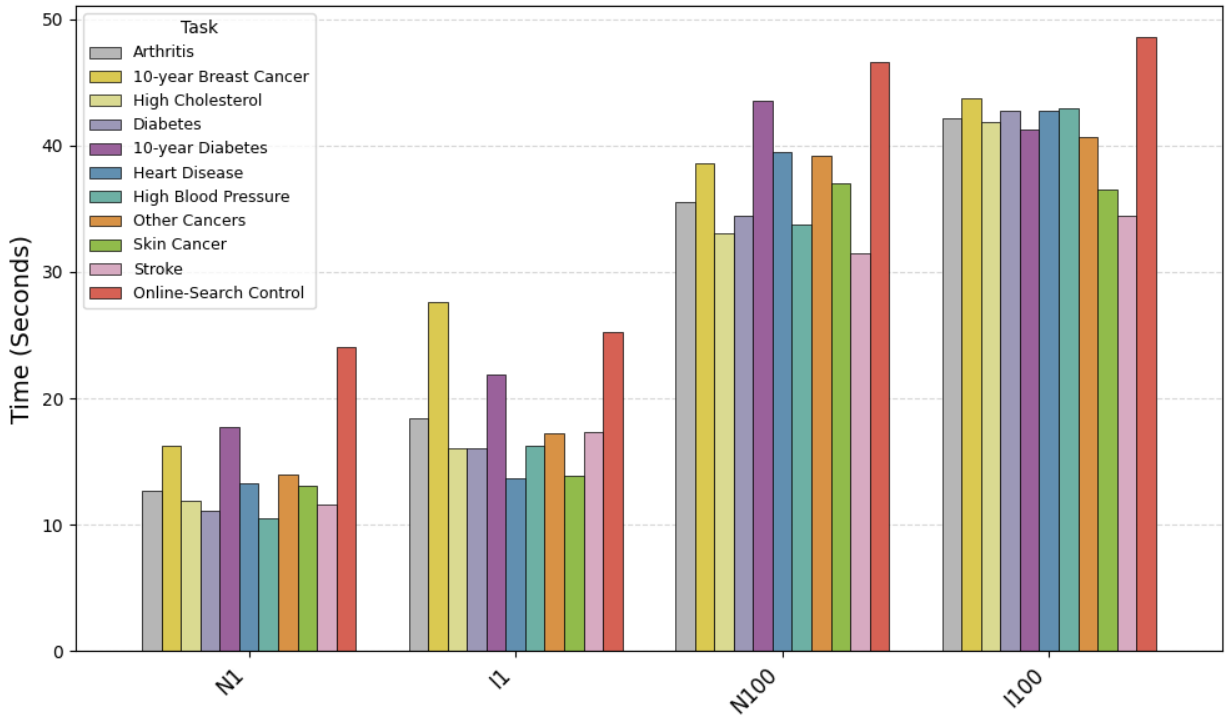


Figure 5: Focused Task Time

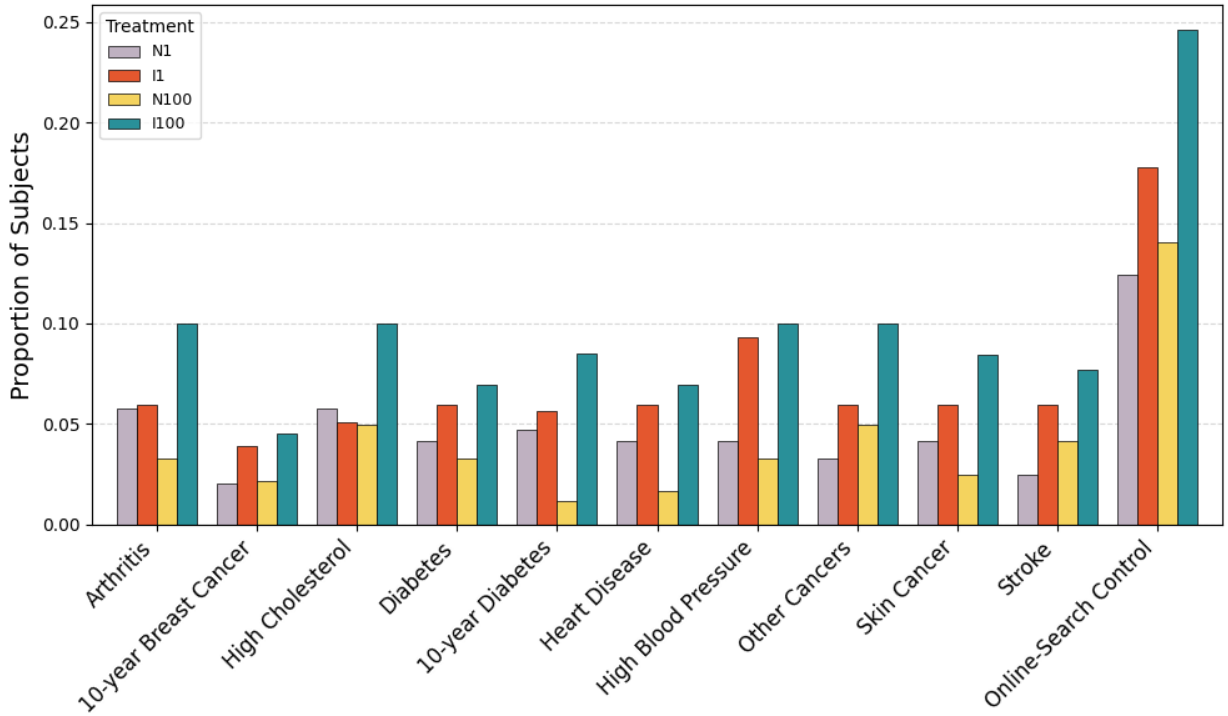


Figure 6: Proportion of Online Searching

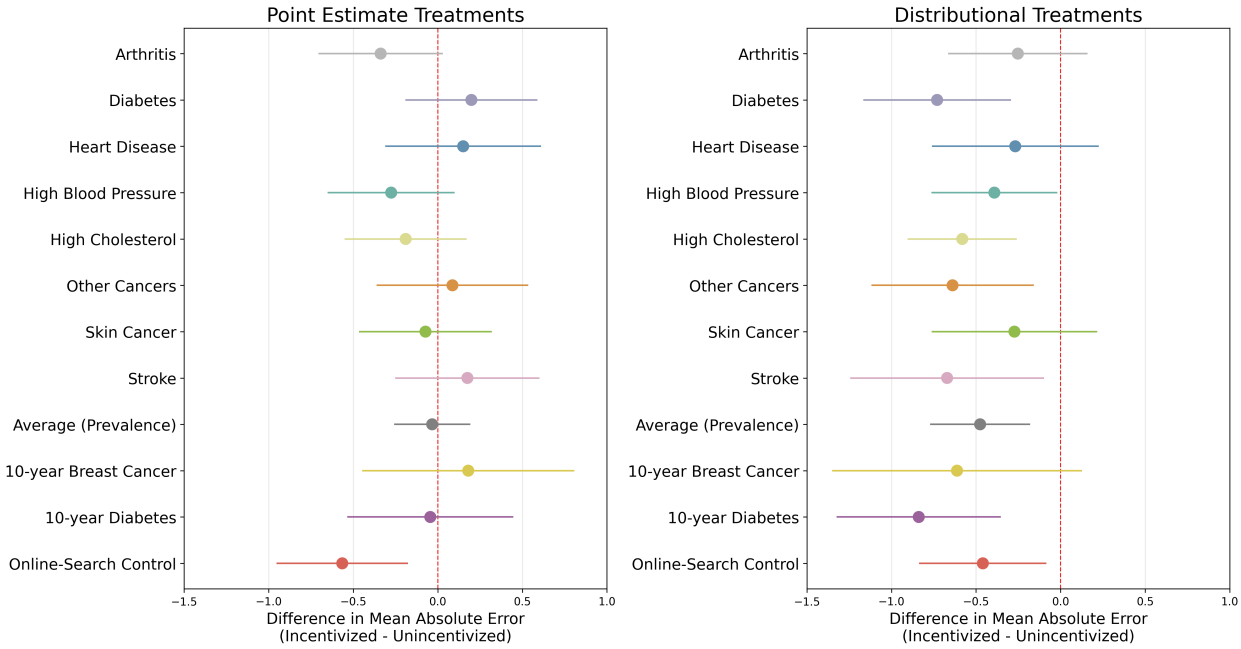


Figure 7: Incentive Effects

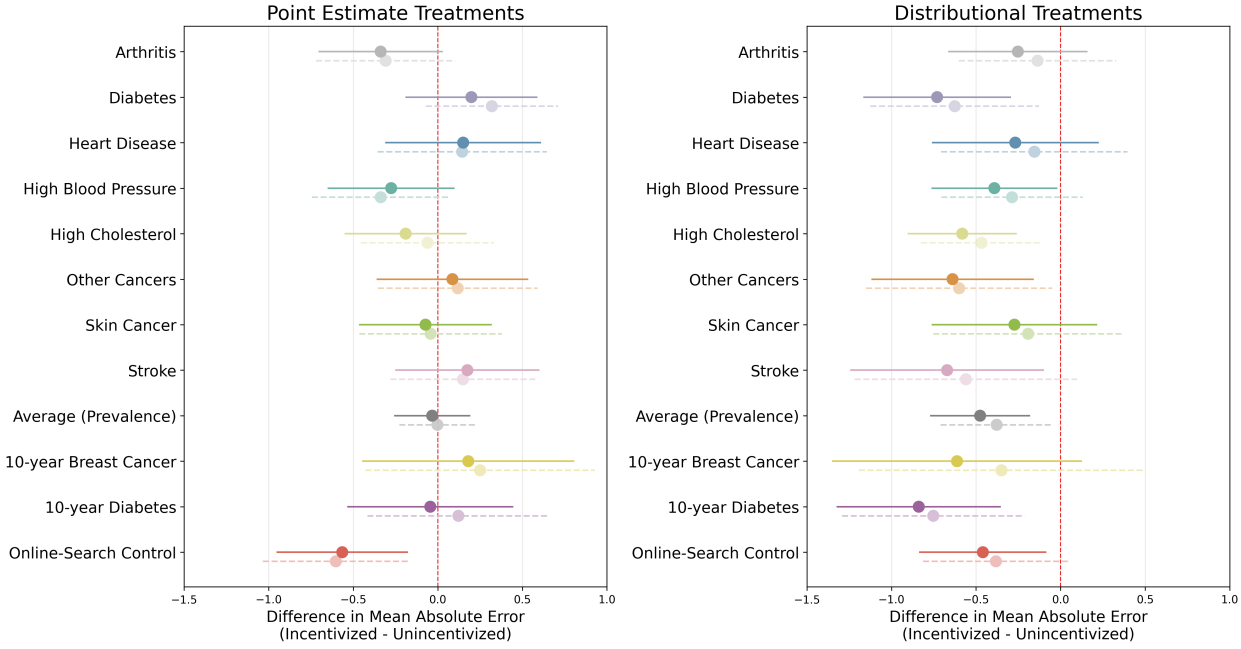


Figure 8: Incentive Effects Subsample Analysis - Sample Excluding Online Searchers shown in Lighter Colors

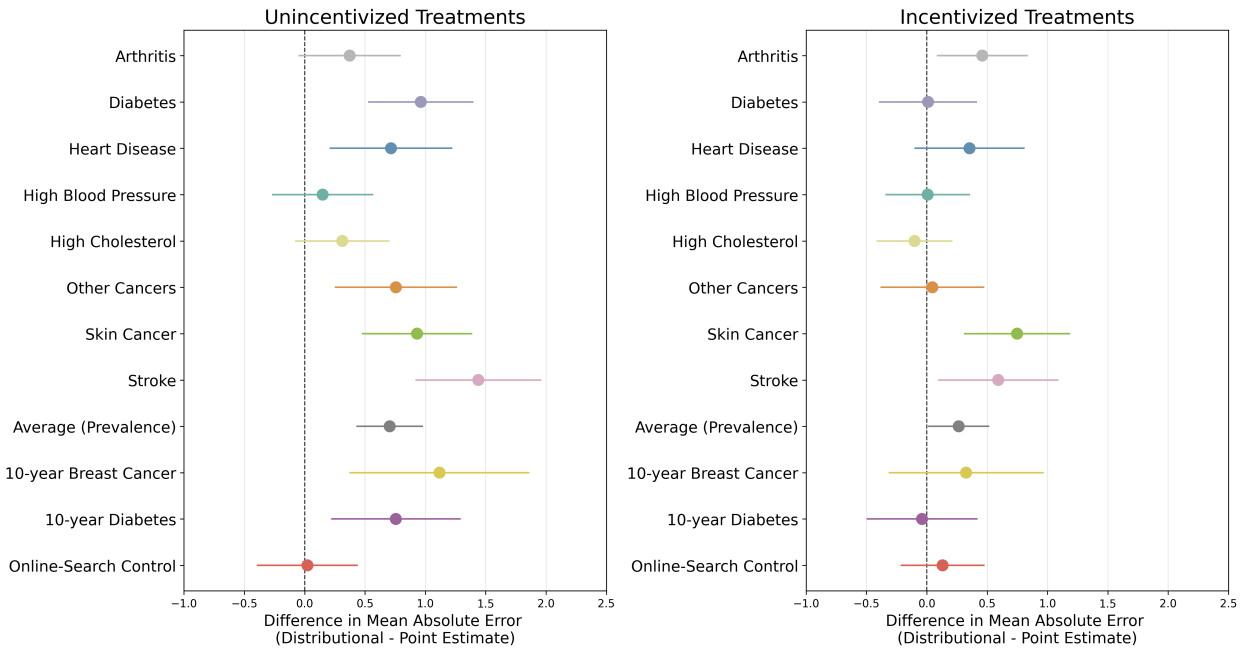


Figure 9: Response Mode Effects

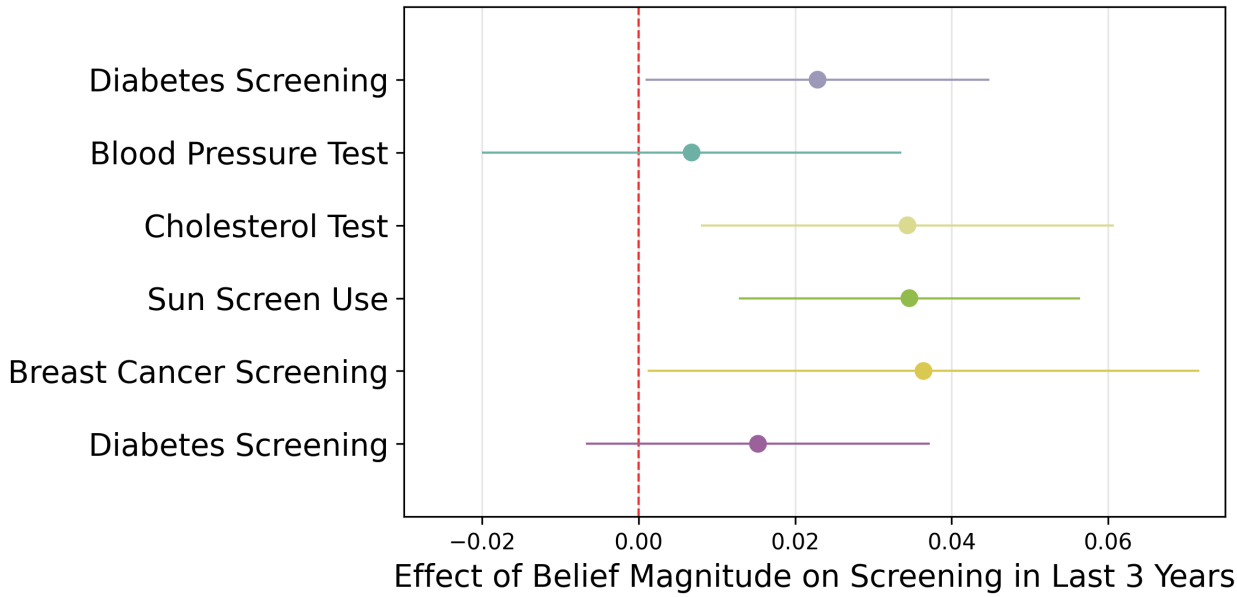


Figure 10: Association between Subjective Beliefs and Preventive Behaviors

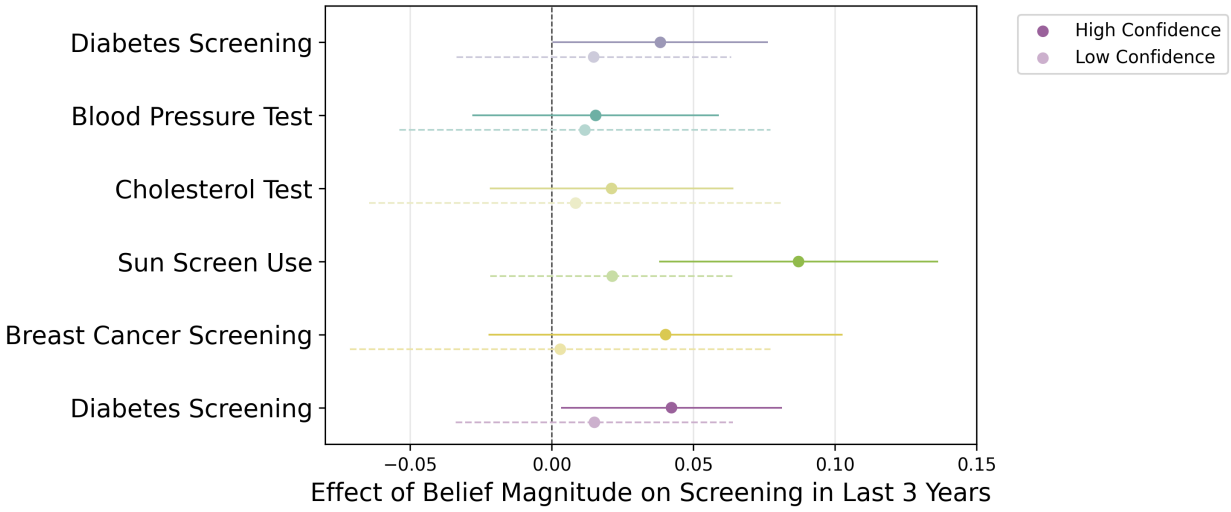


Figure 11: Effects of Confidence on Predictive Power of Beliefs

References

Banegas, M. P., John, E. M., Slattery, M. L., Gomez, S. L., Yu, M., LaCroix, A. Z., Pee, D., Chlebowski, R. T., Hines, L. M., Thompson, C. A., and Gail, M. H. (2017). Projecting Individualized Absolute Invasive Breast Cancer Risk in US Hispanic Women. *Journal of the National Cancer Institute*, 109(2):djw215.

Borsky, A., Zhan, C., Miller, T., Ngo-Metzger, Q., Bierman, A. S., and Meyers, D. (2018). Few Americans Receive All High-Priority, Appropriate Clinical Preventive Services. *Health Affairs*, 37(6):925–928.

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78:1.

Bruine De Bruin, W. and Carman, K. G. (2012). Measuring Risk Perceptions: What Does the Excessive Use of 50% Mean? *Medical Decision Making*, 32(2):232–236.

Burfurd, I. and Wilkening, T. (2022). Cognitive heterogeneity and complex belief elicitation. *Experimental Economics*, 25(2):557–592.

Carman, K. G. and Kooreman, P. (2014). Probability perceptions and preventive health care. *Journal of Risk and Uncertainty*, 49(1):43–71.

Carpenter, C. J. (2010). A Meta-Analysis of the Effectiveness of Health Belief Model Variables in Predicting Behavior. *Health Communication*, 25(8):661–669.

Corbelli, J., Borrero, S., Bonnema, R., McNamara, M., Kraemer, K., Rubio, D., Karpov, I., and McNeil, M. (2014). Use of the Gail Model and Breast Cancer Preventive Therapy Among Three Primary Care Specialties. *Journal of Women's Health*, 23(9):746–752.

Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief Elicitation and Behavioral Incentive Compatibility. *American Economic Review*, 112(9):2851–2883.

Danz, D., Vesterlund, L., and Wilson, A. J. (2024). Evaluating Behavioral Incentive Compatibility: Insights from Experiments. *Journal of Economic Perspectives*, 38(4):131–154.

Delavande, A., Bono, E. D., and Holford, A. (2024). Imprecise health beliefs and health behavior.

Delavande, A., Giné, X., and McKenzie, D. (2011). Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, 26(3):479–497.

Di Girolamo, A., Harrison, G. W., Lau, M. I., and Swarthout, J. T. (2015). Subjective belief distributions and the characterization of economic literacy. *Journal of Behavioral and Experimental Economics*, 59:1–12.

d'Uva, T., O'Donnell, O., and van Doorslaer, E. (2020). Who can predict their own demise? Heterogeneity in the accuracy and value of longevity expectations. *The Journal of the Economics of Ageing*, 17:100135.

Einav, L., Finkelstein, A., Mullainathan, S., and Obermeyer, Z. (2018). Predictive modeling of U.S. health care spending in late life. *Science (New York, N.Y.)*, 360(6396):1462–1465.

Elder, T. E. (2013). The Predictive Validity of Subjective Mortality Expectations: Evidence From the Health and Retirement Study. *Demography*, 50(2):569–589.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters. *Journal of Business & Economic Statistics*, 27(1):30–41.

Erkal, N., Gangadharan, L., and Koh, B. H. (2020). Replication: Belief elicitation with quadratic and binarized scoring rules. *Journal of Economic Psychology*, 81:102315.

Festinger, L. (1962). Cognitive Dissonance. *Scientific American*, 207(4):93–106.

Fischhoff, B., Parker, A. M., de Bruin, W. B., Downs, J., Palmgren, C., Dawes, R., and Manski, C. F. (2000). Teen Expectations for Significant Life Events. *The Public Opinion Quarterly*, 64(2):189–205.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., and Mulvihill, J. J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, 81(24):1879–1886.

Garcia, M. C. (2019). Potentially Excess Deaths from the Five Leading Causes of Death in Metropolitan and Nonmetropolitan Counties — United States, 2010–2017. *MMWR. Surveillance Summaries*, 68.

Gigerenzer, G. and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4):684–704.

Giustinelli, P., Manski, C. F., and Molinari, F. (2022). Precise or Imprecise Probabilities? Evidence from Survey Response Related to Late-Onset Dementia. *Journal of the European Economic Association*, 20(1):187–221.

Graham, M. H. (2024). Detecting and Deterring Information Search in Online Surveys. *American Journal of Political Science*, 68(4):1315–1334.

Grewenig, E., Lergetporer, P., Werner, K., and Woessmann, L. (2022). Incentives, search engines, and the elicitation of subjective beliefs: Evidence from representative online survey experiments. *Journal of Econometrics*, 231(1):304–326.

Gächter, S. and Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3):364–377.

Harrison, G. W. (2014). Hypothetical Surveys or Incentivized Scoring Rules for Eliciting Subjective Belief Distributions?

Harrison, G. W., Hofmeyr, A., Kincaid, H., Monroe, B., Ross, D., Schneider, M., and Swarthout, J. T. (2022). Subjective beliefs and economic preferences during the COVID-19 pandemic. *Experimental Economics*, 25(3):795–823.

Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2013). Inducing risk neutral preferences with binary lotteries: A reconsideration. *Journal of Economic Behavior & Organization*, 94:145–159.

Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., and Ulm, E. R. (2017). Scoring rules for subjective probability distributions. *Journal of Economic Behavior & Organization*, 134:430–448.

Hossain, T. and Okui, R. (2013). The Binarized Scoring Rule. *The Review of Economic Studies*, 80(3 (284)):984–1001.

Hudomiet, P., Hurd, M., and Rohwedder, S. (2018). Measuring Probability Numeracy.

Hudomiet, P., Hurd, M. D., and Rohwedder, S. (2023). Mortality and health expectations. In *Handbook of Economic Expectations*, pages 225–259. Elsevier.

Hurd, M. D. and McGarry, K. (1995). Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study. *The Journal of Human Resources*, 30:S268–S292.

Hurd, M. D. and McGarry, K. (2002). The Predictive Validity of Subjective Probabilities of Survival*. *The Economic Journal*, 112(482):966–985.

Kerwin, J. and Pandey, D. (2023). Navigating Ambiguity: Imprecise Probabilities and the Updating of Disease Risk Beliefs. *SSRN Electronic Journal*.

Khwaja, A., Silverman, D., Sloan, F., and Wang, Y. (2009). Are mature smokers misinformed? *Journal of Health Economics*, 28(2):385–397.

Lindström, J. and Tuomilehto, J. (2003). The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care*, 26(3):725–731.

Makrilakis, K., Liatis, S., Grammatikou, S., Perrea, D., Stathi, C., Tsiligros, P., and Katsilambros, N. (2011). Validation of the Finnish diabetes risk score (FINDRISC) questionnaire for screening for undiagnosed type 2 diabetes, dysglycaemia and the metabolic syndrome in Greece. *Diabetes & Metabolism*, 37(2):144–151.

Manski, C. F. and Molinari, F. (2010). Rounding Probabilistic Expectations in Surveys. *Journal of Business & Economic Statistics*, 28(2):219–231. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/jbes.2009.08098.

Matheson, J. E. and Winkler, R. L. (1976). Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10):1087–1096. Publisher: INFORMS.

Perozek, M. (2008). Using Subjective Expectations to Forecast Longevity: Do Survey Respondents Know Something We Don't Know? *Demography*, 45(1):95–113.

Pruitt, W. R., Samuels, B., and Cunningham, S. (2024). The Gail Model and Its Use in Preventive Screening: A Comparison of the Corbelli Study. *Cureus*.

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change. *The Journal of Psychology*, 91(1):93–114.

Rosenstock, I. M. (1974). The Health Belief Model and Preventive Health Behavior. *Health Education Monographs*, 2(4):354–386.

Scherer, L. D., Ubel, P. A., McClure, J., Greene, S. M., Alford, S. H., Holtzman, L., Exe, N., and Fagerlin, A. (2013). Belief in numbers: When and why women disbelieve tailored breast cancer risk statistics. *Patient Education and Counseling*, 92(2):253–259.

Shah, D., Patel, S., and Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, 1(6):345.

Sloan, F. A. (2024). Subjective beliefs, health, and health behaviors. *Journal of Risk and Uncertainty*.

Smith, V. K., Taylor, D. H., and Sloan, F. A. (2001). Longevity Expectations and Death: Can People Predict Their Own Demise? *The American Economic Review*, 91(4):1126–1134.

Smith, V. L. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review*, 72(5):923–955.

Soni, M. and Varma, D. S. (2020). Diabetes Prediction using Machine Learning Techniques. *International Journal of Engineering Research*, 9(09).

Trautmann, S. T. and van de Kuilen, G. (2015). Belief Elicitation: A Horse Race among Truth Serums. *The Economic Journal*, 125(589):2116–2135.

Wang, S. W. (2011). Incentive effects: The case of belief elicitation from individuals in groups. *Economics Letters*, 111(1):30–33.

Zhang, F. (2020). BCRA: Breast Cancer Risk Assessment. Institution: Comprehensive R Archive Network Pages: 2.1.2.

Zhang, L., Zhang, Z., Zhang, Y., Hu, G., and Chen, L. (2014). Evaluation of Finnish Diabetes Risk Score in Screening Undiagnosed Diabetes and Prediabetes among U.S. Adults by Gender and Race: NHANES 1999-2010. *PLoS ONE*, 9(5):e97865.

Zhong, H., Deck, C., and Henderson, D. J. (2024). Do participation rates vary with participation payments in laboratory experiments? *Experimental Economics*.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., and Tang, H. (2018). Predicting Diabetes Mellitus With Machine Learning Techniques. *Frontiers in Genetics*, 9.

Appendix A Health Condition Definitions

Below are the definitions provided in the subject instructions

Arthritis: Disorders that primarily cause inflammation and swelling in the joints. This includes osteoarthritis, degenerative arthritis, rheumatoid arthritis, and psoriatic arthritis.

Breast Cancer: Any cancer that starts specifically in the breast.

Cancer (other than skin cancer): Diseases in which cells in the body grow out of control. This question includes cancers of the breast, lung, prostate, liver, colon, and bladder. This task does not include skin cancer. We provide a separate task for cancers specifically from the skin.

Diabetes: Prolonged high blood sugar levels due to either the pancreas not producing enough insulin or the cells of the body not responding properly to the insulin produced. Also called sugar diabetes.

Heart Disease: Diseases affecting blood supply to the heart, or the heart's ability to supply blood to the rest of the body. These include heart attack (also called myocardial infarction), coronary heart disease, angina (also called angina pectoris), and congestive heart failure.

High Blood Pressure: When the force of blood against the artery walls is consistently too high. Also called hypertension.

High Cholesterol: Too much buildup of fats in the blood, potentially leading to the growth of plaques and reduced blood flow.

Skin Cancer: Cancers that start in the skin, including melanoma, basal-cell, and squamous-cell skin cancers.

Stroke: Poor blood flow to the brain, causing cell death. This includes both ischemic stroke due to lack of blood flow, and hemorrhagic stroke due to bleeding.

Appendix B Machine-Learning Algorithm

B.1 Data

I trained a machine learning algorithm following the approach of Einav et al. (2018), using data from the National Health and Nutrition Examination Survey (NHANES), produced by the US Centers for Disease Control. The sample was restricted to all adults over the age of 20 with nonmissing covariates from waves 1999-2000 through 2017-2018. Of an available 55,081 survey responses, 15,182 observations contained at least one missing predictor variable, yielding an analysis sample of 39,899. When the relevant outcome of interest for prediction was one of the binary condition diagnoses, this indicator was excluded as a predictor. The predictor variables used for all models are:

- Sex
- Age, age-squared, and categorical age-group
- Race/ethnicity (non-Hispanic White, Hispanic, non-Hispanic Black, other or mixed race)
- Years of education
- Household income category in CPI-adjusted 2023 dollars ($24,999 or less, $25,000 to $49,999, $50,000 to $74,999, $75,000 or more)
- Marital status
- Health insurance status
- Height, Weight, and BMI
- Alcohol use (ever having drank more than 12 drinks of alcohol)
- Chronic alcohol use (ever consumed 4/5 drinks per day, almost every day)
- Current smoking status
- Ever having smoked 100 cigarettes
- Regular moderate physical activity
- Regular vigorous physical activity
- Immediate family history of diabetes
- Immediate family history of heart disease before age 50
- Self-reported general health (Poor, Fair, Good, Very Good, Excellent)
- Number of healthcare visits in the last year
- Diagnosis of:
  - Heart disease
  - Arthritis
  - Skin cancer
  - Cancer (other than skin cancer)
  - High cholesterol
  - High blood pressure
  - Stroke
  - Diabetes

To avoid overfitting, data was divided into 70% train and 30% test samples. Using training data, for each health condition from the prevalence tasks, the likelihood of a diagnosis of the condition was predicted using three separate models: random forest, gradient-boosted regression trees, and LASSO. I estimate each individual model using 5-fold cross-validation. This means that for each set of tuning parameters tested, the training sample is split into 5 equal-sized folds, and estimated 5 times, where one fold is left out and used to test performance. Performance for each model was measured using a negative Brier score loss function.

B.2 Model Estimation and Hyperparameter Tuning

LASSO was estimated using the logistic regression function with L1 penalization from the `scikit-learn` Python package, with maximum iterations set to 1000.

The gradient boosted regression tree model was trained using `XGBoost`. For each health condition, a hyperparameter search was performed with the following grid:

- Number of estimators: 50, 100, 500
- Maximum tree depth: 2, 4, 6
- Learning rate: 0.01, 0.1, 0.2

The random forest model was implemented via `scikit-learn`'s random forest classifier, with the following search grid:

- Number of estimators: 25, 50, 100
- Minimum samples per split: 25, 50, 100

Final hyperparameters for each model varied across health outcome variables.

B.3 Fitting the ensemble model

The predicted probability of each health condition was then estimated separately using each model. The true diagnosis value was then regressed on the three estimated probabilities by specifying a linear OLS regression without a constant. This results in the final ensemble model output being a linear combination of the three individual model predictions, as below:

$$\hat{p}_{\text{ens}} = \hat{\beta}_{\text{rf}}\hat{p}_{\text{rf}} + \hat{\beta}_{\text{gb}}\hat{p}_{\text{gb}} + \hat{\beta}_{\text{lasso}}\hat{p}_{\text{lasso}} \tag{A0}$$

Where $\hat{p}_x$ is the prediction from model x, and $\hat{\beta}_x$ is the associated weight. Final ensemble weights for each health condition are shown below in Table A1.

	Model Weights		
Health Condition	**Random Forest**	**XGBoost**	**LASSO**
Arthritis	0.1759	0.2419	0.5690
Diabetes	0.0954	0.7405	0.1937
Heart Disease	0.2511	0.2710	0.5200
High Blood Pressure	0.1357	0.4149	0.4700
High Cholesterol	0.1757	0.5930	0.2440
Other Cancers	0.0914	0.8708	0.1047
Skin Cancer	0.0000	0.6054	0.4102
Stroke	0.3883	0.0000	0.6710

Table A1: Final Ensemble Weights

B.4 Performance

Ensemble model performance metrics are shown below in Table A2 for both the NHANES test data and the sample of Prolific subjects. Performance is comparable to existing work estimating health risks (Shah et al., 2020; Soni and Varma, 2020; Zou et al., 2018), as well as the clinical risk scores used in the experiment for 10-year risk estimates (Lindström and Tuomilehto, 2003; Banegas et al., 2017).

	NHANES Test Sample			Prolific Sample		
Health Condition	**Brier Score**	**Log Loss**	**AUC-ROC**	**Brier Score**	**Log Loss**	**AUC-ROC**
Arthritis	0.1446	0.4394	0.8277	0.1917	0.5628	0.7401
Diabetes	0.0880	0.2833	0.8785	0.1503	0.4616	0.7904
Heart Disease	0.0642	0.2151	0.8791	0.0982	0.3264	0.7442
High Blood Pressure	0.1572	0.4782	0.8342	0.2235	0.6423	0.7059
High Cholesterol	0.1736	0.5175	0.7845	0.2059	0.5974	0.7426
Other Cancers	0.0613	0.2192	0.7989	0.0683	0.2545	0.6754
Skin Cancer	0.0269	0.1045	0.8754	0.0582	0.2116	0.7687
Stroke	0.0330	0.1248	0.8733	0.0404	0.1679	0.6987

Table A2: Machine Learning Model Performance Metrics

To check that each model is properly calibrated, reliability diagrams for each health outcome are presented in Figure A1 to A8. The figures plot ensemble model predictions against true health condition status for bins of individuals in the NHANES test sample and the Prolific sample. Bins were constructed by sorting individuals according to their predicted probability, and dividing them into equal-sized groups. Then, the average true probability of the condition is calculated within each group. These reliability diagrams demonstrate good concordance between predicted and true prevalence of chronic disease in both samples.

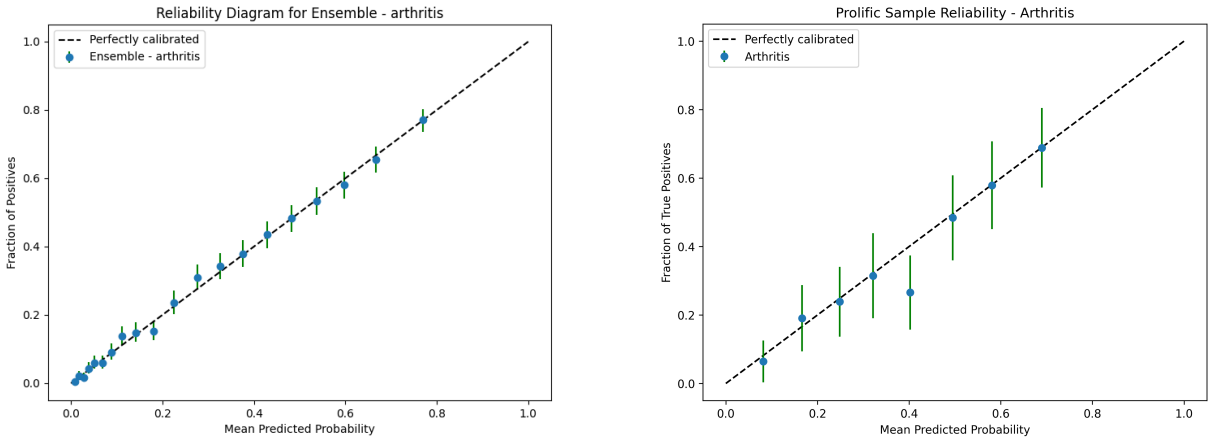


Figure A1: Reliability Diagrams for Arthritis in NHANES (Left) and Prolific (Right) Samples

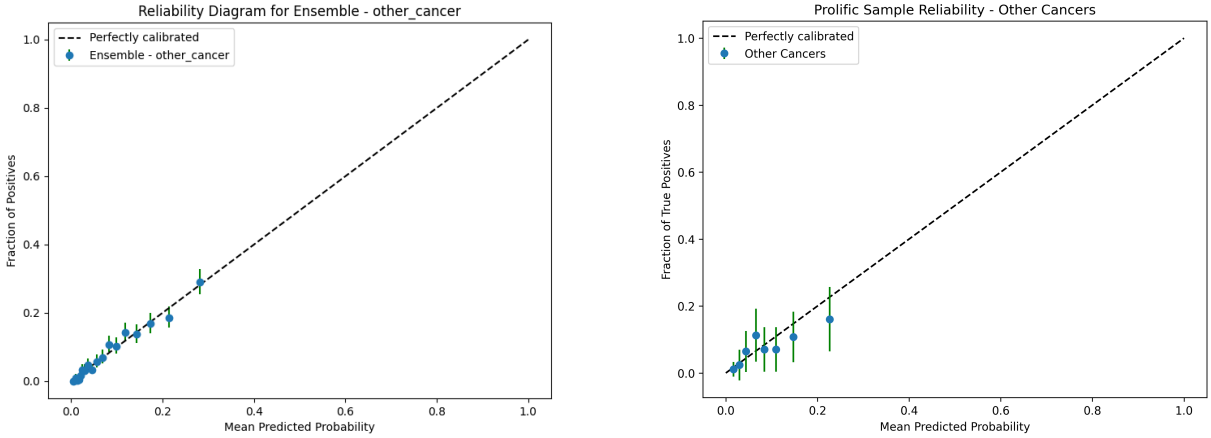


Figure A2: Reliability Diagrams for Other Cancers in NHANES (Left) and Prolific (Right) Samples

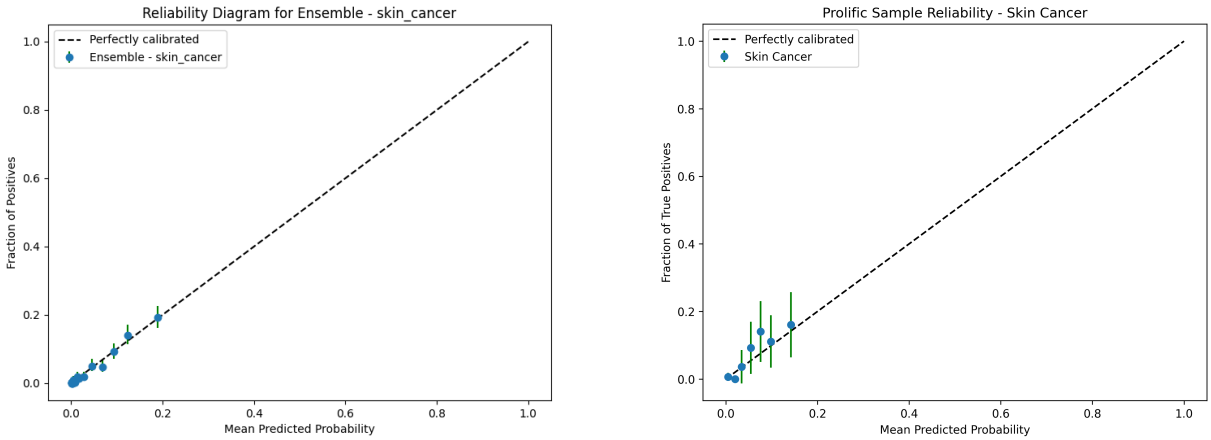


Figure A3: Reliability Diagrams for Skin Cancer in NHANES (Left) and Prolific (Right) Samples

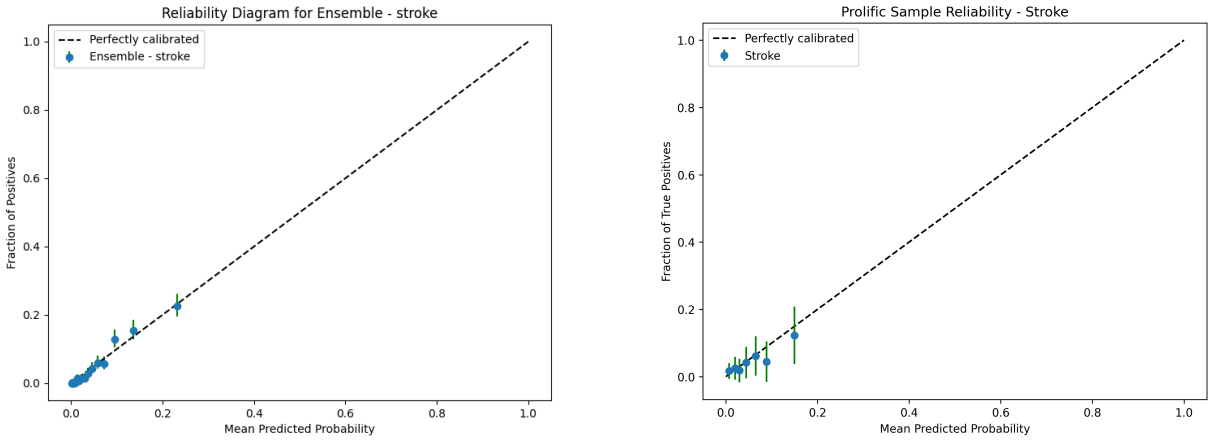


Figure A4: Reliability Diagrams for Skin Cancer in NHANES (Left) and Prolific (Right) Samples

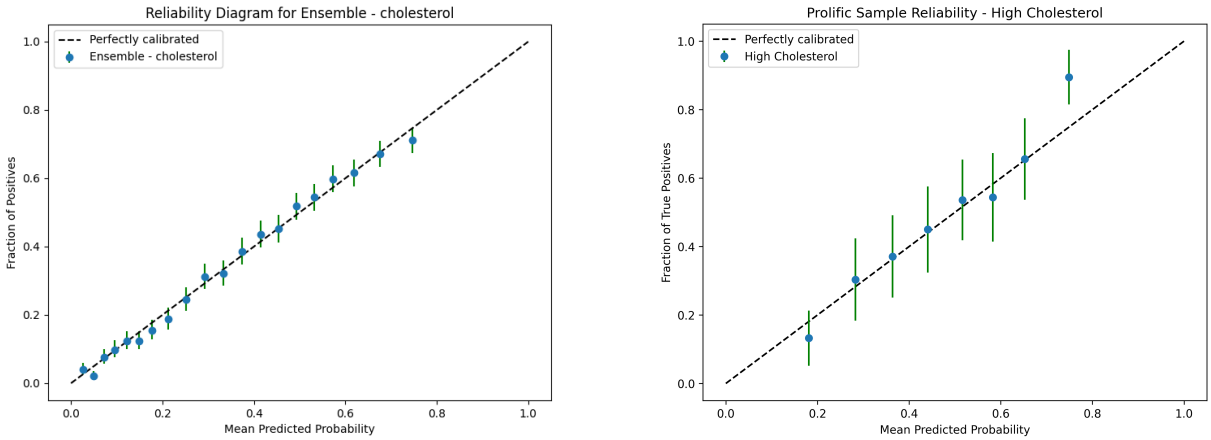


Figure A5: Reliability Diagrams for High Cholesterol in NHANES (Left) and Prolific (Right) Samples

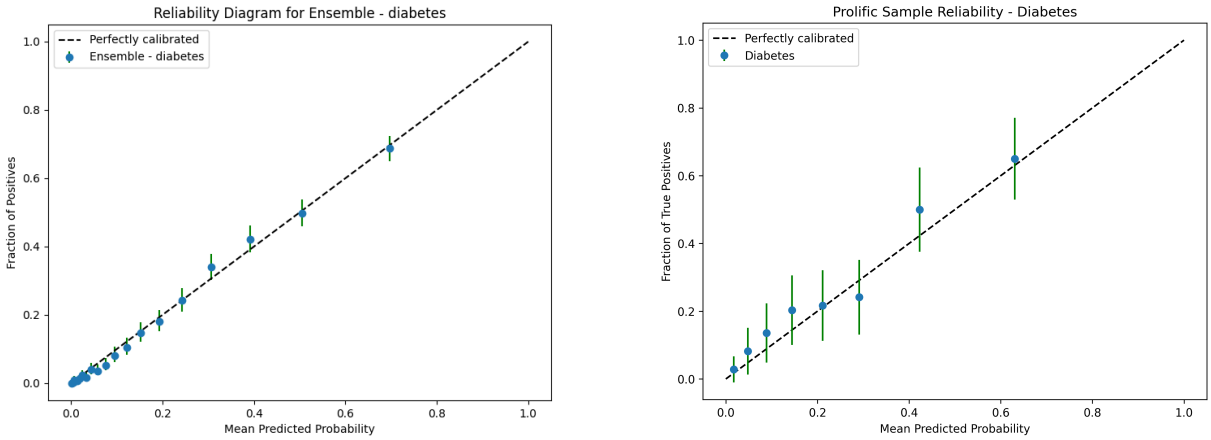


Figure A6: Reliability Diagrams for Diabetes in NHANES (Left) and Prolific (Right) Samples

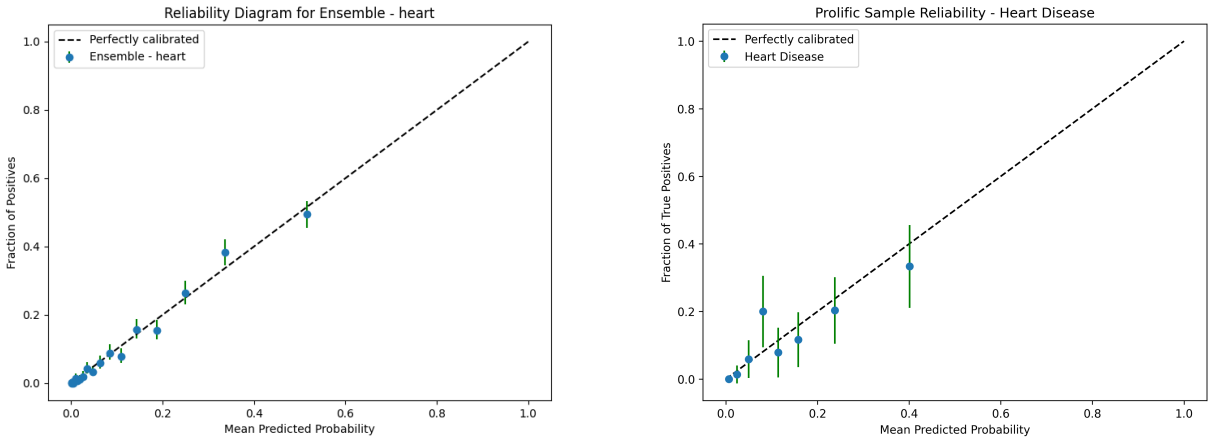


Figure A7: Reliability Diagrams for Heart Disease in NHANES (Left) and Prolific (Right) Samples

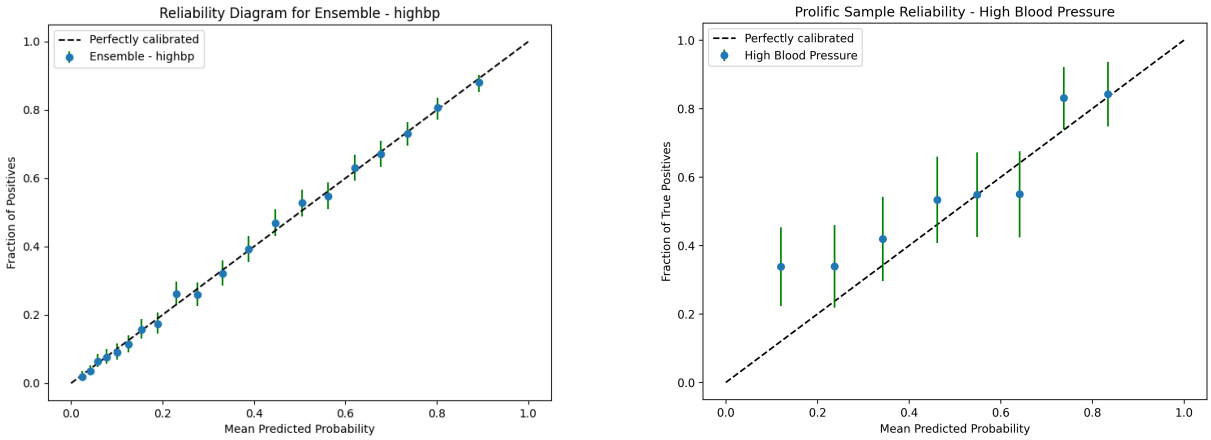


Figure A8: Reliability Diagrams for High Blood Pressure in NHANES (Left) and Prolific (Right) Samples

Appendix C Subject Instructions

Consent Form (*All Treatments*)

Georgia State University: Informed Consent

Title: Health Risk Perceptions
Principal Investigator: Professor James C. Cox
Student Principal Investigator: Ben Mosier IV
Sponsors: GSU Andrew Young School of Policy Studies, GSU Center for Economic Analysis of Risk

Procedures

We invite you to take part in a research study. You will decide if you would like to take part in the study. You must be over the age of 35 to participate. Your participation in this study should take about 30 minutes. The purpose of this study is to learn what people believe about certain things relating to health. We will ask you to answer questions about yourself, your health, your healthcare use, and your attitudes about healthcare. We will also ask you to complete a set of decision-making tasks where you will share your beliefs about health conditions among people like you. This study is not designed to benefit you, though you may learn new information about health risks during your participation.

Compensation

We will pay you through Prolific using your Prolific ID after the experiment has been completed. We will pay you $8 for your participation in this experiment.

Privacy

The information you share in the experiment will not be linked to any information that can personally identify you. You will have complete anonymity other than your Prolific ID.

Voluntary Participation and Withdrawal

If you do not wish to take part in this study, you can choose not to participate at any time. If you choose to withdraw, you will not receive a participation payment.

Contact Information

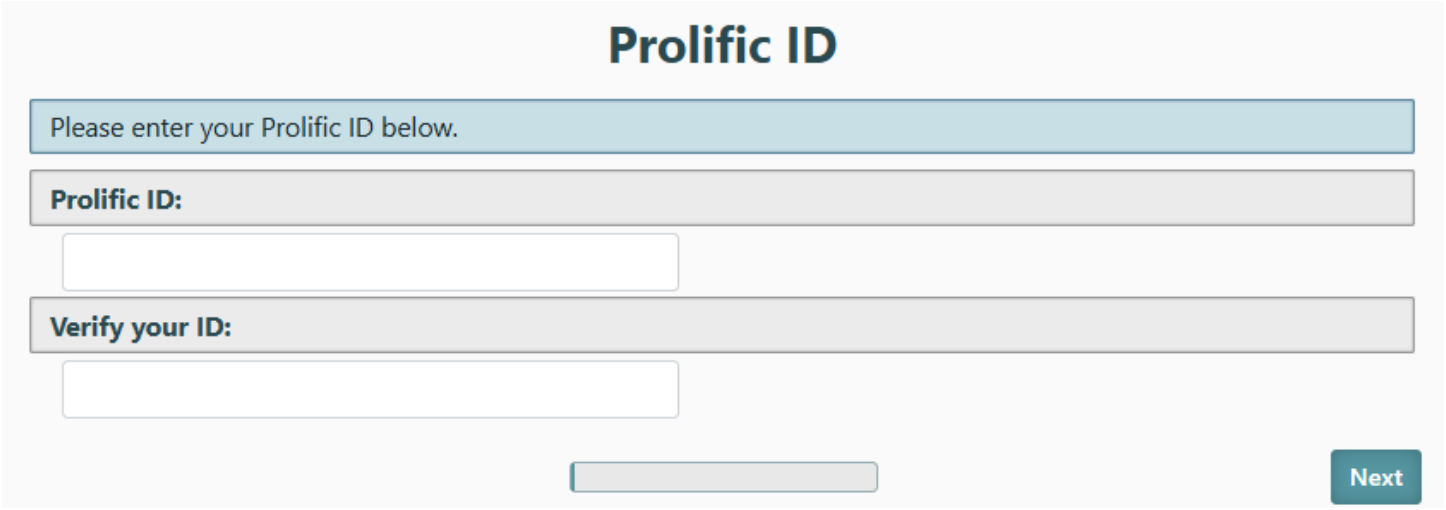
Contact Professor James C. Cox at 404-413-0200 or jccox@gsu if you have questions, concerns, or complaints about this study.

Consent

By clicking on the "I Consent to Participate" button on your screen below, you show your agreement to be in the experiment.

I Consent to Participate

Prolific ID Entry (*All Treatments*)



General Instructions (*Incentivized Treatments*)

General Instructions

You are now participating in a decision-making experiment.

Based on your decisions, you can earn a **bonus payment of $25** that will be paid to you via Prolific.

It is important that you understand all instructions before making your choices in this experiment.

Next

General Instructions

You must complete all four parts of the experiment in order to receive payment:

1. **Pre-Task Survey:** A survey about you and your health
2. **Decision Tasks:** About 10 tasks where you will make decisions related to your beliefs about certain things
3. **Post-Task Survey:** A survey about your healthcare use
4. **Bonus Payment Opportunity:** A chance to earn a bonus payment, where you will be scored and paid based on the accuracy of your decisions in the Decision Tasks

In addition to your participation payment of $8, **you can earn up to $25 from the decision-making tasks.**

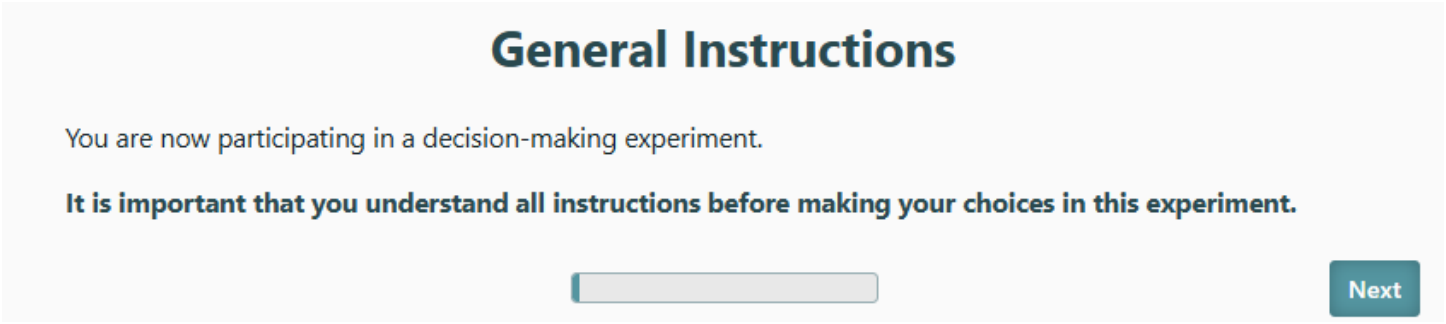
The Decision Tasks and your potential earnings will be explained in detail after the Pre-Task Survey.

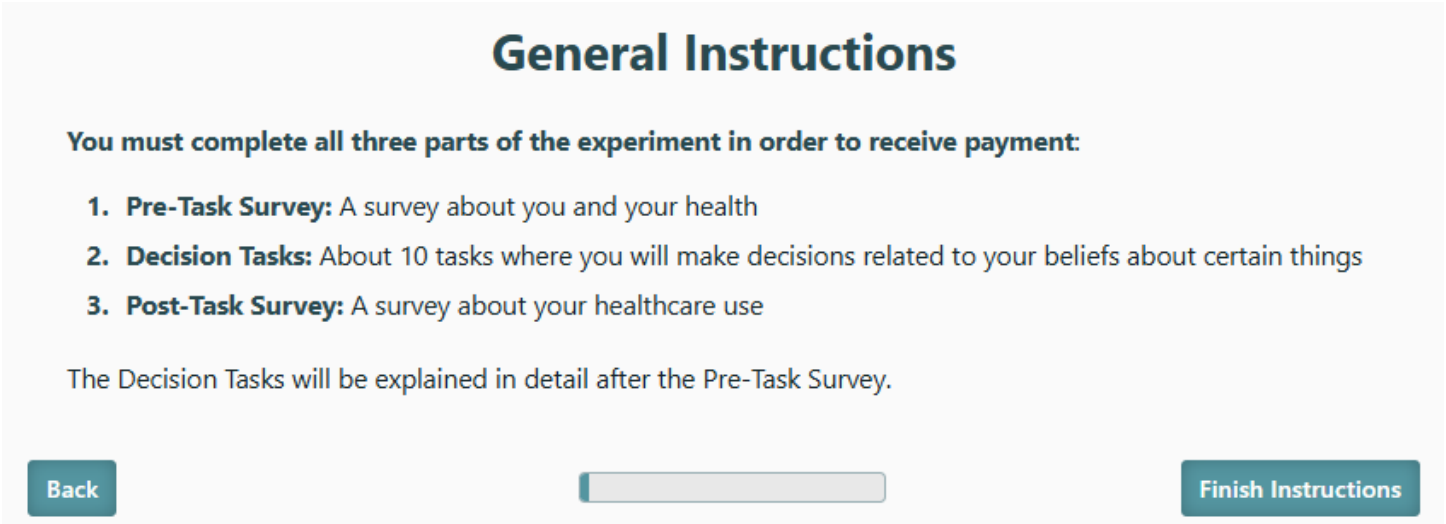
You will receive a bonus payment from one randomly-selected decision task, so you should think carefully about your answer to each question, as any question could be chosen for your payment.

Back

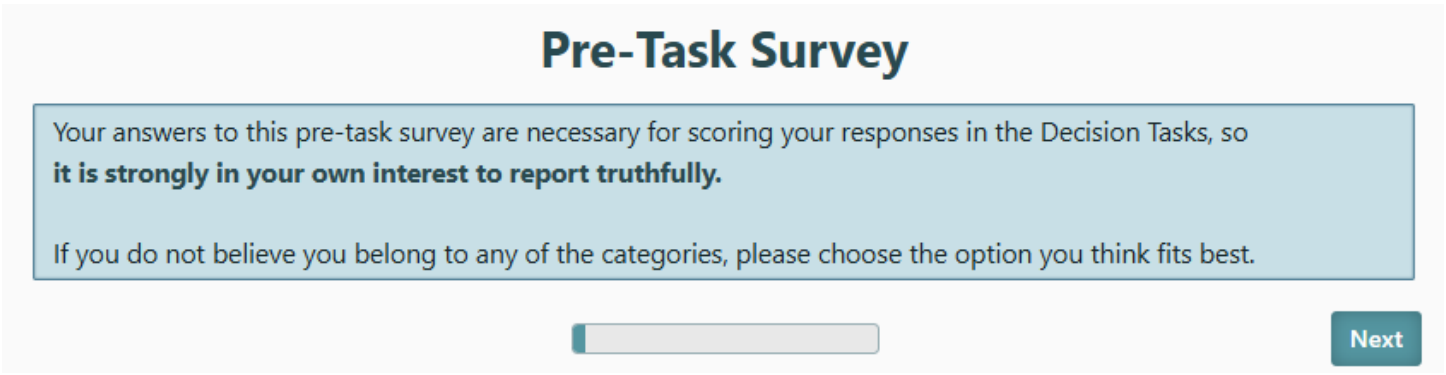
Finish Instructions

General Instructions (*Unincentivized Treatments*)





Pre-Task Survey (*All Treatments*)



Pre-Task Survey: Health Condition Knowledge

In this experiment, we will ask about your beliefs about various health conditions. Before you complete the tasks, we would like to know how familiar these conditions are to you. Below is a list of health conditions and their definition.

Please **read each definition** and **select the number that matches your familiarity** with the causes, risk factors, and symptoms of the condition:

1) **Not at all knowledgeable:** I am not familiar with this health condition at all.
2) **Somewhat knowledgeable:** I know a little bit about this health condition.
3) **Knowledgeable:** I have a good understanding of this health condition.
4) **Very knowledgeable:** I am highly familiar with this health condition.

Condition	Definition	(None) 1	2	3	(Very) 4
High Blood Pressure	When the force of blood against the artery walls is consistently too high. Also called hypertension.	○	○	○	○
High Cholesterol	Too much buildup of fats in the blood, potentially leading to the growth of plaques and reduced blood flow.	○	○	○	○
Heart Disease	Diseases affecting blood supply to the heart, or the heart's ability to supply blood to the rest of the body. These include heart attack (also called myocardial infarction), coronary heart disease, angina (also called angina pectoris), and congestive heart failure.	○	○	○	○
Stroke	Poor blood flow to the brain, causing cell death. This includes both ischemic stroke due to lack of blood flow, and hemorrhagic stroke due to bleeding.	○	○	○	○
Diabetes	Prolonged high blood sugar levels caused by the pancreas not producing enough insulin or the body not responding properly to the insulin produced. Also called sugar diabetes.	○	○	○	○
Cancer	Diseases in which cells in the body grow out of control. Unless specified, this includes any cancer *except skin cancer.* We provide a separate definition for cancers specifically from the skin.	○	○	○	○
Breast Cancer	Any cancer that starts specifically in the breast.	○	○	○	○
Skin Cancer	Cancers that start in the skin, including melanoma, basal-cell, and squamous-cell skin cancers.	○	○	○	○
Arthritis	Disorders that primarily cause inflammation and swelling in the joints. This includes osteoarthritis, degenerative arthritis, rheumatoid arthritis, and psoriatic arthritis.	○	○	○	○

Next

Pre-Task Survey: Sociodemographic Characteristics

The next few questions are about yourself and your sociodemographic background.

What is your **biological sex** (at birth)?

- ○ Male
- ○ Female

What is your age?

Which of the following best describes your **race and ethnicity**?

- ○ White, non-Hispanic
- ○ Black, non-Hispanic
- ○ Hispanic
- ○ Asian American or Pacific Islander
- ○ American Indian or Alaska Native
- ○ Other

What is the **highest grade or level of school** you have completed or the **highest degree** you have received?

- ○ Less than 9th grade
- ○ Some High School
- ○ High School graduate, GED or equivalent
- ○ Some college or Associate's Degree
- ○ College graduate or above

In what range is your **total yearly household income** ?

- ○ $0 to $24,999
- ○ $25,000 to $49,999
- ○ $50,000 to $74,999
- ○ $75,000 or more

What is your **current marital status**?

- ○ Married
- ○ Widowed
- ○ Divorced
- ○ Separated
- ○ Never Married
- ○ Living with Partner

Are you covered by health insurance or some other kind of health care plan? (This includes health insurance through your job, or through government programs like Medicare or Medicaid).

- ○ Yes
- ○ No

During the **past 12 months, how many times have you seen a doctor** or other health care professional about your health? Do not include times you were hospitalized overnight.

- ○ None
- ○ 1 time
- ○ 2 to 3 times
- ○ 4 to 9 times
- ○ 10 or more

Please select your **preferred unit measurement system**:

- ○ U.S. System: inches, pounds
- ○ Metric System: centimeters, kilograms

Next

Pre-Task Survey: Health

The next few questions are about your general health.

How would you say your health is in general?

○ Poor ○ Fair ○ Good ○ Very good ○ Excellent

How tall are you without shoes?

feet inches

How much do you weigh without clothes or shoes?

pounds

Up to the present time, what is the **most you have ever weighed**?

pounds

13. What is your **current waist circumference**?

- ○ Less than 37 inches
- ○ 37 to 40 inches
- ○ Greater than 40 inches

The next few questions are about your personal consumption habits:

Question	Yes	No
In your entire life, have you had at least 12 alcoholic drinks (of any kind of alcohol, not counting small tastes or sips)?	○	○
Was there ever a time in your life when you drank **5 or more alcoholic drinks almost every day** ?	○	○
Have you smoked at least 100 cigarettes in your entire life?	○	○
Do you now smoke cigarettes every day or some days?	○	○
Have you ***ever* taken medication for controlling your blood pressure**?	○	○
Do you **eat fruits and vegetables every day**?	○	○

The next few questions are about your physical activity in a typical week, during any part of your routine.

Does your **typical week** include any of the following:

Question	Yes	No
At least **4 hours of physical activity**?	○	○
Activities that require **moderate physical effort** (cause *small* increases in breathing or heart rate for at least 10 minutes straight)?	○	○
Activities that require **hard physical effort** (cause *large* increases in breathing or heart rate for at least 10 minutes straight)?	○	○

Next

Pre-Task Survey: Health Conditions

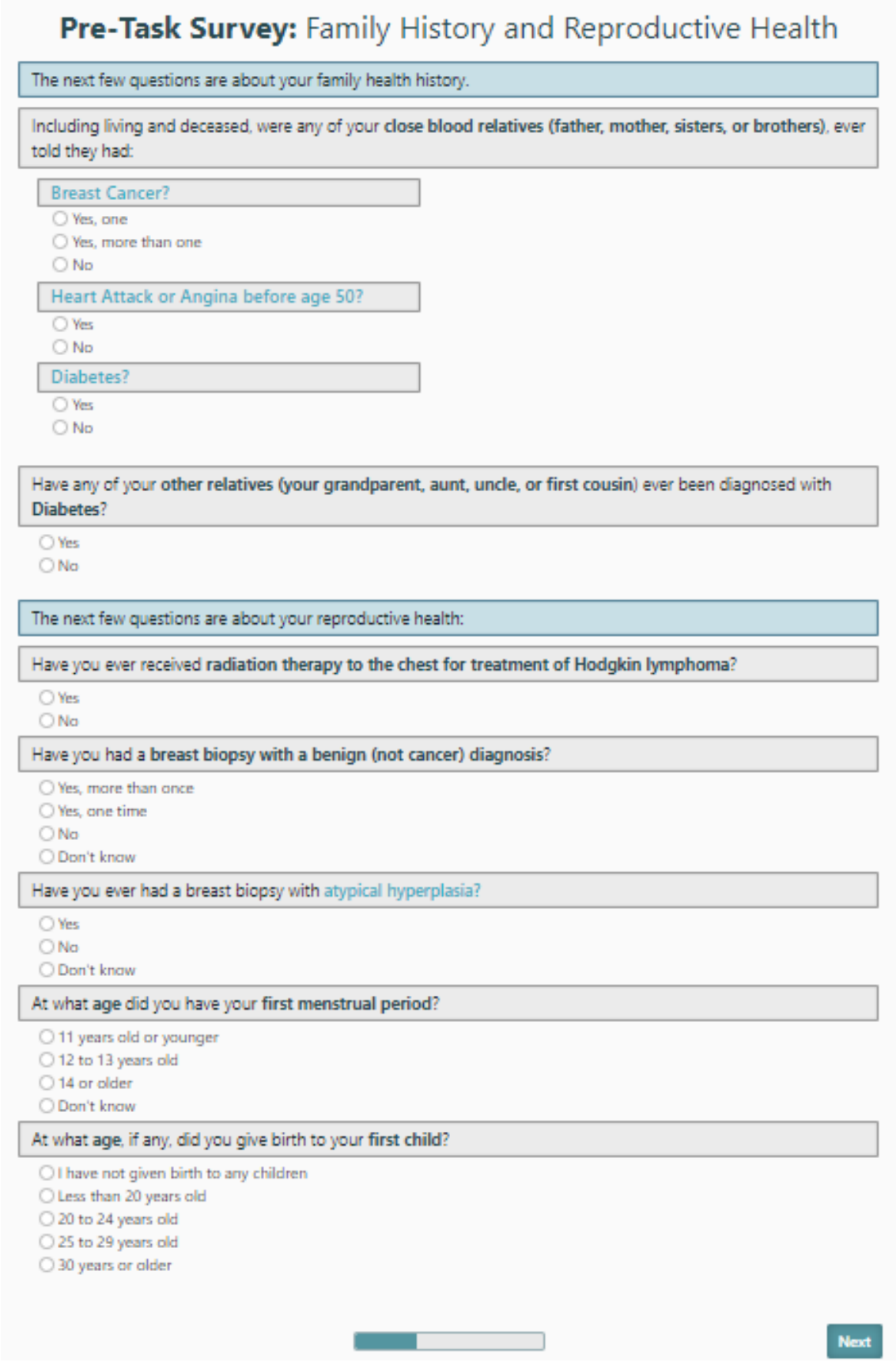
The next few questions are about your current and past health conditions.

Click on a condition name to review its definition.

Has a doctor ever told you that you had any of the following:

Condition	Yes	No
High Blood Pressure	○	○
High Cholesterol	○	○
High Blood Sugar	○	○
Diabetes	○	○
Heart Disease	○	○
Stroke	○	○
Arthritis	○	○
Cancer (other than skin cancer)	○	○
Skin Cancer	○	○

Next



Reproductive health questions are only displayed for respondents who select Female for their biological sex on the Pre-Task Survey: Sociodemographic Characteristics page.

Decision Task Instructions (*Treatment I100*)

Decision Task Instructions

Bonus Payment and Scoring

We will now move on to the Decision Tasks. In these tasks, you will be **scored and paid according to how accurate your beliefs are** about certain things relating to your health.

You can earn a bonus payment of up to **$25** in this experiment. The chance that you will earn $25 depends on the accuracy of your answer, scored in points.

The more points you score, the greater your chance of being paid $25.

At the end of the experiment, your scores for each question will be shown to you, and **your score from one Decision Task will be chosen at random for payment.** To be paid for this task, you will draw a **random number from 1 to 100**, with every number being equally likely.

If the random number you draw is:

- **less** than or equal to your score, you earn **$25**
- **greater** than your score, you earn **$5**

Next

People Like You

When we say that a question is about "**people like you**," this means that the correct answer has been **personalized to you based on your survey responses.**

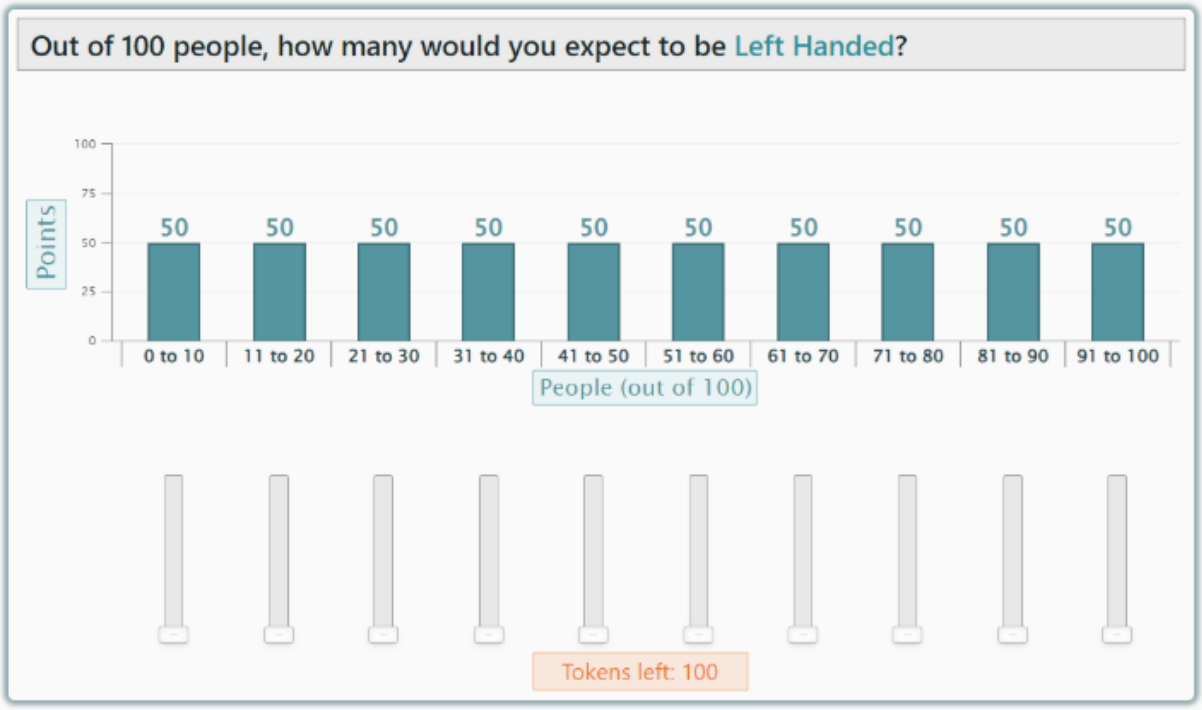
Remember that you answered questions about the following in the Pre-Task Survey:

- Age
- Height and Weight
- Gender
- Diet, Alcohol, and Smoking
- Physical Activity
- Family Health History
- Personal Health Conditions
- Income
- Education
- Race and Ethnicity
- Marital Status
- Insurance Coverage and Healthcare Use

These will all be used to calculate the correct answer to each decision task, using data from the National Health and Nutrition Examination Survey, advanced statistical methods, and medical diagnostic tools.

Next

You will make your decisions by placing **tokens** in bins. Below is an example picture of a Decision Task:



You have 10 sliders to adjust and 100 tokens to allocate to reflect your belief about the answer. **You must use all 100 tokens in order to submit your decision.** As you adjust sliders, the bars showing potential scores on the screen will change. Each bar here shows the points you earn if the correct answer is in the range shown under the bar.

Where you position each slider depends on your beliefs about the correct answer to the question, and your confidence in those beliefs. For the above question, the tokens you placed in each bin would reflect your beliefs about the frequency of left-handedness. The first slider represents your belief that, on average, 0 to 10 out of 100 people are left-handed. The second slider matches your belief that between 11 and 20 people out of 100 are left-handed on average, and so on.

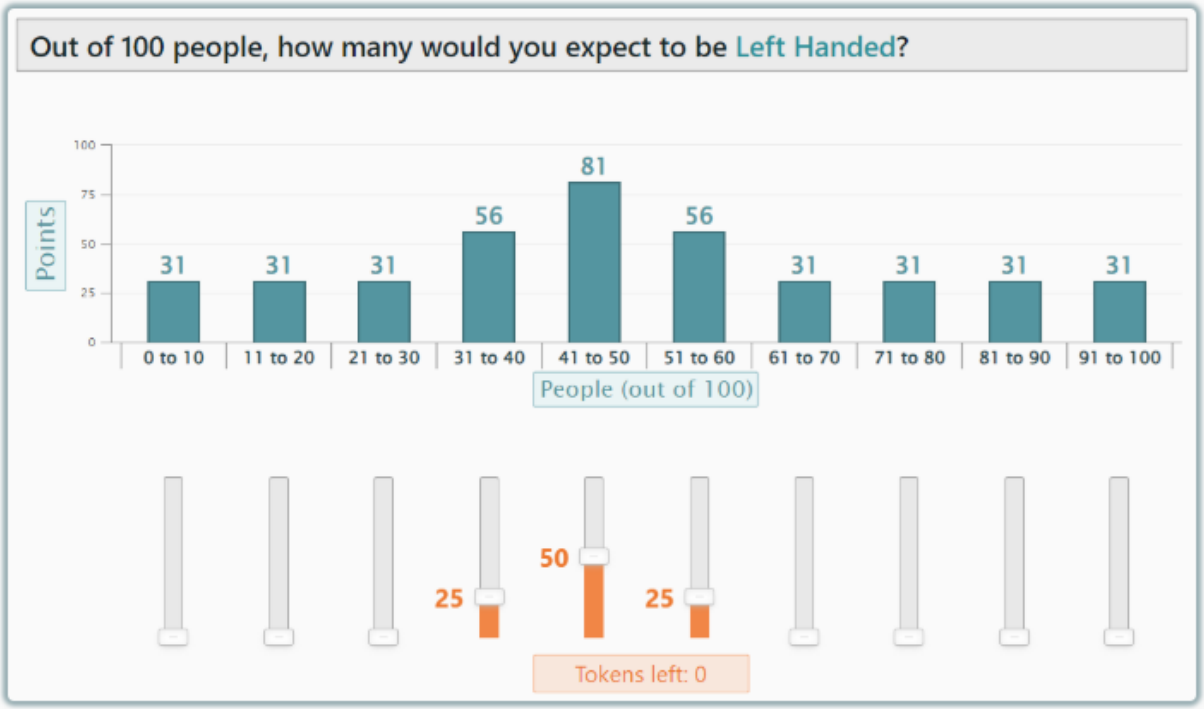
Back

Next

Example 1

To show how to use these sliders, suppose you believe there is a fair chance the true answer is between 41 and 50 out of 100, but it may be a bit lower or higher. Then you might place your 100 tokens in the following way:

- 25 **tokens** to the bin **"31-40"**
- 50 **tokens** to the bin **"41-50"**
- 25 **tokens** to the bin **"51-60"**



From the picture above, you can see that if you placed your tokens this way, you would score:

- 81 **points** if the correct answer is between **41 and 50**
- 56 **points** if the correct answer is between **31 and 40**, or between **51 and 60**
- 31 **points** if the correct answer is **any other number**

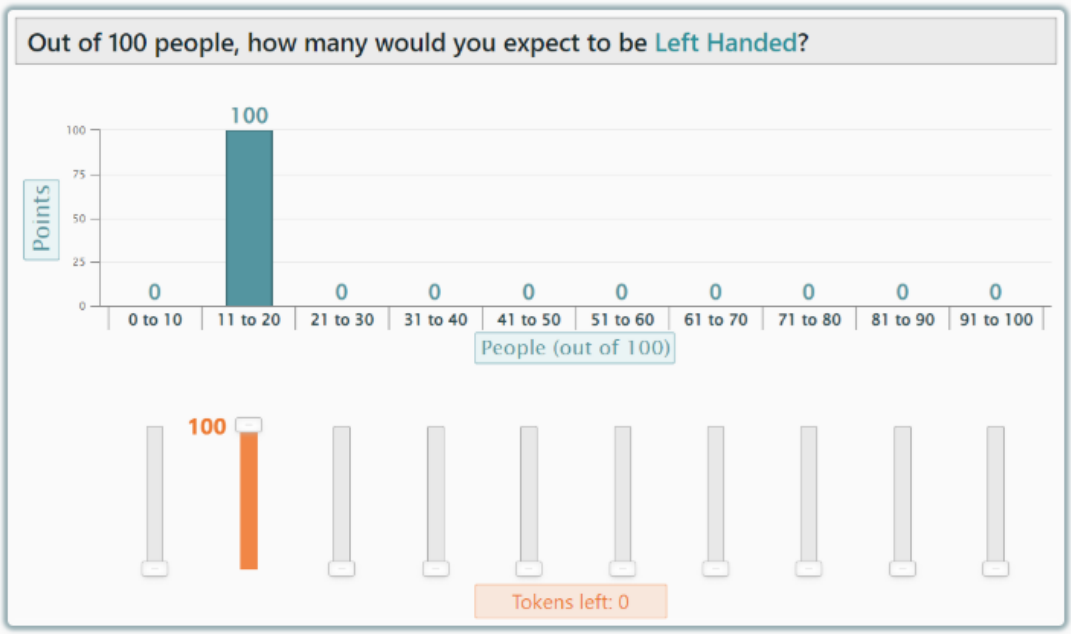
You can adjust the tokens as much as you want to best reflect your personal beliefs and your confidence. For example, if you felt you had truly no idea about the true answer, you could spread your tokens evenly, placing 10 tokens in each bin. Your earnings depend on your reported beliefs and, of course, the true answer.

Example 2

Suppose you "put all of your eggs in one basket" and placed all 100 **tokens** in the bin **"11-20"**.

Then you would have faced the earnings outcomes shown below. If the true answer were 12 (which is the average number of left-handed people out of 100, according to a 2007 study by researchers at the National Bureau of Economic Research), then you would earn the bonus payment of $25 for sure.

Note if you instead placed all 100 tokens in any other bin, your score would be 0 points, and you would receive $5 for sure.



Back

Next

Summary

Keep in mind these important points when making your decisions:

- **The decisions you make are a matter of personal choice,** and your beliefs are judgments based only on your own experience and knowledge. It is up to you to balance the strength of your personal beliefs with the risk of them being wrong.
- **Correct answers are calculated specifically for people like you, using your responses to the Pre-Task Survey,** and will be shown to you at the end of the experiment.
- **You can only earn a bonus payment of either $25 or $5,** depending on your choices and the correct answer for one randomly-selected task.
- **More points increase your chance of being paid $25.** The points you earn will be compared with the outcome of the random number drawn at the end of the experiment.

You have finished the Decision Task Instructions, and **we will now move on to the paid Decision Tasks**.

Finish Instructions

Decision Task Instructions (*Treatment N100*)

Scoring

We will now move on to the Decision Tasks. In these tasks, you will be **scored according to how accurate your beliefs are** about certain things relating to your health.

The more points you score, the greater the accuracy of your belief.

At the end of the experiment, your scores for each question will be shown to you.

Next

People Like You

When we say that a question is about "**people like you**," this means that the correct answer has been **personalized to you based on your survey responses.**

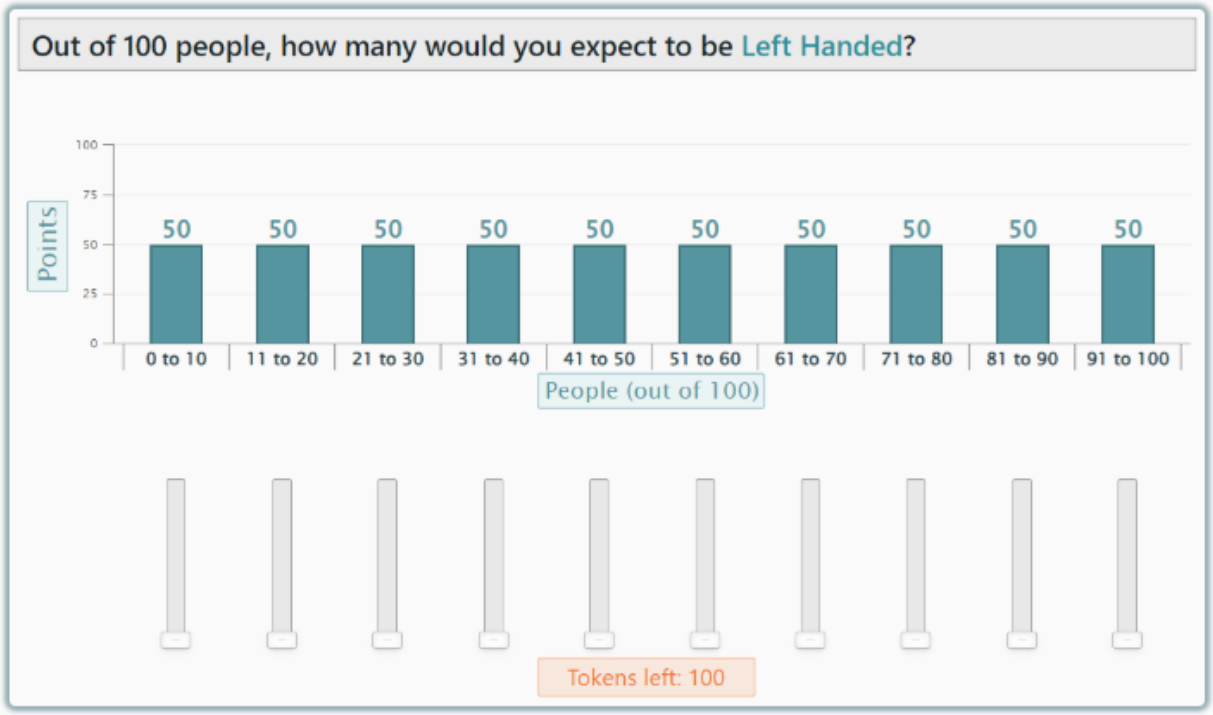
Remember that you answered questions about the following in the Pre-Task Survey:

- Age
- Height and Weight
- Gender
- Diet, Alcohol, and Smoking
- Physical Activity
- Family Health History
- Personal Health Conditions
- Income
- Education
- Race and Ethnicity
- Marital Status
- Insurance Coverage and Healthcare Use

These will all be used to calculate the correct answer to each decision task, using data from the National Health and Nutrition Examination Survey, advanced statistical methods, and medical diagnostic tools.

Next

You will make your decisions by placing **tokens** in bins. Below is an example picture of a Decision Task:



You have 10 sliders to adjust and 100 tokens to place to reflect your belief about the answer. **You must use all 100 tokens in order to submit your decision.** As you adjust sliders, the bars showing potential scores on the screen will change. Each bar here shows the points you earn if the correct answer is in the range shown under the bar.

Where you position each slider depends on your beliefs about the correct answer to the question, and your confidence in those beliefs. For the above question, the tokens you place in each bin would reflect your beliefs about the frequency of left-handedness. The first slider represents your belief that, on average, 0 to 10 out of 100 people are left-handed. The second slider matches your belief that between 11 and 20 people out of 100 are left-handed on average, and so on.

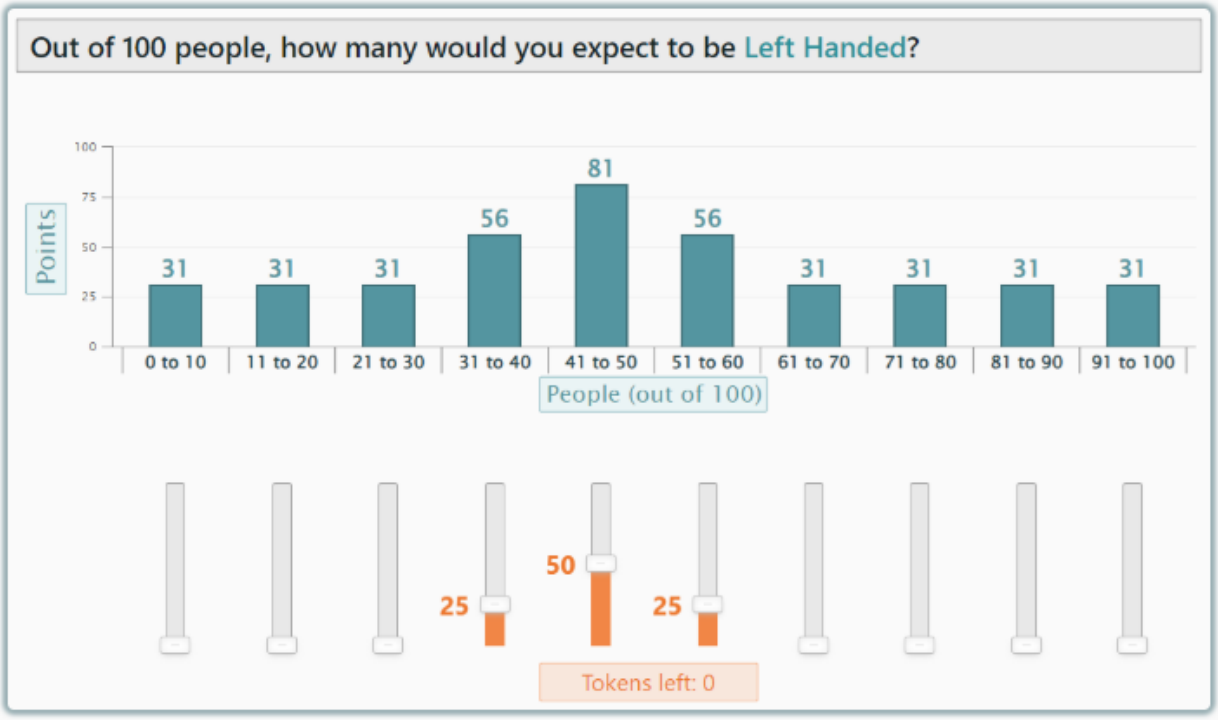
Back

Next

Example 1

To show how to use these sliders, suppose you believe there is a fair chance the true answer is between 41 and 50 out of 100, but it may be a bit lower or higher. Then you might place your 100 tokens in the following way:

- 25 **tokens** to the bin **"31-40"**
- 50 **tokens** to the bin **"41-50"**
- 25 **tokens** to the bin **"51-60"**



From the picture above, you can see that if you placed your tokens this way, you would score:

- 81 **points** if the correct answer is between **41 and 50**
- 56 **points** if the correct answer is between **31 and 40**, or between **51 and 60**
- 31 **points** if the correct answer is **any other number**

You can adjust the tokens as much as you want to best reflect your personal beliefs and your confidence. For example, if you felt you had no idea about the true answer, you could spread your tokens evenly, placing 10 tokens in each bin. Your scores depend on your reported beliefs and, of course, the true answer.

Back

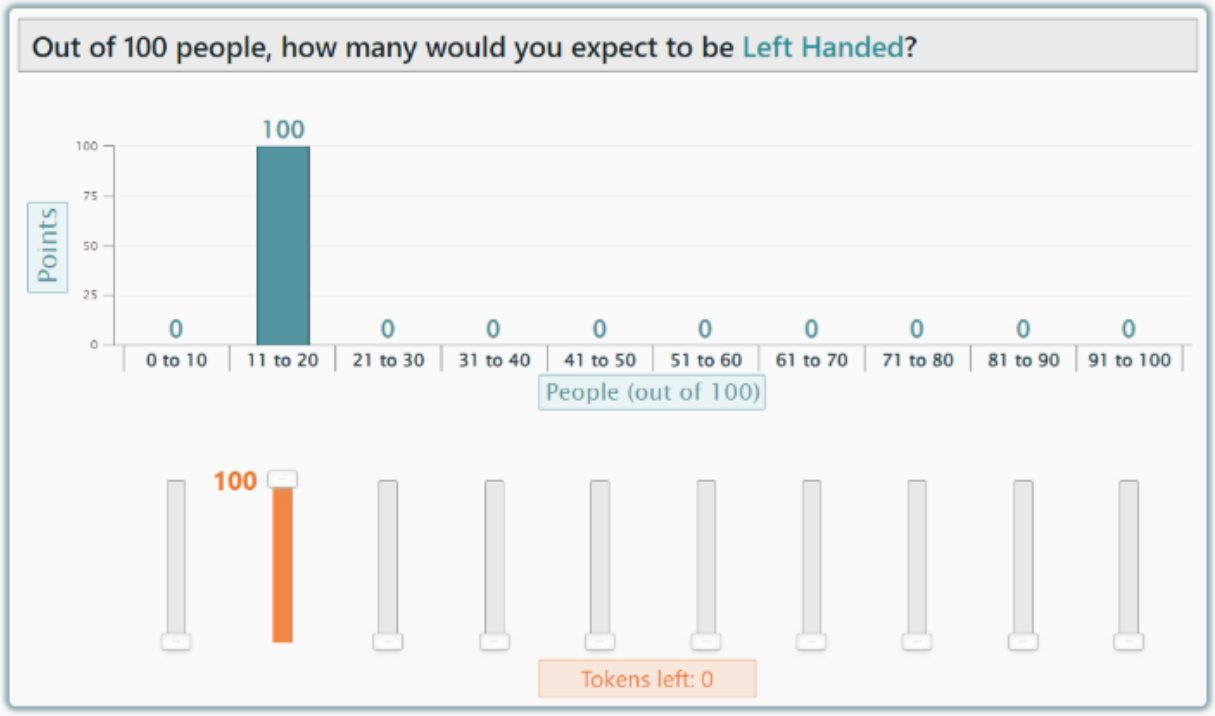
Next

Example 2

Suppose you "put all of your eggs in one basket" and placed all 100 **tokens** in the bin **"11-20"**.

If the true answer were 12 (which is the average number of left-handed people out of 100, according to a 2007 study by researchers at the National Bureau of Economic Research), then you would score 100 points for sure.

Note if you instead placed all 100 tokens in any other bin, your score would be 0 points for certain.



Back

Next

Summary

Keep in mind these important points when making your decisions:

- **The decisions you make are a matter of personal choice,** and your beliefs are judgments based only on your own experience and knowledge.
- **Correct answers are calculated specifically for people like you, using your responses to the Pre-Task Survey,** and will be shown to you at the end of the experiment.

You have finished the Decision Task Instructions, and **we will now move on to the real Decision Tasks**.

Finish Instructions

Decision Task Instructions (*Treatment I1*)

Bonus Payment and Scoring

We will now move on to the Decision Tasks. In these tasks, you will be **scored and paid according to how accurate your beliefs are** about certain things relating to your health.

You can earn a bonus payment of up to **$25** in this experiment. The chance that you will earn $25 depends on the accuracy of your answer.

At the end of the experiment, your scores for each question will be shown to you, and **one Decision Task will be chosen at random for payment.**

- If your answer to this task is **correct**, you will earn a bonus payment **$25**
- If you are **incorrect**, you will earn a bonus payment of **$5**

Next

People Like You

When we say that a question is about "**people like you**," this means that the correct answer has been **personalized to you based on your survey responses.**

Remember that you answered questions about the following in the Pre-Task Survey:

- Age
- Height and Weight
- Gender
- Diet, Alcohol, and Smoking
- Physical Activity
- Family Health History
- Personal Health Conditions
- Income
- Education
- Race and Ethnicity
- Marital Status
- Insurance Coverage and Healthcare Use

These will all be used to calculate the correct answer to each decision task, using data from the National Health and Nutrition Examination Survey, advanced statistical methods, and medical diagnostic tools.

Next

You will make your choices by placing **tokens** in bins. Below is an example picture of a Decision Task:

You have 10 sliders to adjust and 1 token to place to reflect your belief about the answer. **You must use your token to submit your decision.**

Where you place your token depends on your beliefs about the correct answer to the question.

For the above question, the token would reflect your beliefs about the frequency of left-handedness. The first slider represents your belief that, on average, 0 to 10 out of 100 people are left-handed. The second slider matches your belief that between 11 and 20 people out of 100 are left-handed on average, and so on.

Back

Next

Example

To show how to use these sliders, suppose you believe the true frequency of left-handedness is between 11 and 20. You could then place your **token** in the bin **"11-20"**.

If the true answer were 12 (which is the average number of left-handed people out of 100, according to a 2007 study by researchers at the National Bureau of Economic Research), then you would earn a bonus payment of $25 for sure.

Note if you instead placed your token in any other bin you would earn a bonus payment of $5 for sure.

Back

Next

Summary

Keep in mind these important points when making your decisions:

- **The decisions you make are a matter of personal choice,** and your beliefs are judgments based only on your own experience and knowledge. It is up to you to balance the strength of your personal beliefs with the risk of them being wrong.
- **Correct answers are calculated specifically for people like you, using your responses to the Pre-Task Survey,** and will be shown to you at the end of the experiment.
- **You can only earn a bonus payment of either $25 or $5,** depending on your choices and the correct answer for one randomly-selected task.

You have finished the Decision Task Instructions, and **we will now move on to the paid Decision Tasks**.

Finish Instructions

Decision Task Instructions (*Treatment N1*)

Decision Task Instructions

Scoring

We will now move on to the Decision Tasks. In these tasks, you will be **scored according to how accurate your beliefs are** about certain things relating to your health.

At the end of the experiment, **the correct answer for each question will be shown to you.**

Next

People Like You

When we say that a question is about "**people like you**," this means that the correct answer has been **personalized to you based on your survey responses.**

Remember that you answered questions about the following in the Pre-Task Survey:

- Age
- Height and Weight
- Gender
- Diet, Alcohol, and Smoking
- Physical Activity
- Family Health History
- Personal Health Conditions
- Income
- Education
- Race and Ethnicity
- Marital Status
- Insurance Coverage and Healthcare Use

These will all be used to calculate the correct answer to each decision task, using data from the National Health and Nutrition Examination Survey, advanced statistical methods, and medical diagnostic tools.

Next

You will make your decisions by placing **tokens** in bins. Below is an example picture of a Decision Task:

You have 10 sliders to adjust and 1 token to place to reflect your belief about the answer. **You must use your token to submit your decision.**

Where you place your token depends on your beliefs about the correct answer to the question.

For the above question, the token would reflect your beliefs about the frequency of left-handedness. The first slider represents your belief that, on average, 0 to 10 out of 100 people are left-handed. The second slider matches your belief that between 11 and 20 people out of 100 are left-handed on average, and so on.

Back Next

Example

To show how to use these sliders, suppose you believe the true frequency of left-handedness is between 11 and 20. You could then place your **token** in the bin **"11-20"**.

If the true answer were 12 (which is the average number of left-handed people out of 100, according to a 2007 study by researchers at the National Bureau of Economic Research), then you would earn 100 points for sure.

Note if you instead placed your token in any other bin, your score would be 0 points for sure.

Back Next

Summary

Keep in mind these important points when making your decisions:

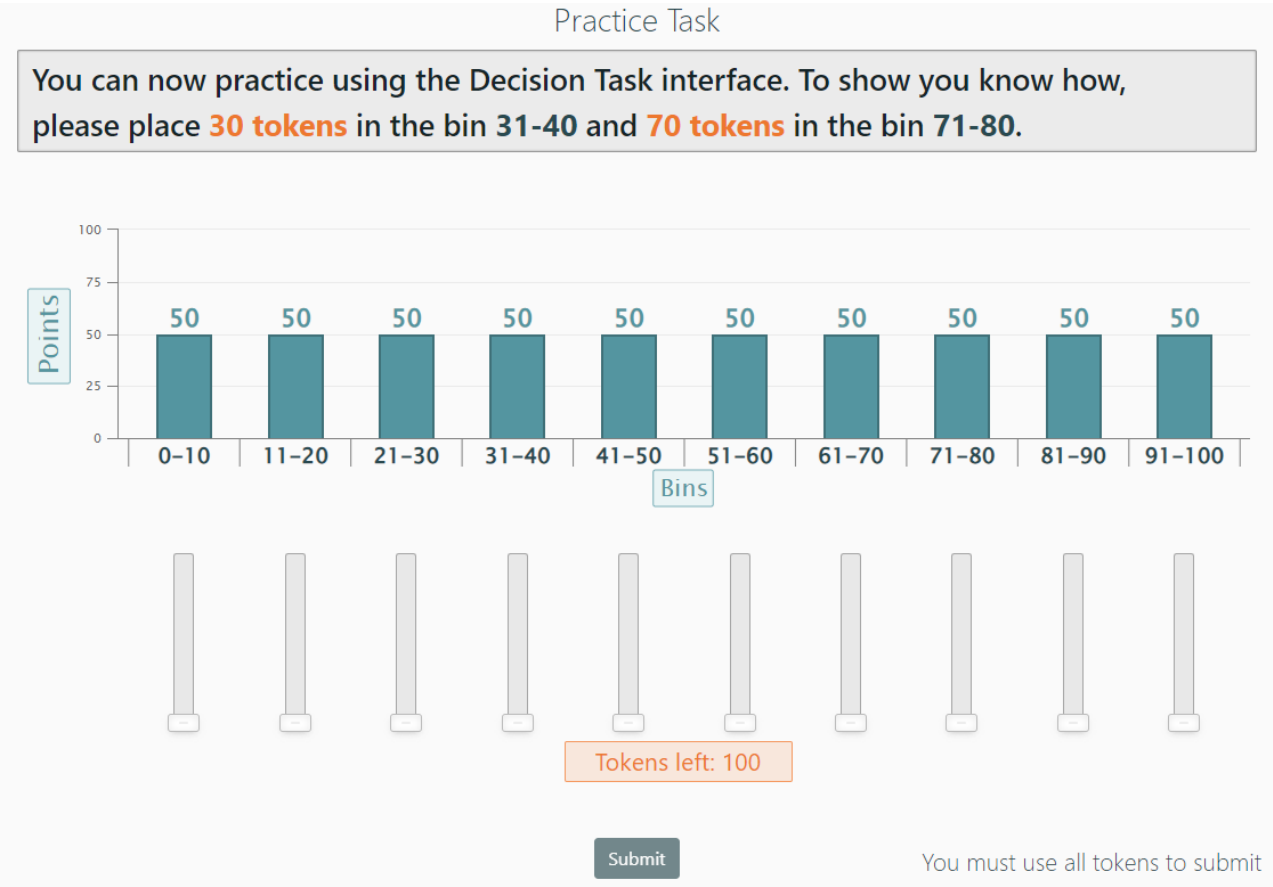
- **The decisions you make are a matter of personal choice,** and your beliefs are judgments based only on your own experience and knowledge.
- **Correct answers are calculated specifically for people like you, using your responses to the Pre-Task Survey,** and will be shown to you at the end of the experiment.

You have finished the Decision Task Instructions, and **we will now move on to the real Decision Tasks**.

Finish Instructions

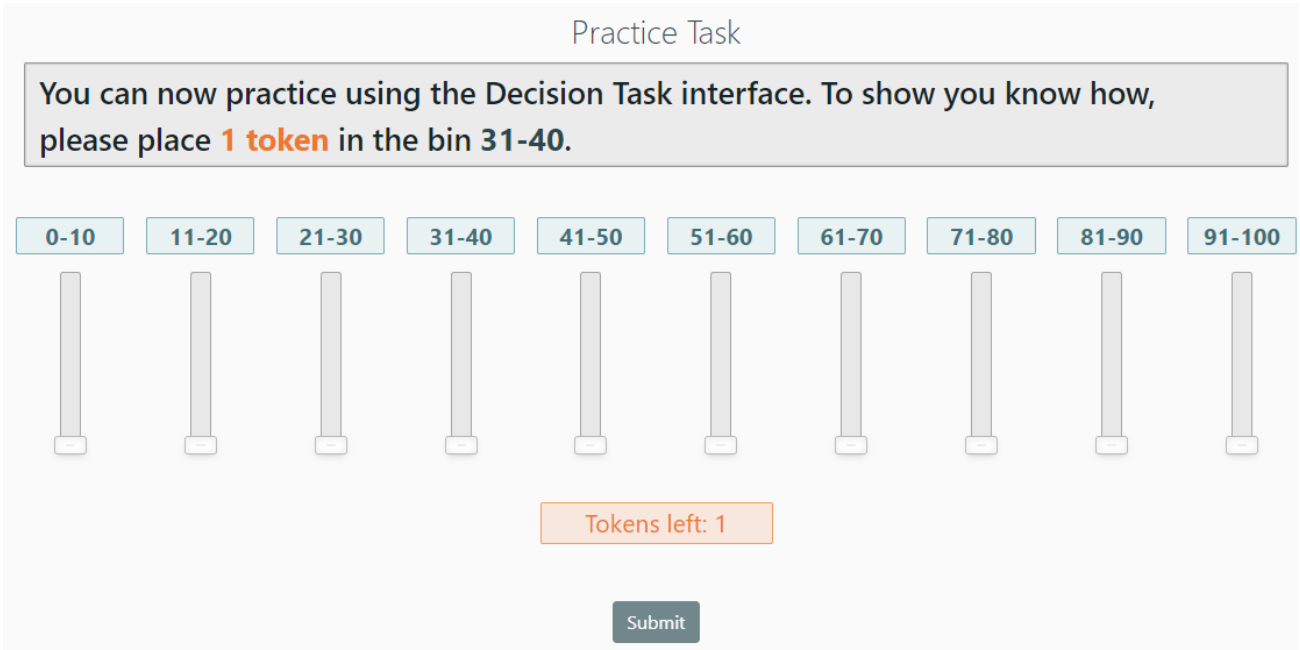
Decision Task Comprehension Check

(Distributional Treatments)

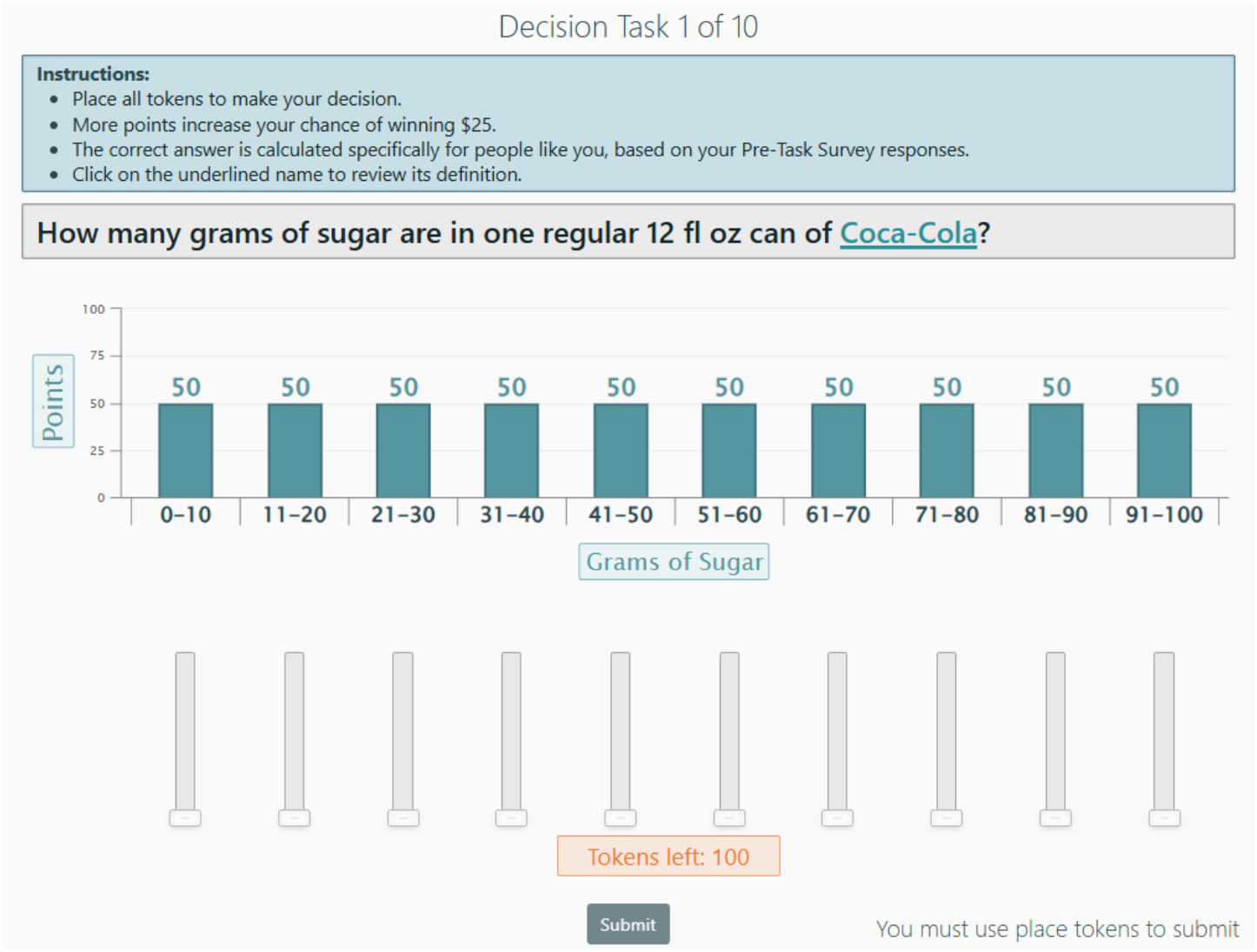


Decision Task Comprehension Check

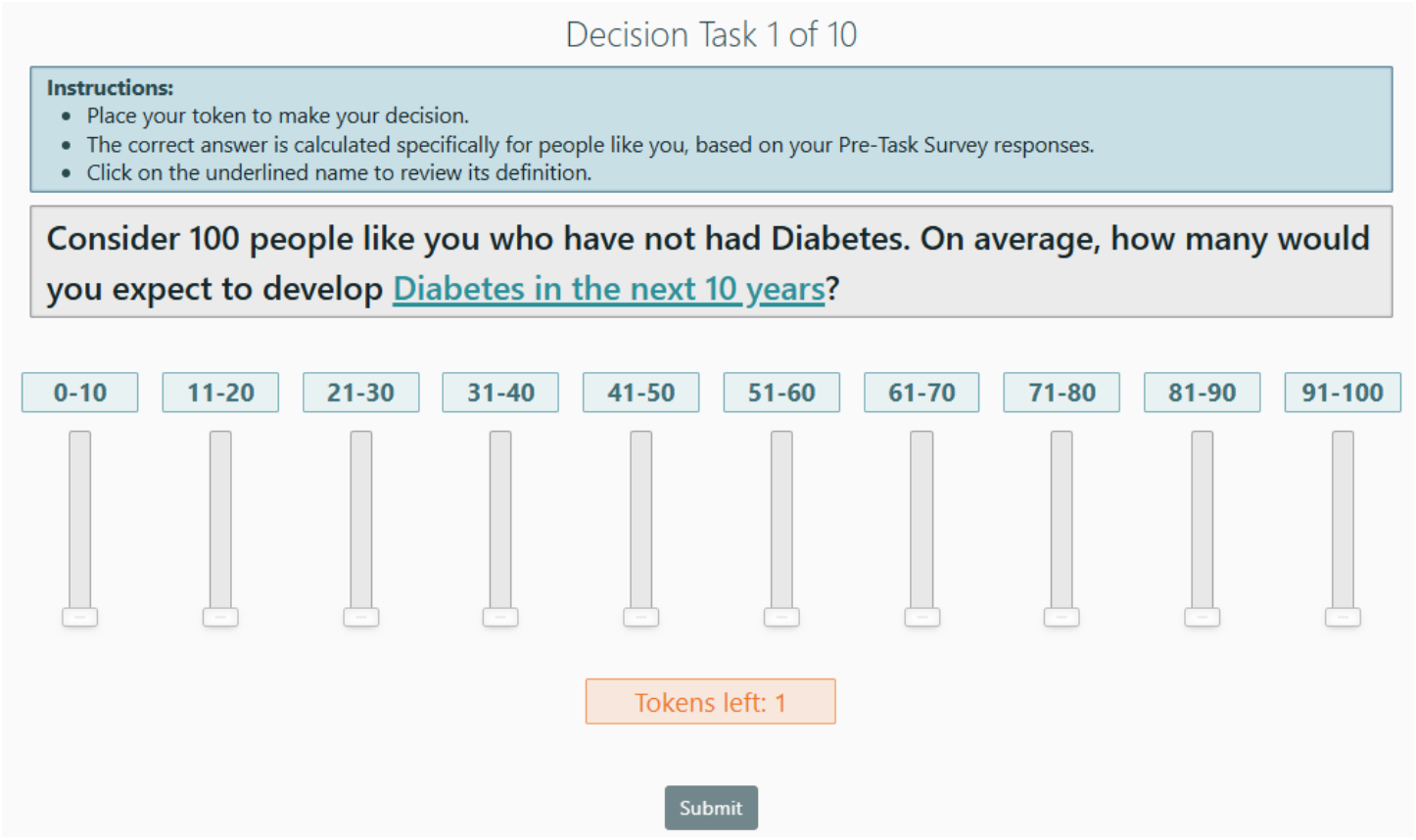
(Point Estimate Treatments)



Decision Task (*Distributional Treatments*)



Decision Task (*Point Estimate Treatments*)



Post-Task Survey (*All Treatments*)

Post-Task Survey: Health Care Use

The following post-task survey asks you about your healthcare use and your attitudes about health.

Please mark the time since your last screening or test for the following health conditions:

Condition	Less than 1 year	1 to 2 years	2 to 3 years	3 years or more	Never
High Cholesterol	○	○	○	○	○
High Blood Pressure	○	○	○	○	○
Diabetes or High Blood Sugar	○	○	○	○	○
Colon Cancer	○	○	○	○	○
Breast Cancer	○	○	○	○	○

On days when you expect to be outdoors for more than 15 minutes in the sun, **how frequently do you protect your skin** using sunscreen, protective clothing, or other measures?

- ○ Always
- ○ Often
- ○ Sometimes
- ○ Rarely
- ○ Never

Next

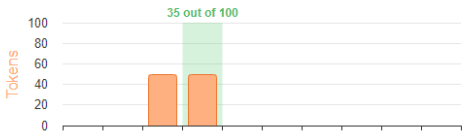
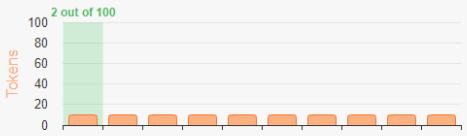
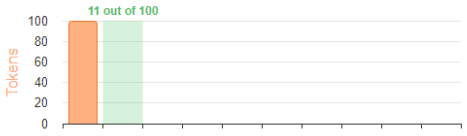
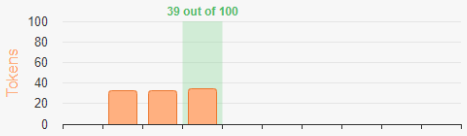
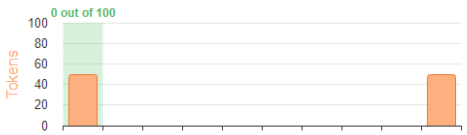
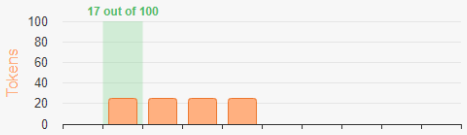
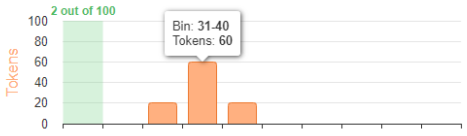
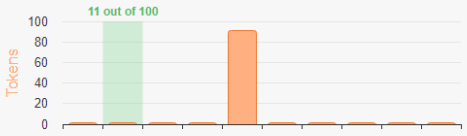
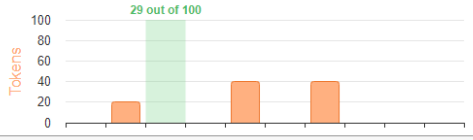
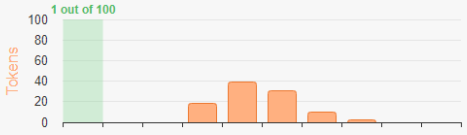
Post-Task Survey: Health Attitudes

Please mark your level of agreement with each statement:

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
My health is important to me.	○	○	○	○	○
I seek medical care and advice more often than others like me.	○	○	○	○	○
The consequences of having a health condition like diabetes, heart disease, or cancer would be severe for my daily life and well-being.	○	○	○	○	○
Taking preventative health actions significantly reduces my risk of developing health conditions.	○	○	○	○	○
I am confident in my own ability to reduce my risk of developing health conditions.	○	○	○	○	○
I am confident in the healthcare system's ability to help reduce my risk of developing health conditions.	○	○	○	○	○
There are significant barriers that prevent me from taking preventative health actions.	○	○	○	○	○

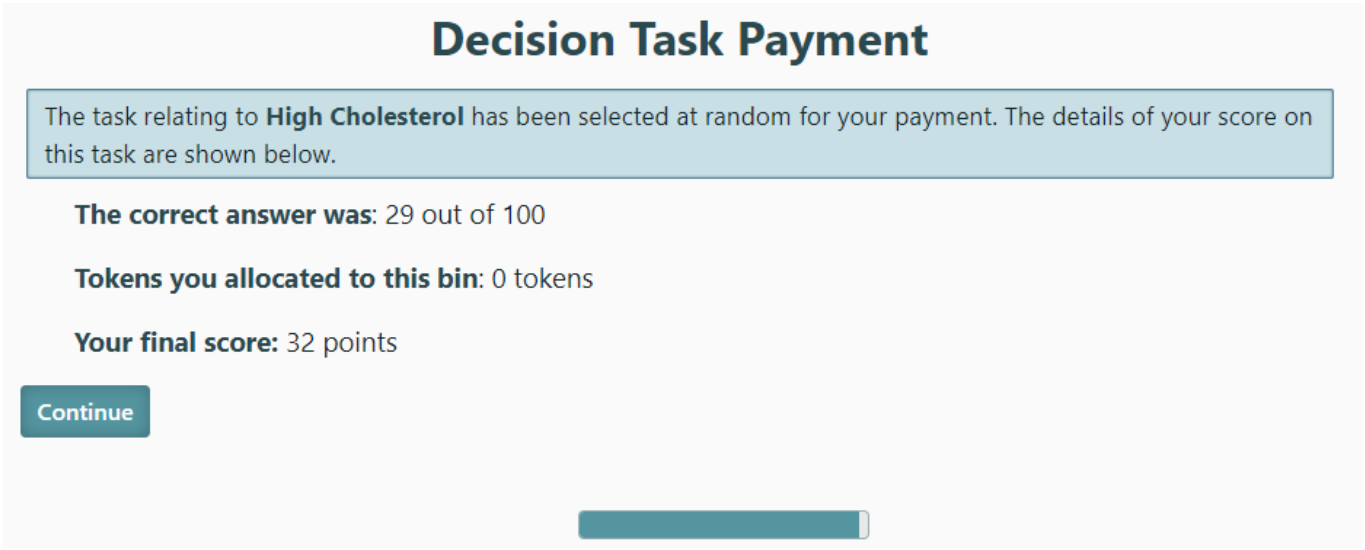
Next

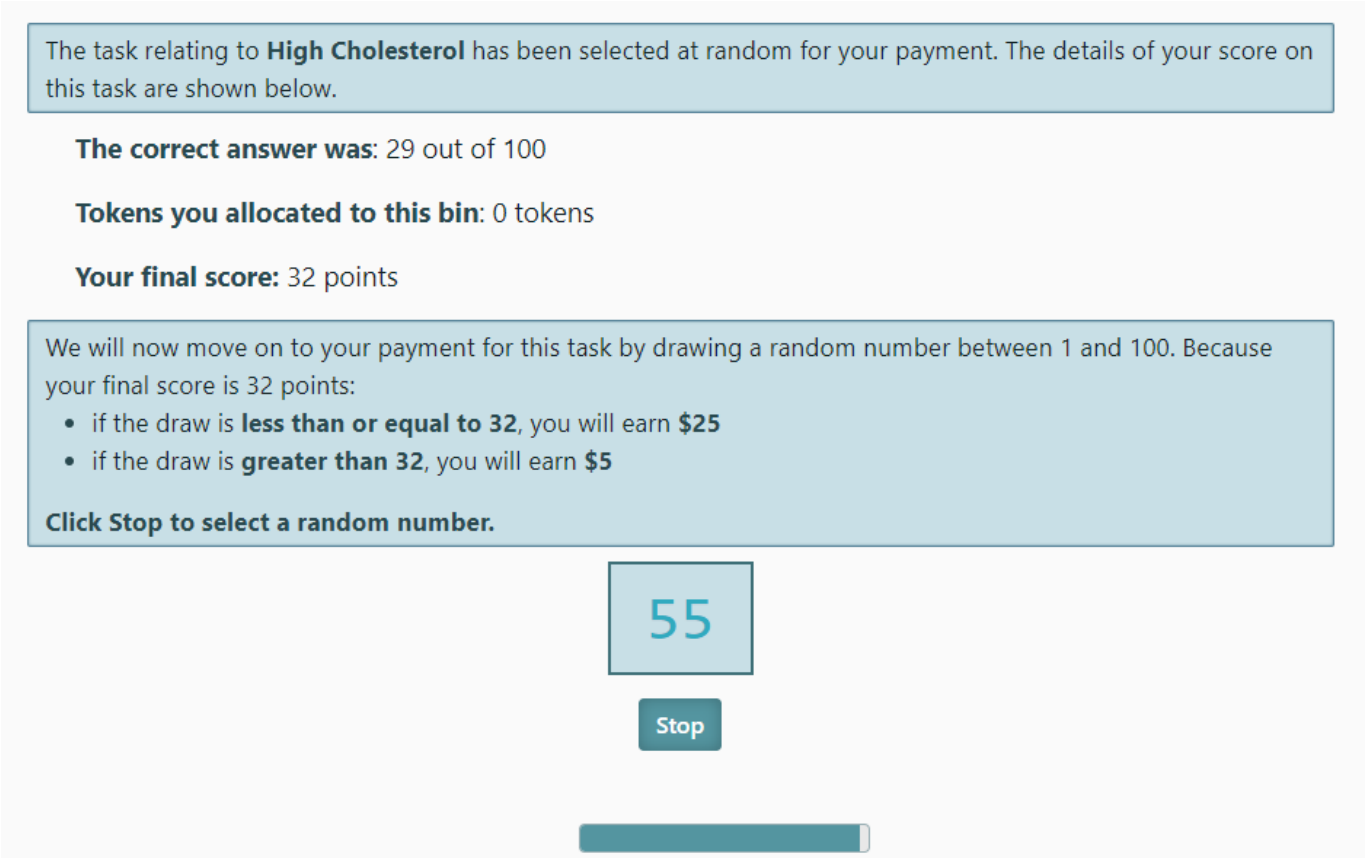
Decision Task Results (*All Treatments*)

Health Condition	Your Decision	Correct Answer (out of 100)	Your Allocation (tokens)	Your Score (points)
High Blood Pressure		35	50	75
Heart Disease		2	10	55
Diabetes		11	0	0
Coca-Cola		39	34	67
Skin Cancer		0	50	75
Diabetes in the next 10 years		17	25	63
Cancer (other than skin cancer)		2	0	28
Arthritis		11	1	10
High Cholesterol		29	0	32
Stroke		1	0	35

Next

Decision Task Payment (*Incentivized Treatments*)





The task relating to **High Cholesterol** has been selected at random for your payment. The details of your score on this task are shown below.

The correct answer was: 29 out of 100

Tokens you allocated to this bin: 0 tokens

Your final score: 32 points

We will now move on to your payment for this task by drawing a random number between 1 and 100. Because your final score is 32 points:

- if the draw is **less than or equal to 32**, you will earn **$25**
- if the draw is **greater than 32**, you will earn **$5**

Click Stop to select a random number.

81

Stop

Your random number is **81**. This is **greater than 32**, so you will receive a bonus payment of **$5**.

Next

End Screen (*All Treatments*)