Received: 27 October 2021 | Revised: 27 May 2022 | Accepted: 2 August 2022

DOI: 10.1111/jori.12402

ORIGINAL ARTICLE

Journal of Risk and Insurance

Insurance demand experiments: Comparing crowdworking to the lab

Johannes G. Jaspersen[1] | **Marc A. Ragin**[2] | **Justin R. Sydnor**[3]

[1]Munich School of Management, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Department of Insurance, Legal Studies, and Real Estate, Terry College of Business, University of Georgia, Athens, Georgia, USA

[3]Department of Risk and Insurance, Wisconsin School of Business, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Marc A. Ragin, Department of Insurance, Legal Studies, and Real Estate, Terry College of Business, University of Georgia, Moore-Rooker A417, 610 S Lumpkin Street, Athens, GA 30602, USA.
Email: mragin@uga.edu

Funding information
Alfred P. Sloan Foundation, Grant/Award Number: G-2016-7312

Abstract

We analyze an insurance demand experiment conducted in two different settings: in-person at a university laboratory and online using a crowdworking platform. Subject demographics differ across the samples, but average insurance demand is similar. However, choice patterns suggest online subjects are less cognitively engaged—they have more variation in their demand and react less to changes in exogenous factors of the insurance situation. Applying data quality filters does not lead to more comparable demand patterns between the samples. Additionally, while online subjects pass comprehension questions at the same rate as in-person subjects, they show more random behavior in other questions. We find that online subjects are more likely to engage in "coarse thinking," choosing from a reduced set of options. Our results justify caution in using crowdsourced subjects for insurance demand experiments. We outline some best practices which may help improve data quality from experiments conducted via crowdworking platforms.

KEYWORDS

experimental measurement, insurance demand, insurance experiments, online research

JEL CLASSIFICATION

C90, D00, D12, D81, G52

1 | INTRODUCTION

The question of why and how individuals purchase insurance contracts is one of the most widely studied issues in risk management and insurance research. In behavioral theory, insurance demand often serves as an introductory decision problem to exemplify new theoretical frameworks (see, e.g., Kahneman & Tversky, 1979; Kőszegi & Rabin, 2007). Insurance demand is a popular focus in experimental economics due to the relative simplicity of insurance decisions and the importance of insurance markets for economic welfare (for a review, see Jaspersen, 2016). Research evaluates many important and varied topics, including insurance demand in various situations (e.g., Bajtelsmit et al., 2015; Laury et al., 2009; Zimmer et al., 2018), behavioral patterns such as reactions to framing (e.g., Powell & Ansic, 1997), the tendency to make decisions in isolation (Gottlieb & Mitchell, 2020), and the predictive validity of entire theories of decision-making (Jaspersen et al., 2022). However, experimental research on insurance demand does not come without complications. Two common problems are that such studies can be comparatively costly and that the subject pools are typically small and drawn from student populations, which raises questions of representativeness.

Online crowdworking platforms offer a potential solution to both problems. Researchers can post surveys and experimental tasks on such platforms and let the crowdworking population be their, usually much larger, subject pool. While the workers on such platforms are not fully representative of the general population, they offer more variable demographic characteristics than the average student population. Further, since hourly wages on crowdsourcing platforms are lower than those for experimental laboratories, experiments can be conducted for less money. These advantages have led many researchers to use crowdsourcing platforms for their surveys and experimental studies in economics in general (DellaVigna & Pope, 2022; Goodman et al., 2013) and insurance demand in particular (Abito & Salant, 2019; Ragin et al., 2021). This adoption has spurred some criticisms as it is unclear whether the results obtained from studies using online participants can be considered to have the same validity as those obtained in in-person experimental laboratories.

In this paper, we report the results of a large-scale ($n = 1730$) insurance demand experiment carried out both online with crowdworkers and in a physical laboratory with a student population.[1] The incentivized task required participants to select a level of insurance coverage from 0% to 100% of a fixed potential loss amount. Subjects made choices in 12 scenarios that differed in the price of insurance (loading) and probability of suffering the loss. We describe the two different samples and analyze the patterns of insurance choices across the lab and online populations.

In some ways, the lab and online participants make similar choices. Overall, the difference on average insurance demand between the samples is small. Both groups show the same directional responses to changes in prices and probabilities. These similarities are in line with the study by Goodman et al. (2013) that is largely supportive of crowdworking populations as a subject pool for behavioral experiments. Goodman et al. highlight that such online participants tend to produce "reliable results" that are broadly consistent with average behavior observed in

[1]Two companion papers use data from this experiment, Jaspersen et al. (2020) and Jaspersen et al. (2022). Both papers analyze research questions distinct from the one analyzed here, and simply combine the online and in-person subjects in their analyses. Because these papers use data from the same experiment, we use similar language in describing the research setting and experimental design (Section 3 and in the Supporting Information). The theory on preference elicitation (in the Supporting Information) is the same.

other studies on a number of dimensions, including well-documented decision biases. Our results suggest that studying insurance demand with online participants can be a reliable gauge of average demand and some directional comparative statics.

Despite the similarity on average choices, however, the online participants display a series of patterns indicative of reduced cognitive engagement and lower decision quality. The online group has a higher standard deviation within each insurance scenario and reacts less strongly to exogenous changes in the parameters of the insurance decision. Online participants are also significantly more likely to choose the "corner solutions" of no insurance or full insurance than the lab participants. Additionally, the online subjects violate first-order stochastic dominance (FOSD) more frequently than in-person subjects. Online subjects violate the law of demand (LoD) less often, but this is primarily due to choosing the same coverage level in many different scenarios. Differences persist even when controlling for demographic characteristics and risk preferences.

We find that standard attention checks are an insufficient filter to limit the sample of online subjects to only attentive ones. Applying additional filters to the sample, such as removing those with faster-than-average response time or who found the experiment difficult to understand, also does not seem to be able to diminish the difference between the samples consistently. Furthermore, applying multiple filters leads to marked reductions in the sample size.

The specific patterns in our study offer some new insights about the performance of online participants relative to laboratory subjects that prior studies have not fully revealed. A well-cited study by Hauser and Schwarz (2016) finds that experienced online participants appeared to be more attentive than laboratory participants: they pass attention-check questions at higher rates and are more responsive to an experimental variation that relies on small changes to wording in a paragraph. In our study, we similarly find that online participants pass attention checks at nearly the same rates as the laboratory participants and on average have similar insurance demand. Those patterns imply that online participants offer similar response quality as laboratory subjects, consistent with the overall message in studies such as Goodman et al. (2013). Yet, we find that the online participants are significantly less responsive to changes in the underlying parameters of the insurance scenario and are significantly more likely to select the extremes of full or no insurance. Our results show that standard attention-check questions and even measures such as response time are not necessarily sufficient to detect the underlying cognitive engagement with a potentially challenging task like making insurance choices.

We investigate the nature of our online respondents' seemingly reduced attention using a structural modeling approach. One potential reason online participants choose corner solutions more often is that they are engaging in "coarse thinking," limiting their decision to a reduced set of options (e.g., Sovinsky Goeree, 2008). Such a decision process also might explain the more muted response to changes in economic parameters by the online subjects. If coarse thinking was the primary reason for the differences we observe, one way to encourage similarity between online and lab participants in future studies would be to offer only coarse choice sets or filter out subjects who appear to be considering only a coarse subset of choices.

To determine whether coarse thinking plays a role, we develop a discrete choice model that incorporates a *probability of coarse thinking*—the likelihood an agent only considers the two extreme options from the broader choice set. We estimate the model using our set of insurance decisions, allowing for different probabilities of coarse thinking and different underlying elasticities with respect to prices and loss probabilities across the two samples. We find that online participants are significantly more likely to engage in coarse thinking, but controlling for coarse thinking does not meaningfully diminish the estimated differences between populations in how they respond to changes in prices and loss probabilities. This suggests that while coarse thinking is an important part of behavior for online participants, there are differences in the degree of cognitive engagement even among participants who appear to be considering the entire range of options. Importantly, this is not simply a problem of "noise" in online samples which can be solved with a larger sample size—this is a difference in the underlying decision process which must be considered when evaluating existing results and designing new experiments.

Beyond our contributions to the literature on the quality of online versus in-person subject pools, our investigation into coarse thinking also reveals an interesting pattern in how people make insurance decisions. We find that coarse thinking is more common for insurance choices when the probability of loss is low. The design of our experiment does not allow us to disentangle the exact reason for these differences, but the pattern itself suggests that people may be less cognitively engaged in the insurance decision process when considering low-probability events. Our findings can thus inform on the ongoing debate of the seemingly irrational tendency to insure high-probability, low-consequence risks rather than low-probability, high-consequence risks (Browne et al., 2015; Laury et al., 2009; Slovic et al., 1977).

Our results contribute to a broader literature that investigates the quality of online study populations. The prior literature has mixed results, with some studies suggesting that online subjects display less attention and provide lower quality responses than laboratory subjects, while others find minimal differences (see our literature review in Section 2). Our results provide an additional data point into this general discussion. We also contribute to a subset of studies that use incentivized tasks with variation in the parameters of the task to investigate how the populations differ in their response.

Finally, we aim to provide a methodological contribution for researchers conducting online experiments on insurance decisions. Experiments with online subject pools are increasing in popularity, but there are relatively few central resources to help researchers use the platform effectively. Our guidelines are specific to Amazon Mechanical Turk (MTurk), where our study was conducted, but many of our recommendations translate easily to other platforms, such as Prolific Academic and CrowdFlower. We cover topics such as avoiding bots in the subject pool, automating postings and payment, monitoring subjects' communication with each other, and what stake size to use for experimental rewards. For future insurance demand experiments, the distributional parameters of our insurance demand variables can be used to conduct power tests. Further, our results on coarse thinking highlight the benefits of offering multiple different insurance-choice problems. When subjects make multiple insurance decisions, it is possible to identify patterns of coarse thinking and disentangle coarse thinking from an underlying strong preference for one of the extreme options.

2 | RELATED LITERATURE

Here, we discuss the literature most relevant to the major contributions of our study. We focus exclusively on studies which compare behaviors and characteristics of online subjects to those of subjects recruited elsewhere, with special attention to risk attitudes and decisions under risk.[2] Our results speak to cognitive engagement for individual decision tasks; we omit

[2]The review in this section highlights the most relevant comparisons of online versus in-person experiments directly involving risky decisions. We offer a more comprehensive tabulated summary of articles in in Table A1 of Appendix A.

studies of interactive games from this review, as they are likely to involve a different level of engagement. This section is ordered from the narrowest case to the broadest: insurance demand experiments, risky decisions outside insurance, risk attitudes as a demographic characteristic, the quality of decisions in complex tasks, and more general differences between the samples.

Only a few studies in the existing literature conduct insurance demand experiments with both in-person and online samples. Osberghaus and Reif (2021) explore the role of governmental relief schemes and loss experience in the decision to insure against natural disasters. They use a university sample and also recruit online subjects from a previous survey of households in flood-prone areas; their main results are similar across the groups and so they report detailed results for the student sample only. Burkovskaya et al. (2021) and Ragin et al. (2021) each conduct insurance demand experiments with a university sample and an online sample recruited via MTurk. Both studies frame the online experiment as a robustness test of their laboratory results and find support for their main result, though neither study explicitly tests for differences in insurance demand between the two groups. Our study is the first to explicitly compare insurance demand between online and in-person samples, both on average demand and demand elasticity over prices and loss probabilities.

Some studies use both in-person and online samples to explore decision-making under risk outside of an insurance setting. Eben et al. (2020) ask subjects to make a series of choices between safe and risky gambles where outcomes are immediately observable. Both the laboratory and online groups begin the next task more quickly after a loss, and are more likely to choose the risky option again after losing. Alós-Ferrer and Ritschel (2022) investigate whether salience theory can explain preference reversals in risky lotteries. They observe different baseline preference reversal rates for online and laboratory subjects, though the groups considered different lotteries. In both groups, the effect of salience in explaining reversals is null. Johnson and Ryan (2020) examine self-reported risk attitudes over time, finding that online and laboratory samples exhibit similar consistency in their responses. They conduct an incentivized lottery experiment with the online sample and find that self-reported risk attitudes correlate with those implied by their risky choices. The decision under risk in our experiment is the amount of insurance to buy; even when we control for risk attitudes, online and in-person samples differ in their demand elasticity with respect to price and loss probability.

Many more studies measure risk attitudes as a demographic characteristic, but do not focus on its role in the decision-making process. A common method of measuring risk attitudes is to use the self-reported general risk question (GRQ) of Dohmen et al. (2011), as in Buso et al. (2021). Holt and Laury's (2002) incentive-compatible lottery also appears in a number of studies, including Hergueux and Jacquemet (2015) and Li et al. (2021). Many articles in the political science literature, such as Berinsky et al. (2012) and Coppock (2019), index a subject's "risk acceptance" using answers to the seven hypothetical questions outlined in Kam and Simas (2010). Snowberg and Yariv (2021) elicit risk attitudes with three different methods (the procedure in Gneezy & Potters, 1997, a multiple price list of risky lotteries, and the GRQ). For the most part, these studies find that the risk attitudes of online subjects differ at least slightly from in-person subjects. Several note that online samples appear closer to the U.S. population than student samples. While risk attitudes alone are not the purpose of this paper, we contribute to this literature. Univariate tests (reported in the Supporting Information) show that online and in-person samples do not significantly differ in their risk attitudes when utility curvature is the underlying preference motive. However, the samples do significantly differ when other motives—namely, probability weighting and loss aversion—drive their risk attitudes.

A few studies find that online subjects exhibit behaviors similar to in-person subjects, but their behaviors tend to diverge when engaging in more complex tasks. Li et al. (2021) ask laboratory and online subjects to make a series of resource allocation decisions, assigning subjects to either a simple or complex condition. Both subject pools perform worse in the more complex condition, but online subjects have a greater drop-off in performance. The authors suggest that this difference is due to lower attention and cognitive engagement by online subjects. Lee et al. (2018) replicate three experiments in behavioral operations management and find that online subjects perform similarly on average but appear to learn more slowly. The psychology literature also documents larger drop-offs in performance, understanding, and learning for online participants when tasks are complex (Crump et al., 2013; Finley & Penningroth, 2015). Our structural estimation identifies one underlying reason for the difference in performance—online subjects are more likely to simplify a complex decision task.

Ever since experiments using online subjects have been technologically feasible, researchers in many different fields have worked to understand how online samples compare to traditional samples. This literature is too large to summarize in a single paper; here, we offer a sampling of interesting articles and discuss common trends in their findings. Some studies compare descriptive information on subjects, such as demographics (Difallah et al., 2018; Paolacci et al., 2010), political ideologies (Berinsky et al., 2012; Clifford et al., 2015; Coppock, 2019; Huff & Tingley, 2015), and personality traits (Colman et al., 2018; Holden et al., 2013). These studies tend to find that online samples are more reflective of the general population than university laboratory samples, though some differences exist. In other work, researchers use online subjects to replicate behavioral experiments in fields such as advertising (Kees et al., 2017), economics (Amir et al., 2012; Hergueux & Jacquemet, 2015; Horton et al., 2011), and psychology (Crump et al., 2013). In these, online subjects tend to exhibit behaviors (i.e., respond to treatment) in the same direction as laboratory subjects, though the degree of response is not always equivalent. Others focus on the quality of online subject pools with respect to attention (Hauser & Schwarz, 2016), honesty (Peer et al., 2017), costly effort (Farrell et al., 2017), problem behaviors (Necka et al., 2016), and so forth. Our study adds to this literature in the field of behavioral economics, specifically as it applies to insurance decisions. We compare demographics (Table 2), attention (Table 5), and risk attitudes (Table A1). We also show that online and in-person samples have similar levels of average insurance demand, but they differ in response to changes in parameters of the insurance decision.

3 | RESEARCH SETTING AND DESIGN

We conducted a large, incentive-compatible insurance demand experiment both online (via MTurk) and in an experimental laboratory at a large Midwestern public university. In total, our subject pool consisted of 1730 subjects—1352 online and 378 in the traditional laboratory setting. Online subjects were promised $1 in compensation for completing the survey and were made aware that they would earn additional money based on their actions and choices. Subjects in the laboratory were also told that they would earn money based on their actions and choices and faced the same lotteries with the same stakes as online subjects. To comply with compensation standards of the laboratory, subjects were paid an additional $6 after the experiment. Subjects were unaware

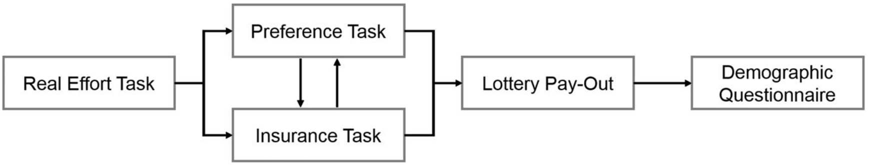

FIGURE 1 Structural outline of the experimental design

of this payment before they received it. We provide an overview of the theoretical considerations and a detailed accounting of the experimental design in the Supporting Information.

3.1 | Experimental design

Our experiment consisted of four stages, which are illustrated in Figure 1. After the introductory screen with a consent form, subjects completed a real-effort task in which they could earn \$5. This task consisted of typing two text passages selected randomly from a set of four possibilities. We showed subjects an image with the written passage, which then had to be typed verbatim into a text box. Subjects took an average of 5.6 minutes for this task. Once the typing task was complete, subjects next faced a preference task and an insurance task in random order.

The preference task consisted of a total of six lottery choice tables, five of which had 16 choices, and one which had 21 choices. The six tables were used to elicit six different motives of underlying risk preferences.[3] These were utility curvature and likelihood insensitivity (each measured in both the gain domain and the loss domain), loss aversion, and a preference for certainty. Additional details on the preference measurement tables and how they translate into preference scales are given in the Supporting Information. Importantly for our later analysis, five of the six tables contain choice options which are dominated in the sense of FOSD. Since this criterion is often used as baseline threshold for rationality (e.g., Bhargava et al., 2017), choices by subjects for the dominated options can be used as a measure of attentive choices.

In the insurance task, subjects were faced with the possibility of losing \$3 of their initial income of \$5 in 12 different insurance scenarios, and they could purchase insurance against the loss. The relative price of insurance (m) and the probability of the loss (p) differed between the scenarios. Losses were determined by the virtual draw of a red ball from an urn filled with 20 balls, some red and some white. The relevant loss probability was represented by the proportion of red balls in the displayed urn. Below the urn, the loss probability was also described as a percentage (e.g., "you have a 10% chance of drawing a red ball and losing \$3.00"). The different p, m pairs are summarized in Table 1. Insurance was available in the form of a partial insurance rate α which could be freely set between 0% and 100% in increments of 1% with the help of a slider. The proportion of the \$3 loss insured, the loading, and the loss probability combine to set the price of insurance at $3\alpha mp$. A summary of the price and payouts was displayed above the slider; it updated each time the subject selected a different coinsurance level. Figure 2 shows an example of this screen.

In the third stage of the experiment, a choice from one of the prior stages (either preference tables or insurance scenarios) was randomly selected for pay-out. The scenario was played out

[3]Tables were displayed to the subjects in random order. We also randomized which choices were shown first and which option was shown on the left side of the screen.

TABLE 1 Insurance scenarios faced by the experimental subjects

	Loading factor *m*				
	0.80	**1.00**	**1.25**	**1.50**	**2.50**
Loss probability p (%)					
5				×	×
10		×	×	×	×
20			×	×	
40				×	
70	×	×	×		

Note: Subjects faced 12 insurance scenarios with a potential loss of $3 and the loss probability and loading factor displayed above.

(virtually) and the subject was informed about the outcome and their final earnings. The fourth stage of the experiment consisted of a demographic questionnaire in which subjects in both subject pools were asked for their age, gender, education, and ethnicity. Online subjects were asked for their annual income in a categorical answer format. Income information was not elicited from the in-person sample, as most university students are unlikely to generate meaningful income. All experimental instructions, including a screenshot for the insurance task, can be found in the Supporting Information.

3.2 | Implementation on MTurk

The MTurk platform can be thought of as a job board, where "requesters" ask "workers" to complete "Human Intelligence Tasks (HITs)." When creating a new MTurk posting, the requester must set several parameters. These parameters include the reward per response, the number of unique workers desired, the time allotted per worker, and the worker qualifications required to participate.

For insurance experiments, researchers should set incentives such that an uninsured loss results in an undesirable hourly wage for the worker. At the same time, the experiment should target average earnings consistent with MTurk norms. In a study using data collected between 2014 and 2017, Hara et al. (2018) find that the mean (median) hourly wage for completed tasks is $6.53 ($3.31) per hour.[4] Most studies have not found that higher payments increase response quality (e.g., Horton & Chilton, 2010; Mason & Watts, 2009). One exception is Litman et al. (2015), who show that larger payments increase the internal consistency of responses for workers in India, but not for US-based workers (who had a higher overall consistency score).

[4]At the time of this writing, Amazon does not allow requesters to explicitly offer variable rewards in the HIT. Instead, requesters must specify a baseline reward per response (i.e., the minimum a participant can earn in the experiment) and then pay any earnings in excess of the minimum as a "bonus" to the worker. Using the term "bonus" may make experimenters uncomfortable. This is a known limitation of the platform and experienced workers appear to understand that variable payments must be made in this way. Nevertheless, experimenters should be careful to clearly explain how the final payment is allocated between the baseline reward and the bonus. In our study, the baseline reward was $1.00.

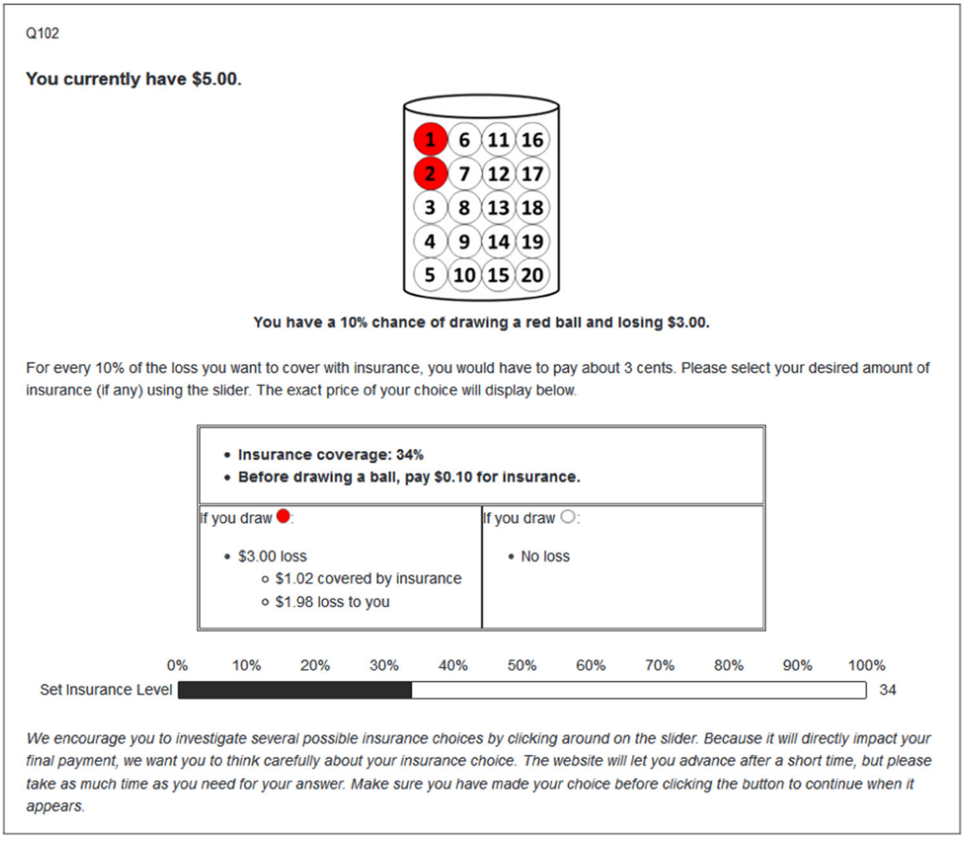

FIGURE 2 Screenshot from the insurance decision task [Color figure can be viewed at wileyonlinelibrary.com]

Amazon's fee is charged as a percent of the total payments to each worker, so this should be accounted for in setting the budget for the experiment.

Setting the required worker qualifications is an important decision for experimenters using MTurk. MTurk allows requesters to limit the visibility of their HIT to workers who meet certain conditions. Two measures which are important in maintaining quality responses are the number of HITs approved (a proxy for the worker's experience with MTurk tasks) and the HIT approval rate (a proxy for worker quality). Experimenters must set these high enough to generate high-quality participation without setting them so high that few workers will qualify.[5] Our experiment required that workers must have completed at least 100 prior HITs with an approval rate of 95% or greater. MTurk has since grown, and there is now a large worker pool with 500 prior HITs and approval rates of 98%. The requirement should be reasonably high, but

[5]Another consideration is that workers with significant experience in economics studies may be "trained" in certain tasks (e.g., Chandler et al., 2014). This is particularly important when using established measurement tools, such as the financial literacy questions of Lusardi and Mitchell (2007). If low-financial-literacy subjects know correct answers simply from experience, their responses will significantly bias results. Chandler et al. (2014) show this effect with the cognitive reflection test of Frederick (2005), a tool common in behavioral finance (e.g., Corgnet et al., 2018).

continuing to increase the necessary qualifications likely has diminishing returns on response quality.

Several considerations are important when designing the HIT posting in MTurk. A common setup is to use the "Survey Link" template in MTurk, which includes a link to an external survey (e.g., to Qualtrics) and an open-response text box for the worker to submit a unique validation code after they have completed the HIT. Survey software can be set up to provide a random code once the worker has completed the experiment, which the worker enters in MTurk. This arrangement links experiment outcomes to the worker's completed HIT for payment. It also is good practice to ask the worker to enter their worker ID at the beginning of the survey. This serves as a backup to the validation code method, and also allows for partial payment if the worker encounters technical errors while participating in the experiment.[6] Finally, qualified workers can view a HIT but are not committed to it until they "accept" it. If workers are able to peruse an experiment before committing to it, it might bias the sample. One common solution is to include Javascript code which hides the hyperlink until the HIT is accepted.

As with any experiment, extensive testing is important. Because MTurk workers may be located around the world on many different types of devices, MTurk experiments should be tested on many different browsers, operating systems, and devices, particularly if treatments rely on the display of information. MTurk offers requesters a "sandbox" for testing the display of their HIT request. In the experiment, a "browser check" question at the beginning can also help to filter out bots. Many survey platforms offer a CAPTCHA question, though it may be useful to tailor this check to the experiment itself.[7]

When an MTurk experiment is live, workers may post some information about it in online forums. Most of these forums have some form of a "daily HITs" thread where workers can post earnings opportunities they find. Often, they post the description of the experiment, their earnings, and the time it took to get their work approved by the requester. There are typically rules that they should not post information that allows others to "game the system." Monitoring these forums can help identify any technical issues, as well as determine whether an experiment is properly incentivized. Prompt responses to worker emails, (e.g., with questions or technical issues) as well as prompt HIT approvals and payment, result in positive feedback in the forums and better engagement from study participants.

Manually approving HITs and making payments is feasible for pilot testing or experiments with a manageable number of participants. With variable payments, experiment outcomes must be matched to worker IDs to determine bonus payments. As n increases, this process becomes laborious and prone to error. MTurk offers an API where requesters can automate HIT approvals and bonus payments using the Amazon Web Services command line interface. Programmers have also developed packages that integrate popular software, such as R and Python, with MTurk's API.

[6]It is common practice, particularly for in-person experiments, to condition payment on successful completion of an experiment. Because experimenters have less control over the conditions in an online experiment (e.g., browser versions, language), there is more room for technical problems or other issues. MTurk workers often participate in online forums where they share their experiences with certain requesters. Partial payments for technical problems can help to maintain a good reputation in these communities.

[7]In our study, we asked participants to move a slider (which looked similar to our insurance questions) to 75% and report the number displayed above the slider. We used Javascript to display a number as a linear function of the slider value, and ended the experiment for any workers who did not provide the correct value. This was preferable to the usual CAPTCHA because it also ensured technical compatability with the real-time display of the parameters of the insurance contract (see Figure 2).

TABLE 2 Summary statistics for demographic variables

	Online		In-person		Difference	
Demographic	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**	**Mean**	**Standard error**
Age	36.14	10.81	21.48	2.34	14.66	0.32***
Female	0.40	0.49	0.72	0.45	−0.32	0.03***
White	0.69	0.46	0.69	0.46	0.01	0.03
Black	0.06	0.24	0.04	0.20	0.02	0.01
Asian	0.21	0.41	0.29	0.45	−0.07	0.03***
Latino	0.06	0.25	0.02	0.15	0.04	0.01***
High school grad	0.99	0.07	1.00	0.00	−0.01	0.00
Some college	0.90	0.30	0.76	0.43	0.14	0.02***
College grad	0.70	0.46	0.19	0.39	0.51	0.02***
Grad school	0.14	0.35	0.04	0.19	0.11	0.01***

Note: For age, the significance of differences is calculated using a t test with unequal variances. For the other demographic variables, the significance of differences is calculated using a two-sample test of proportions. Stars *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. The online sample size is 1352 and the in-person sample size is 378.

4 | RESULTS

4.1 | Differences between crowdworking and the lab

Demographics of the two subject pools differed in a number of ways. We report summary statistics for these demographic characteristics in Table 2. Online subjects are significantly older than subjects in the student pool and are close on average age to the general US population (for a comprehensive discussion of MTurk demographics, see, Difallah et al., 2018). The university sample was primarily female, while the majority of online subjects were male. The proportion of white and black participants was similar between the samples, and the online group had significantly fewer Asian and significantly more Latino subjects.

The (self-reported) educational attainment of the student population was, unsurprisingly, less varied than that of the online population. Online subjects overwhelmingly reported to be college educated, with only 10% of the subjects reporting no college education at all. More than 70% of online subjects reported being college graduates, and another 14% reported having a master's, doctoral, or professional degree (e.g., JD and MD). The difference in age between the populations is a likely factor in the educational differences. Lastly, online subjects reported a fairly wide spread of annual income with the most commonly reported category being between \$25,000 and \$49,999 (28% of online subjects).

We begin our analysis of insurance demand with a descriptive comparison across the two samples. Figure 3 shows the insurance demand distributions by subject sample for the 12 different insurance scenarios. Subjects in the online sample tended to choose the corner options of 0% or 100% more often compared to the in-person sample. For interior choices, demand

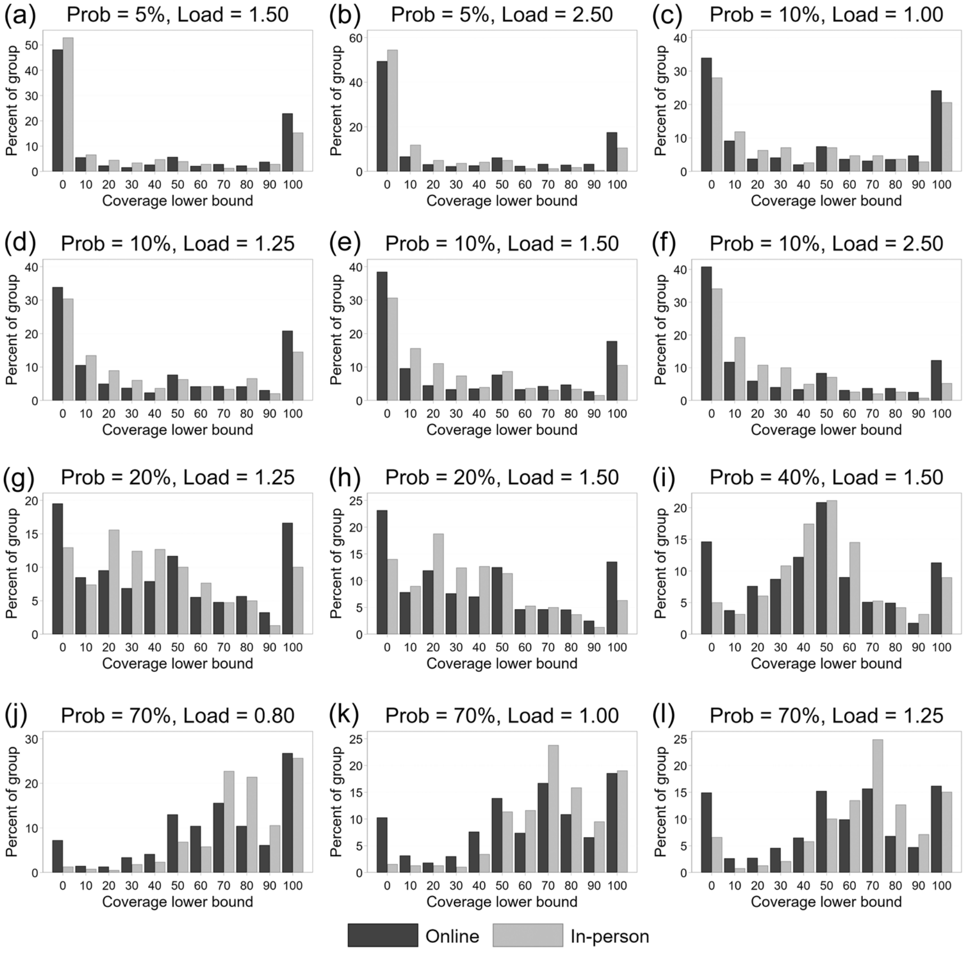

FIGURE 3 Distributions of observed coverage levels chosen by scenario. In each histogram, the *x*-axis is the lower bound of the selected coverage level. The *y*-axis is the percent of subjects in the given sample who selected a coverage level in that bin. Note that the scales of the *y*-axes differ for the different scenarios.

patterns look fairly similar between the two groups. Consistent with the LoD, insurance demand is generally decreasing as prices increase. Contrary to the predictions of most standard models, however, insurance demand is increasing in the loss probability.[8]

To identify whether any differences in the demand patterns aggregated over all insurance scenarios exist, we make further descriptive comparisons in Figure 4. Panel (a) displays the average insurance demand over all considered scenarios. On average, there seems to be very little difference between the two populations. Both have a mean insurance demand slightly above 45% (45.6% for in-person and 46.3% for online) and the standard deviation of the online population is only slightly higher at 38.0% than that of the in-person subjects at 35.4%.

[8]While this behavior is not consistent with theoretical models of insurance demand, it is consistent with empirically observed choices. For example, it has been documented that demand is greater for high-probability, low-impact risks than for low-probability, high-impact risks (e.g., Browne et al., 2015). Jaspersen et al. (2022) identify this as a primary disconnect between what theory predicts and the choices people make in an experiment.

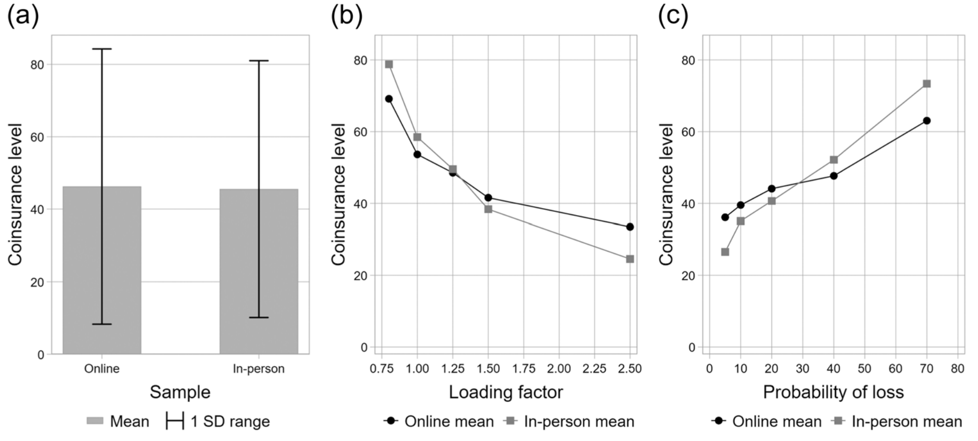

FIGURE 4 Demand for insurance by subject pool: (a) insurance demand across all scenarios, (b) insurance demand by loading factor, and (c) insurance demand by loss probability. Figures report simple arithmetic means of insurance demand for the two subject pools overall (a), at different loading factors (b), and at different loss probabilities (c).

However, this first impression is misleading. Panels (b) and (c) of Figure 4 display the average insurance demand for different loading factors and probabilities of loss, respectively. We can see that both populations react to these factors in the same direction: insurance demand is decreasing in price and increasing in the probability of loss. However, for both factors, we see that the online group is less responsive than the in-person group.

How can the two populations have similar average descriptive statistics but have different reactions to insurance scenario parameters? To answer this question, we conduct a variance decomposition exercise. Denoting $\bar{y}$ the overall average demand and $\bar{y}_j$ the average insurance demand in the jth of the 12 insurance scenarios, we can rewrite the variance of insurance demand as

$$Var(y) = \frac{1}{12}\sum_{j=1}^{12}\frac{1}{n}\sum_{i=1}^{n}(y_{i,j} - \bar{y}_j)^2 + \frac{1}{12}\sum_{j=1}^{12}(\bar{y}_j - \bar{y})^2. \tag{1}$$

The first term is the average variance of subject responses within each of the j insurance scenarios. The second term is the variance of average demand across scenarios. When calculating these values separately for each population, we can see that for the in-person population, 74.6% of the variance is from across-subject variation within the insurance scenarios and 25.4% of the variance can be explained due to different reactions across insurance scenarios. In the online population, the corresponding figures are 91.9% and 8.1%. This implies that, for the in-person population, about one-quarter of the variation in insurance demand is due to reactions to parameter changes across the insurance scenarios. For the online population, less than 10% of the variation in insurance demand can be attributed to such reactions. This is also evident in Table 3, which lists the mean and variance of the insurance demand response separately for each insurance scenario. Despite similar variance over all insurance choices, the online sample has a higher variance than the in-person sample for every individual scenario. The online variance is always higher than the in-person variance, ranging

TABLE 3 Mean and variance of insurance demand by scenario and subject sample

Probability (%)	Loading factor	Online Mean	Online Variance	In-person Mean	In-person Variance
5	1.50	38.11	1819.60	29.65	1463.52
5	2.50	34.08	1632.11	23.39	1122.83
10	1.00	44.57	1717.45	43.28	1518.78
10	1.25	42.00	1627.91	38.19	1363.33
10	1.50	38.64	1568.87	33.12	1133.75
10	2.50	32.87	1347.93	25.71	811.90
20	1.25	46.29	1253.38	42.88	897.76
20	1.50	41.91	1195.98	38.43	780.03
40	1.50	47.67	904.33	52.16	625.03
70	0.80	69.15	843.68	78.77	408.63
70	1.00	62.73	952.51	73.71	452.05
70	1.25	57.31	1044.27	67.60	679.92
Total		46.28	1442.41	45.57	1255.54

Note: Mean and variance are calculated separately for the 1352 online subjects and the 378 in-person subjects.

from 13% higher (for Prob = 10%, Load = 1.00) to 111% higher (for Prob = 70%, Load = 1.00). Thus, while the overall standard deviation in insurance demand seemingly implies similar behavior by both populations, the decomposition analysis shows that behavior differs systematically.

To analyze the difference in insurance demand between the two populations, we use a multivariate regression analysis with insurance demand as the dependent variable. For ease of interpretation, we use a standard ordinary least-squares model with standard errors clustered on the subject level.[9] We report the results of this multivariate analysis with the full sample of participants in Table 4. In column (1), we evaluate how subjects change their insurance demand in response to loss probability and loading, and include an indicator for online subjects to account for differences in the level of coverage between settings. The regression results are consistent with the trends illustrated in Figure 4. Subjects increase their insurance demand with the probability of loss and decrease it with price. Online subjects do not appear significantly different from in-person subjects in their baseline level of insurance demand. In column (2), we interact the online dummy with loss probability and the loading factor. Results show that online subjects are significantly less responsive to the exogenous factors of the insurance scenario. On average, their reaction is moderated by about one-third in comparison to the in-person subjects.

[9]Even though insurance coverage must be nonnegative and cannot be greater than the value lost, our dependent variable might still be considered as left- and right-censored. We thus repeat all our analyses with two-sided Tobit regressions and report results in the Supporting Information. The results are consistent in sign and comparable in significance.

TABLE 4 Regression analysis of insurance demand

Dependent variable: Insurance demand	**(1)**	**(2)**	**(3)**	**(4)**
Probability of loss	0.365***	0.553***	0.538***	0.521**
	(0.016)	(0.029)	(0.201)	(0.209)
Loading factor	−0.082***	−0.106***	−0.118**	−0.112**
	(0.004)	(0.009)	(0.048)	(0.050)
Online	0.706	2.993	1.722	−1.376
	(1.239)	(2.921)	(4.109)	(4.168)
Online × Probability of loss		−0.240***	−0.248***	−0.196***
		(0.035)	(0.051)	(0.052)
Online × Loading factor		0.031***	0.038***	0.039***
		(0.010)	(0.014)	(0.014)
In-person mean	45.573	45.573	45.573	45.573
Demographics and interactions	No	No	Yes	Yes
Preferences and interactions	No	No	No	Yes
R^2	0.106	0.112	0.121	0.135
N subjects	1730	1730	1730	1730
N choices	20,760	20,760	20,760	20,760

Note: The table displays the results of a linear estimation with insurance demand as the dependent variable. When control variables are included, their interactions with loss probability and loading factor are also included. Demographic controls are age, gender, education, and race. Risk preferences are utility curvature and likelihood insensitivity in the gain and the loss domain, loss aversion, and certainty preference. Stars *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors clustered by subject are in parentheses.

These differences in response could potentially be due to the different demographic compositions of the two populations. In column (3) of Table 4, we add controls for demographic variables; we also interact those control variables with both loss probability and loading. Demographic controls include the continuous variable age and indicators for gender, education, and race. These controls do not change the initial results reported in column (2)—online subjects are still less reactive to the exogenous variables. The results in column (3) indicate that the difference in insurance demand behavior between online and in-person subjects does not stem from demographic differences between the two populations.

Even after controlling for demographics, differences in the distributions of risk preferences could still explain different reactions to the exogenous factors of the insurance scenarios. To account for this possibility, we include nonparametric measures of the six risk preferences, as well as their interactions with loss probability and loading. We report the results using this specification, which also includes the demographic variables and their interactions, in column (4) of Table 4. Differences in risk preferences are not able to explain the lesser reaction of online subjects to changes in loss probability and loading. While the coefficient on Online × Probability of loss is reduced in comparison to the other specifications, the reaction of online subjects is still more than one-third lower than that of the in-person subjects.

TABLE 5 Summary statistics for attention variables

Measure	Online		In-person		Difference	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard error
Any wrong attention check	0.27	0.44	0.26	0.44	0.00	0.03
Time taken	10.60	5.84	11.71	4.33	−1.10	0.27***
Study not rated easy	0.12	0.33	0.17	0.38	−0.05	0.02**
Always 0% or 100% insurance	0.10	0.30	0.03	0.18	0.07	0.01***
Violated FOSD	0.30	0.46	0.14	0.34	0.16	0.02***
Num strong LoD violations	1.60	1.40	1.90	1.28	−0.30	0.08***

Note: Row one reports a dummy equal to one if the subject answered any attention-check question wrong. Row two reports the total time taken to complete the insurance and lottery tasks, in minutes. Row three reports a dummy equal to one if the subject did not rate the study as "easy" or "very easy" to understand. Row four reports a dummy equal to one if the subject always chose a corner option (0% or 100% coverage). Row five reports the share of subjects who violated FOSD at least once in their lottery selections. Row six reports the number of times the subject (strongly) violated the law of demand in making insurance choices. For Time taken and Num strong LoD violations, the significance of differences is calculated using a t test with unequal variances. For the other attention variables, the significance of differences is calculated using a two-sample test of proportions. Stars *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. The online sample size is 1352 and the in-person sample size is 378.

Abbreviations: FOSD, first-order stochastic dominance; LoD, law of demand.

4.2 | Randomness in insurance choices

Since neither demographic differences nor differences in risk preferences can explain the observed differences in behavior, another explanation is that they are caused by more random decision behavior of online subjects. Specifically, more random initial choices by online subjects lead to a larger variance across individuals. At the same time, inertia leads many online subjects to choose the same insurance coverage in every scenario, leading to a comparably small between-scenario variance. To evaluate this possibility, we construct several measures of random choice behaviors and summarize them in Table 5. To more robustly evaluate the prevalence of each measure across the populations, we regress each measure on an indicator for online subjects and control for demographic characteristics. We report the results of these regressions in Table 6. To ease interpretation, we estimate linear models for this purpose.[10]

Experimenters often ask "attention-check" questions about the experiment after the instructions are given, which helps flag inattentive subjects and identify potential random choice behavior. Our experiment included five such questions; we logged subjects who answered incorrectly but allowed them to proceed once they provided the correct answer. The "Any wrong attention-check" measure in Tables 5 and 6 is an indicator equal to one if the subject answered any of the five questions incorrectly (and zero otherwise). Consistent with the results in Hauser and Schwarz (2016), online subjects do not perform significantly worse than in-person subjects in answering attention checks correctly. This may be because their continuous work on MTurk HITs has made them particularly aware of such questions—some

[10]In the Supporting Information, we report results of logit regressions on the models with binary dependent variables (all models except column 2). Results are consistent in both sign and significance.

TABLE 6 Regression analysis of random choice behavior

Dependent variable	Any wrong attention check (1)	Time taken (2)	Study not rated easy (3)	Always 0% or 100% insurance (4)	Violated FOSD (5)	Num strong LoD violations (6)
Online	0.018	−2.493***	−0.098***	0.059**	0.193***	−0.138
	(0.034)	(0.527)	(0.028)	(0.021)	(0.033)	(0.108)
Constant	0.423**	7.311***	0.149	0.085	0.271	1.990***
	(0.193)	(1.072)	(0.144)	(0.140)	(0.192)	(0.383)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.017	0.032	0.039	0.017	0.035	0.033
N	1730	1730	1730	1730	1730	1730

Note: Dependent variables are denoted in each column header and are defined as follows. Column (1) is a dummy equal to one if the subject answered any attention-check question wrong. Column (2) is the total time taken to complete the insurance and lottery tasks, in minutes. Column (3) is a dummy equal to one if the subject did not rate the study as "easy" or "very easy" to understand. Column (4) is a dummy equal to one if the subject always chose a corner option (0% or 100% coverage). Column (5) is a dummy equal to one if the subject violated FOSD in any of their lottery selections. Column (6) is the number of times the subject (strongly) violated the law of demand in their insurance choices. Demographic controls are age, gender, education, and race. Stars *, **, and *** denote statistical significance with Šidák–Holm-corrected p-values at the 0.10, 0.05, and 0.01 levels, respectively. Standard errors clustered by subject are in parentheses.

Abbreviations: FOSD, first-order stochastic dominance; LoD, law of demand.

HITs automatically exclude workers without payment if the attention checks are not answered correctly.

While there is no significant difference between the populations according to standard attention-check questions, other patterns in the experiment tell a slightly different story. On average, online subjects completed the experiment about one minute faster than in-person subjects. After controlling for demographics in the multivariate analysis, this difference is 2.5 minutes (column 2).[11] Further, fewer online subjects reported difficulty in understanding the experiment. Column (3) of Table 6 indicates that they were almost 10 percentage points less likely to rate the study as anything other than "easy" or "very easy" to understand in the poststudy questionnaire. It is likely that online subjects understood the experimental instructions sufficiently well, but made more inattentive choices because they wanted to finish the experiment as quickly as possible. Given that many MTurk workers participate in the platform as a means to generate income, they will be focused on finishing each HIT as quickly as possible to increase their HITs per hour and increase their hourly wage. Since we required MTurk workers to have completed at least 100 prior HITs, they may have more experience with similarly structured economics experiments. For the

[11]Differences between the univariate tests in Table 5 and the multivariate tests in Table 6 are due to the multivariate models controlling for demographic characteristics and the sometimes systematic difference in demographics between the two populations. The most sizable difference appears for the "time taken" measure. Here, age had a strongly positive and significant effect on the outcome. Since online subjects are on average almost 15 years older than in-person subjects, the coefficient on the dummy for Online subjects is larger in the multivariate analysis than the univariate difference would imply. The other, less sizable differences between Tables 5 and 6 are due to similar reasons.

predominantly student-based subject pools of in-person laboratories, time and prior experience likely play a smaller role.

Specific choice patterns in the experiment also provide a measure of random choice behavior. For example, online subjects were about 6 percentage points more likely to choose a corner solution ("Always 0% or 100% insurance" in column 4). Online subjects were about twice as likely as in-person subjects to make lottery choices which violated FOSD (column 5). Violating FOSD is a strong indicator of more random choice behavior; adhering to FOSD is often seen as one of the most basic principles of rationality (Bhargava et al., 2017). Only 14% of in-person subjects violated this principle, while about 30% of online subjects made an FOSD-violating choice. The one measure in which online subjects fared better than in-person subjects was the number of LoD violations when purchasing insurance (column 6). The LoD states that insurance demand should be decreasing in price. We counted the LoD as "strongly" violated by a subject if for two questions with equal loss probability, they chose strictly more insurance coverage for the one with a higher loading factor. The insurance scenarios in our experiment provided seven opportunities for subjects to violate the LoD. One possible reason why we saw fewer such violations among the online subjects is that we only counted strict violations. Subjects who always chose the same level of insurance coverage, of whom there were more in the online population, were not counted as violating the LoD.

Given that the less reactive choices of the online subjects seem to be caused by more random choice behavior of subjects, it stands to reason that one could exclude the subset of subjects with such behavior and obtain a sample which reacts similarly to the in-person population. Using the same measures of random choice behavior reported in Tables 5 and 6, we report six such analyses in columns (1)–(6) of Table 7. Specifically, we repeat the regression reported in column (2) of Table 4, excluding subjects who display the random choice behaviors summarized in Table 5. In each column, we describe the subsample below the coefficient estimates.

We report the results of two additional analyses in columns (7) and (8) of the table. Our study was conducted in March 2018, just before concerns emerged about a "quality crisis" in MTurk in the summer of 2018 (Chmielewski & Kucker, 2020). The primary culprit in this crisis is thought to be primarily overseas workers using virtual private systems to pose as US-based workers (Kennedy et al., 2020). Responses from these workers tend to be of significantly lower quality. To address this, we evaluate two additional subsamples, one excluding 39 online subjects from "blocked" IP addresses according to IP Hub (http://iphub.info) and one excluding the 253 online workers with IP addresses outside the United States.

The results reported in the table indicate that filtering on any single measure of random behavior is unlikely to generate similar responses between in-person subjects and online subjects. In particular, online subjects are always less reactive to changes in loss probability than in-person subjects, regardless of the filter being applied. The only filter for which the price elasticity is not significantly different is conditioning on those who took more time in the experiment. Note, however, that this filter results in the smallest sample size (making statistical significance less likely) and that online subjects have a higher baseline level of demand than in-person subjects. Additionally, and unsurprisingly, decreasing the sample of analyzed subjects increases the standard errors of all coefficients and thus leads to more variable results. It thus does not seem like there is a subgroup of more randomly behaving online subjects which can easily be filtered out. Rather, behavior by online subjects seems generally more variable and less reactive than that of in-person subjects.

The regressions reported in Table 7 do not include controls for demographics or risk preferences. Including those controls such that the regressions follow column (4) of Table 4 does

TABLE 7 Regression analysis of insurance demand, subsamples excluding random choice subjects

Dependent variable: Insurance demand	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Probability of loss	0.506***	0.531***	0.561***	0.549***	0.564***	0.500***	0.553***	0.553***
	(0.034)	(0.035)	(0.032)	(0.030)	(0.031)	(0.037)	(0.029)	(0.029)
Loading factor	−0.121***	−0.115***	−0.108***	−0.104***	−0.106***	−0.138***	−0.106***	−0.106***
	(0.011)	(0.012)	(0.010)	(0.009)	(0.010)	(0.011)	(0.009)	(0.009)
Online	0.695	7.765**	2.851	2.750	5.766*	1.049	3.076	0.886
	(3.470)	(3.731)	(3.171)	(2.954)	(3.183)	(3.692)	(2.935)	(3.023)
Online × Probability of loss	−0.212***	−0.293***	−0.238***	−0.215***	−0.250***	−0.224***	−0.238***	−0.207***
	(0.041)	(0.045)	(0.038)	(0.036)	(0.039)	(0.044)	(0.035)	(0.036)
Online × Loading factor	0.033***	0.022	0.028**	0.028***	0.020*	0.036***	0.030***	0.024**
	(0.013)	(0.014)	(0.011)	(0.010)	(0.011)	(0.012)	(0.010)	(0.011)
Demographics and interactions	No	No	No	No	No	No	No	No
Preferences and interactions	No	No	No	No	No	No	No	No
R^2	0.107	0.113	0.116	0.130	0.126	0.103	0.114	0.128
Subsample	Attention checks correct	Time above median	Study rated easy	Some interior choices	No FOSD violations	Two or fewer LoD violations	Clean IP addresses	US subjects only
N in-person subjects	278	250	313	366	326	264	378	378
N online subjects	993	621	1187	1212	950	992	1313	1099
N choices	15,252	10,452	18,000	18,936	15,312	15,072	20,292	17,724

Note: The table displays the results of a linear estimation with insurance demand as the dependent variable. Each column is a subsample excluding subjects who exhibited random choice behavior by different criteria. Column (1) includes only subjects who answered all attention-check questions correctly on the first try. Column (2) includes only subjects who took longer than the median total time for the lottery and insurance choices (9.5 min). Column (3) includes only subjects who rated the study as "easy" or "very easy" to understand. Column (4) includes only subjects who chose an interior insurance option at least once. Column (5) includes only subjects who never violated FOSD in any of their lottery selections. Column (6) includes only subjects who violated the law of demand two or fewer times in their insurance choices. Column (7) includes only subjects with "clean" IP addresses according to the blacklist of iphub.info. Column (8) includes only subjects with a US IP address. Stars *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors clustered by subject are in parentheses.

Abbreviations: FOSD, first-order stochastic dominance; IP, Internet Protocol; LoD, law of demand.

not substantially change our findings (results in the Supporting Information). Further, the regressions in Table 7 apply only one filter at a time. Even when applying *all* the above-referenced filters together, the coefficient on Online × Probability of loss remains negative and significant. Only 302 subjects (101 in-person and 201 online) remain when all filters are applied together.

4.3 | Structural estimation using choice sets

One pattern stands out from our analysis of inattentive behavior in the previous section: crowdworking subjects were more likely to choose corner solutions when making insurance decisions. However, Figure 3 shows that 0% coverage and 100% coverage were popular options for both groups. In line with recent work on consumer decisions and decision-making under risk (Barseghyan et al., 2021; Sovinsky Goeree, 2008), one possible reason for these patterns is that some subjects use a different choice set than others. Specifically, subjects sometimes appear to focus more on the corner solutions as feasible choices. In some scenarios, particularly those with low loss probabilities, these choice patterns occur more often.

The analysis framework used thus far is incapable of differentiating between choices made from specific sets. Even when considering all available options (i.e., 0%–100% in 1% increments), a subject still might choose a corner solution as the best available option. Indeed, many established models of choice would predict these corner solutions for some of the commonly assumed combinations of preference parameters (Jaspersen et al., 2022). Thus the filter used in column (4) of Table 7 does not adequately categorize subjects based on their likely choice sets. To more fully explore the possible role of different choice sets we conduct a structural estimation, which considers choice sets explicitly. In addition to comparing the two subject populations, this approach also has the advantage that it allows us to learn about which other factors lead people to make insurance choices from the coarse choice set, thus providing a contribution to the general literature on insurance demand.

For the structural estimation, we follow the typical approach and consider a stochastic choice environment. We thus differentiate the utility obtained by subject i from choosing coinsurance level α in scenario j into a deterministic component and an unobserved stochastic component such that

$$u_{ij}(\alpha) = E[U_i(\alpha, j)] + \epsilon_{ij\alpha}. \tag{2}$$

$E[U_i(\alpha)]$ is the expected utility of subject i derived from choosing coinsurance level α. $\epsilon_{ij\alpha}$ is a mean-zero stochastic term independently and identically distributed (i.i.d.) type I extreme value. Within a choice set A, each coinsurance level h thus has the probability

$$\frac{exp(E[U_i(\alpha = h, j)])}{\sum_{k \in \mathrm{A}} exp(E[U_i(\alpha = k, j)])} \tag{3}$$

of being chosen.

The expected utility of coinsurance level α in scenario j is given by

$$E[U_i(\alpha, j)] = p_j U_i(w - (1 - \alpha)L - \alpha m_j p_j L) + (1 - p_j)U_i(w - \alpha m_j p_j L). \tag{4}$$

Using a second-order Taylor approximation and suppressing the indices on p and m for legibility, we expand this expression around w to gain

$$E[U(\alpha,j)] \approx \underbrace{\frac{U_i(w) - U_i'(w)p_jL - \frac{1}{2}U''(w)p_jL^2}{U_i'(w)p_jL}}_{\beta_0} + \underbrace{\left[1 - m_j + 2(m_jp_j - L)\frac{U_i''(w)}{U_i'(w)}L\right]}_{\beta_1}\alpha$$
$$+\underbrace{\left[(1-m_j)(p_j - m_jp_j)\frac{U_i''(w)}{U_i'(w)}L\right]}_{\beta_2}\alpha^2. \tag{5}$$

Given the choice probability in Equation (3), β_0 has no effect and thus does not play a role in the estimation of the model.[12] β_1 and β_2 depend on the probability and the loading in the decision situation, as well as on the risk aversion coefficient and thus the preferences of the decision-maker. We assume that preferences are homogeneous within the lab population and across the crowdworkers, but that they can differ between the two samples. We thus approximate the two coefficients (ignoring interactions between p and m in the interest of parsimony) by setting $\beta_l(i,j) = \beta_l^0 + \beta_l^p p_j + \beta_l^p m_j + \beta_l^{MT} MT_i + \beta_l^{MTp} MT_i p_j + \beta_l^{MTm} MT_i m_j$ for $l \in \{1, 2\}$, where MT indicates an MTurk subject.[13] The last two terms are included to reflect the fact that the crowdworking population seems to react differently to the loading and loss probability of each insurance scenario. We thus estimate the observed component of the utility obtained by subject i of choosing α in scenario j as

$$\begin{aligned} E[U_i(\alpha,j)] = {} & \beta_1^0\alpha + \beta_1^p p_j\alpha + \beta_1^m m_j\alpha + \beta_1^{MT} MT_i\alpha + \beta_1^{MTp} MT_i p_j\alpha + \beta_1^{MTm} MT_i m_j\alpha \\ & + \beta_2^0\alpha^2 + \beta_2^p p_j\alpha^2 + \beta_2^m m_j\alpha^2 + \beta_2^{MT} MT_i\alpha^2 + \beta_2^{MTp} MT_i p_j\alpha^2 \\ & + \beta_2^{MTm} MT_i m_j\alpha^2. \end{aligned} \tag{6}$$

To reflect that some decision-makers seem to choose from a coarse choice set, while others choose from a choice set that includes all possible coinsurance levels, we adopt the stochastic choice set model of Sovinsky Goeree (2008). Decision-makers either choose from the coarse set $C = \{0, 100\}$ or the full set $F = \{0, 1, \ldots, 100\}$. The probability that subject i chooses from the coarse choice set in scenario j is equal to $\phi^C(\omega_{ij})$ and depends on the attention ω_{ij} that the subject spends on the given scenario. Combining these elements, we can calculate the probability that subject i chooses coinsurance level $\alpha = h$ in scenario j as

$$\begin{aligned} s_{ij}(\alpha = h) = {} & \mathbf{1}_{h \in C}\phi^C(\omega_{ij})\frac{exp(E[U_i(\alpha = h, j)])}{\sum_{k \in C} exp(E[U_i(\alpha = k, j)])} \\ & + (1 - \phi^C(\omega_{ij}))\frac{exp(E[U_i(\alpha = h, j)])}{\sum_{k \in \mathrm{F}} exp(E[U_i(\alpha = k, j)])}. \end{aligned} \tag{7}$$

[12]Within an insurance scenario, β_0 is the same for all coverage levels. The choice probability in Equation (3) for coverage level h thus reads $\frac{exp(\beta_0 + \beta_1 h + \beta_2 h^2)}{\sum_{k \in \mathrm{A}} exp(\beta_0 + \beta_1 k + \beta_2 k^2)])} = \frac{exp(\beta_0) \times exp(\beta_1 h + \beta_2 h^2)}{\sum_{k \in \mathrm{A}} exp(\beta_0) \times exp(\beta_1 k + \beta_2 k^2)])} = \frac{exp(\beta_1 h + \beta_2 h^2)}{\sum_{k \in \mathrm{A}} exp(\beta_1 k + \beta_2 k^2)])}$ and is not affected by β_0.

[13]Note that our experiment did not vary L between different insurance scenarios. Without any variation in the variable, its influence is not identified separately, but rather subsumed in the β coefficients.

Here, $\mathbf{1}_{h\in C}$ is an indicator function which shows whether or not coinsurance level h is in the coarse choice set. Attention is a function of the scenario attributes as well as of the subject population such that

$$\omega_{ij} = \gamma_0 + \gamma^p p_j + \gamma^m m_j + \gamma^{MT} MT_i. \tag{8}$$

It translates to the choice set probability via the equation $\phi^C(\omega_{ij}) = \frac{1}{1+exp(\omega_{ij})}$ as is suggested by Sovinsky Goeree (2008).

We estimate the decision model by maximizing the log-likelihood of the 20,760 insurance choices observed in our data. Specifically, if we denote the coinsurance level chosen by subject i in scenario j as α_{ij}, then we find those β and γ coefficients which maximize the expression $\sum_{i=1}^{N}\sum_{j=1}^{12} log(s_{ij}(\alpha = \alpha_{ij}))$. Table 8 shows the resulting coefficient estimates and block-bootstrapped standard errors sampled on the subject level. The β coefficient estimates are somewhat difficult to interpret outside of the model structure. However, we can immediately see the strong influence of the loss probability on insurance demand from the large and significant estimates of β_1^p and β_2^p. The effect of the loading, on the other hand, is still present but shows smaller levels of statistical significance. The influence of both p and m on the utility of insurance demand is smaller in the online population, because β_1^{MTp}, β_1^{MTm}, β_2^{MTp}, and β_2^{MTm} are of the opposite sign as their respective counterparts among $\beta_1^p, \beta_1^m, \beta_2^p$, and β_2^m. γ coefficients are easier to interpret because a higher value of ω_{ij} implies a higher likelihood of choosing from the full choice set. The negative sign on γ_{MT} indicates that online subjects are much less likely to consider the full choice set. The positive sign on γ_p indicates that, for all subjects, the likelihood of considering the full choice set increases with the loss probability.

For a simple interpretation of the model's results, we use the coefficient estimates in Table 8 to simulate choice behavior according to the structural model. A selection of the results is illustrated in Figure 5. In the first two panels of the figure, we focus on the demand pattern if subjects consider the full choice set. The main results of the descriptive observations in Figure 4

TABLE 8 Coefficient estimates in the structural decision model

Coefficient	Estimate	Standard error	Coefficient	Estimate	Standard error
β_1^0	−4.336***	(1.144)	β_2^m	−0.956*	(0.523)
β_1^p	0.254***	(0.020)	β_2^{MT}	−0.235	(1.291)
β_1^m	−0.152	(0.501)	β_2^{MTp}	0.016	(0.019)
β_1^{MT}	0.366	(1.312)	β_2^{MTm}	0.501	(0.582)
β_1^{MTp}	−0.060**	(0.022)	γ_0	0.615***	(0.091)
β_1^{MTm}	0.111	(0.561)	γ_p	0.014***	(0.001)
β_2^0	4.069***	(1.126)	γ_m	−0.011	(0.023)
β_2^p	−0.184***	(0.017)	γ_{MT}	−0.545***	(0.090)

Note: The table displays the coefficient estimates of the structural decision model given in Equations (6)–(8). The subscript on each β denotes the exponent on α in the term, while the superscript indicates the other variables in the term (p, m, and MT). Coefficients are estimated using a maximum likelihood procedure and standard errors are block-bootstrapped with 1000 iterations. Stars *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors clustered by subject are in parentheses.

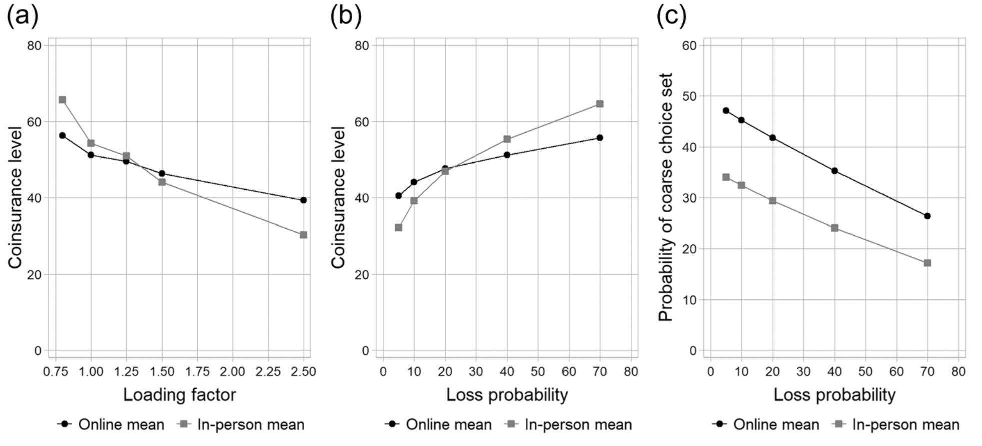

FIGURE 5 Simulated choice behavior according to the estimation results of the structural model. (a) average simulated insurance demand of subjects choosing in the full choice set by loading factor, (b) average simulated insurance demand of subjects choosing in the full choice set by loss probability, and (c) simulated probability of choosing the coarse choice set by loss probability. Graphs use the model outlined in Equations (6)–(8) and the coefficient estimates from Table 8.

are maintained—demand decreases in loading, increases in probability, and the online population is less sensitive in both. If anything, the difference between the two subject populations becomes even more pronounced. Online subjects seem to respond even less to the changes in the insurance scenario than if the choice sets are not taken into consideration.

Panel (c) of Figure 5 displays the probability of choosing from the coarse choice set as a function of loss probability and subject population. It reflects the results on the γ coefficients graphically. Online subjects are much more likely to use the simpler choice set, thus suggesting that they spend less attention on the insurance decisions, on average. There is also a strong and negative effect of the loss probability on the likelihood of choosing from the coarse choice set. This suggests that as the expected loss increases, subjects more carefully consider their options and thus make a more informed decision. Higher expected loss thus increases the care with which an insurance decision is made. On the other hand, we find no such influence of the loading factor (not reported). So our findings do not support a well-calibrated model of rational inattention as, for example, proposed by Caplin and Dean (2015). These findings have important implications for insurance demand research outside an experimental setting.

5 | DISCUSSION AND CONCLUSIONS

In this study, we compare behavior in an incentive-compatible insurance demand experiment between subjects in a university laboratory and subjects recruited through an online crowdworking platform. We show that while aggregate demand looks similar between both populations, online subjects provide more variable answers within each insurance scenario and react less strongly to changes in the parameters of the insurance scenarios. Further, online subjects make more choices which are commonly associated with irrational behavior, such as choices that violate stochastic dominance. Differences in demographic characteristics or risk preferences cannot explain the differences in observed behavior. Simple attention checks in the form of instruction comprehension questions are insufficient to filter out more randomly behaving online subjects.

Our results are seemingly at odds with studies such as that by Hauser and Schwarz (2016, and references within), who show that crowdworking populations are more attentive to survey instructions than participants from student populations. However, a deeper look shows that the results are actually consistent. As Hauser and Schwarz demonstrate, these online subjects have learned to be attentive to instructions because they are often punished (e.g., excluded from a study without payment) if they are not. We also observe this effect in our experiment. Online subjects are as good as in-person subjects in answering questions about the experimental instructions. However, when it comes to the choices within the experiment, choices for insurance coverage or between risky gambles are likely not common enough for online subjects to have consistently learned about them. Even if they were common, the feedback provided to the subjects is not as direct as it is for the instruction comprehension questions or true/false questions tested by Hauser and Schwarz (2016). Subjects are merely informed about the outcome of the random variable in question, observe the choice they made in the respective experimental question and are informed about the resulting payment. No statement about the correctness of their choice is made. So to observe that a mistake was made, for example, in the form of an FOSD violation, subjects would have to pay close attention at the pay-out screen. This indirect form of feedback likely hampers learning by subjects and thus explains the difference between our results on attention checks and our results on the other forms of random choice behavior.

Our results provide some reasons for caution when recruiting crowdworkers as subjects for insurance demand experiments. On the one hand, since their insurance demand reactions to exogenous factors are consistent in direction with those observed by in-person subjects, we can expect that effects shown in online populations should also appear in in-subject populations. On the other hand, the more random choices suggest that a larger sample would be necessary to detect any effects. At the same time, however, behavior typically reported as irrational is more often observed in the online group. Our evidence suggests that this increase is due to more random choices by subjects rather than fundamentally different preferences. This is also supported by the analysis of Chandler et al. (2014). They report that crowdworkers are likely to do other activities (such as watching TV or using their phones) while participating in an experiment. If that is indeed the case, then studies using these populations which report irrational insurance demand behavior might overstate the prevalence of such effects in comparison to studies using in-person laboratories. The fact that excluding subjects which show some kind of random behavior does not consistently make reactions of online subjects more similar to in-person subjects suggests that the behavioral differences are not due to a small subset of subjects, but are endemic to the entire population. For studies of insurance demand choice anomalies, results using crowdworkers as subjects should be validated with results from an in-person experiment.

Future research in this area might evaluate differences between the samples with respect to other behaviors in the insurance decision process. For example, insurance consumers often have private information about their risk, which can manifest into an adverse selection problem for insurance pools. This has been experimentally investigated with subjects in the laboratory (Riahi et al., 2013) and online (Zhang & Palma, 2021). No study has tested whether these samples differ in their use of private information, and so their results cannot be directly compared. A similar hole in the literature exists for other areas of research in insurance economics, such as moral hazard and providing biased advice. Evaluating such behaviors in crowdworking samples relative to more traditional samples is a promising avenue for future research and, similar to our study, will give researchers in the field essential tools for assessing existing results and designing new studies.

ACKNOWLEDGEMENTS

We thank the Alfred P. Sloan Foundation for financial support under grant number G-2016-7312. This study was pre-registered with the American Economic Association's randomized controlled trial registry under ID AEARCTR-0002783 (https://doi.org/10.1257/rct.2783-1.0). For helpful feedback and comments, we thank Lan Zou and seminar participants at the American Risk and Insurance Association annual meeting. We are grateful to the JRI editorial team and two anonymous reviewers for their comments, which substantially improved the manuscript. All remaining errors are ours.

ORCID

Johannes G. Jaspersen http://orcid.org/0000-0002-3599-8988
Marc A. Ragin http://orcid.org/0000-0003-2370-410X
Justin R. Sydnor http://orcid.org/0000-0003-0532-1949

REFERENCES

Abito, J. M., & Salant, Y. (2019). The effect of product misperception on economic outcomes: Evidence from the extended warranty market. *Review of Economic Studies*, *86*(6), 2285–2318.

Alós-Ferrer, C., & Ritschel, A. (2022). Attention and salience in preference reversals. *Experimental Economics*, *25*, 1024–1051.

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the Internet: The effect of $1 stakes. *PLoS ONE*, *7*(2), e31461.

Bajtelsmit, V., Coats, J. C., & Thistle, P. (2015). The effect of ambiguity on risk management choices: An experimental study. *Journal of Risk and Uncertainty*, *50*(3), 249–280.

Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, *89*(5), 2015–2048.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

Bhargava, S., Loewenstein, G., & Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated options. *The Quarterly Journal of Economics*, *132*(3), 1319–1372.

Bieniek-Tobasco, A., Rimal, R. N., McCormick, S., & Harrington, C. B. (2020). The power of being transported: Efficacy beliefs, risk perceptions, and political affiliation in the context of climate change. *Science Communication*, *42*(6), 776–802.

Bolton, G., Dimant, E., & Schmidt, U. (2020). When a nudge backfires: Combining (Im)plausible deniability with social and economic incentives to promote behavioral change (CESifo Working Paper No. 8070), Center for Economic Studies and ifo Institute (CESifo), Munich. pp. 1–40.

Browne, M., Knoller, C., & Richter, A. (2015). Behavioral bias, market intermediaries, and the demand for bicycle and flood insurance. *Journal of Risk and Uncertainty*, *50*, 141–160.

Burkovskaya, A., Teperski, A., & Atalay, K. (2021). Framing and insurance choices. *Journal of Risk and Insurance*, *89*(2), 311–337.

Buso, I. M., Di Cagno, D., Ferrari, L., Larocca, V., Lorè, L., Marazzi, F., Panaccione, L., & Spadoni, L. (2021). Lab-like findings from online experiments. *Journal of the Economic Science Association*, *7*(2), 184–193.

Caplin, A., & Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, *105*(7), 2183–2203.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473.

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), 1–9.

Colman, D. E., Vineyard, J., & Letzring, T. D. (2018). Exploring beyond simple demographic variables: Differences between traditional laboratory samples and crowdsourced online samples on the Big Five personality traits. *Personality and Individual Differences*, *133*, 41–46.

Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, *7*(3), 613–628.

Corgnet, B., Desantis, M., & Porter, D. (2018). What makes a good trader? On the role of intuition and reflection on trader performance. *The Journal of Finance*, *73*(3), 1113–1137.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*(3), e57410.

DellaVigna, S., & Pope, D. (2022). Stability of experimental results: Forecasts and evidence. *American Economic Journal: Microeconomics*, *14(3)*, 889–925.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February 5–9). Demographics and dynamics of Mechanical Turk workers. In *Proceedings of WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining* (pp. 135–143). ACM, CA, USA.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Eben, C., Chen, Z., Vermeylen, L., Billieux, J., & Verbruggen, F. (2020). A direct and conceptual replication of post-loss speeding when gambling. *Royal Society Open Science*, *7*(5), 200090.

Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment and Decision Making*, *9*(2), 14.

Farrell, A. M., Grenier, J. H., & Leiby, J. (2017). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review*, *92*(1), 93–114.

Finley, A., & Penningroth, S. (2015). Online versus in-lab: Pros and cons of an online prospective memory experiment. *Advances in Psychology Research*, *113*, 135–162.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*(2), 631–645.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224.

Gottlieb, D., & Mitchell, O. S. (2020). Narrow framing and long-term care insurance. *Journal of Risk and Insurance*, *87*(4), 861–893.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Hergueux, J., & Jacquemet, N. (2015). Social preferences in the online laboratory: A randomized experiment. *Experimental Economics*, *18*(2), 251–283.

Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon's Mechanical Turk. *Computers in Human Behavior*, *29*(4), 1749–1754.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Horton, J. J., & Chilton, L. B. (2010). The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce* (pp. 209–218). ACM. 00271.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3), 1–12.

Jaspersen, J. G., Ragin, M. A., & Sydnor, J. R. (2020). Linking subjective and incentivized risk attitudes: The importance of losses. *Journal of Risk and Uncertainty*, *60*, 187–206.

Jaspersen, J. G. (2016). Hypothetical surveys and experimental studies of insurance demand: A review. *Journal of Risk and Insurance*, *83*(1), 217–255.

Jaspersen, J. G., Ragin, M. A., & Sydnor, J. R. (2022). Predicting insurance demand from risk attitudes. *Journal of Risk and Insurance*, *89*(1), 63–96.

Johnson, D., & Ryan, J. B. (2020). Amazon Mechanical Turk workers can provide consistent and economically meaningful data. *Southern Economic Journal*, *87*(1), 369–385.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kam, C. D., & Simas, E. N. (2010). Risk orientations and policy frames. *The Journal of Politics*, *72*(2), 381–396.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, *8*(4), 614–629.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, *97*(4), 1047–1073.

Laury, S. K., McInnes, M. M., & Swarthout, J. T. (2009). Insurance decisions for low probability losses. *Journal of Risk and Uncertainty*, *39*(1), 17–44.

Lee, Y. S., Seo, Y. W., & Siemsen, E. (2018). Running behavioral operations experiments using Amazon's Mechanical Turk. *Production and Operations Management*, *27*(5), 973–989.

Li, Z., Lin, P.-H., Kong, S.-Y., Wang, D., & Duffy, J. (2021). Conducting large, repeated, multi-game economic experiments using mobile platforms. *PLoS ONE*, *16*(4), e0250668.

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. *Behavior Research Methods*, *47*(2), 519–528.

Lusardi, A., & Mitchell, O. S. (2007). *Financial literacy and retirement planning: New evidence from the Rand American Life Panel* [Michigan Retirement Research Center, Research Paper No. WP 2007–157]. University of Michigan.

Mason, W., & Watts, D. J. (2009). Financial incentives and the "performance of crowds." In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 77–85). ADM, NY, USA. https://doi.org/10.1145/1600150.1600175

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PLoS ONE*, *11*(6), e0157732.

Osberghaus, D., & Reif, C. (2021). How do different compensation schemes and loss experience affect insurance decisions? Experimental evidence from two independent and heterogeneous samples. *Ecological Economics*, *187*, 107087.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Powell, M., & Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, *18*(6), 605–628.

Ragin, M. A., Collier, B., & Jaspersen, J. G. (2021). The effect of information disclosure on demand for high-load insurance. *Journal of Risk and Insurance*, *88*(1), 161–193.

Riahi, D., Lévy-Garboua, L., & Montmarquette, C. (2013). Competitive insurance markets and adverse selection in the lab. *The Geneva Risk and Insurance Review*, *38*(1), 87–113.

Slovic, P., Fischhoff, B., Lichtenstein, S., Corrigan, B., & Combs, B. (1977). Preference for insuring against probable small losses: Insurance implications. *Journal of Risk and Insurance*, *44*(2), 237–258.

Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, *111*(2), 687–719.

Sovinsky Goeree, M. (2008). Limited information and advertising in the U.S. personal computer industry. *Econometrica*, *76*(5), 1017–1074.

Zhang, P., & Palma, M. A. (2021). Compulsory versus voluntary insurance: An online experiment. *American Journal of Agricultural Economics*, *103*(1), 106–125.

Zimmer, A., Gründl, H., Schade, C. D., & Glenzer, F. (2018). An incentive-compatible experiment on probabilistic insurance and implications for an insurer's solvency level. *Journal of Risk and Insurance*, *85*(1), 245–273.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jaspersen, J. G., Ragin, M. A., & Sydnor, J. R. (2022). Insurance demand experiments: Comparing crowdworking to the lab. *Journal of Risk and Insurance*, 89, 1077–1107. https://doi.org/10.1111/jori.12402

APPENDIX A: TABULATED LITERATURE REVIEW

TABLE A1 Articles involving risk that compare online to in-person samples

Author	Main comparison topic	Treat versus control?	Incentive-compatible	Comparison groups	Main result(s)
Alós-Ferrer and Ritschel (2022)	Preference reversals due to salience	Yes, alternative lottery to create salience	Yes	Prolific, laboratory	Salience theory does not explain preference reversals in risky lotteries. Baseline preference reversal rates were higher for online than in-person.
Bieniek-Tobasco et al. (2020)	Climate risk perception	Yes, content of video	No	MTurk, laboratory	Online has higher climate risk perception and beliefs about the efficacy of combating climate change. Demographics also differ. Watching climate change documentary and political affiliation are associated with efficacy beliefs and risk perception. No direct comparison of groups.
Bolton et al. (2020)	Inducing prosocial behavior, risk-taking behavior	Yes, observability of behavior, inequality, normative beliefs	Yes	MTurk, laboratory	No significant differences in risk-taking behavior. Online similar to initial lab results. Reputation increases prosocial behavior, social observation does not. Inequity plays a role.
Burkovskaya et al. (2021)	Insurance deductible choice	Yes, broad versus narrow framing of insurance choices	Yes	MTurk, laboratory	WTP was similar between treatments. Broad framing preferred more similar deductibles. No direct comparison of groups.

(Continues)

TABLE A1 (Continued)

Author	Main comparison topic	Treat versus control?	Incentive-compatible	Comparison groups	Main result(s)
Eben et al. (2020)	Speed of taking on another task after a risky loss	No	Yes	Prolific, laboratory	Lab and online have similar patterns in speed to next task: fastest after risky loss and slowest after choosing safe option. Both groups more likely to gamble again after risky loss than after risky win or choosing safe option.
Evans and Krueger (2014)	Role of risk attitudes and temptation in the trust game	Yes, riskiness of trustor's payoff and temptation of trustee	Yes	MTurk, laboratory	Online more likely to trust and reciprocate. Role of risk and temptation similar between groups.
Goodman et al. (2013)	Attention, demographics, risk attitudes	No	No	MTurk, community survey, laboratory	Online similar to community in many demographics and risk attitudes. Online pays less attention, uses internet more, has lower self-esteem, and is more introverted. Differences were often larger when comparing online to students.
Johnson and Ryan (2020)	Impulsiveness, self-reported risk attitudes (GRQ), and risky decisions, consistency of responses over time	No	No for risk attitude, yes for risky decisions	MTurk, existing lit	Temporal consistency of self-reported risk attitude is similar between groups. Self-reported risk preference correlates with incentivized risky choices. Authors cite this as evidence that online workers are not using a decision rule to complete study quickly.

TABLE A1 (Continued)

Author	Main comparison topic	Treat versus control?	Incentive-compatible	Comparison groups	Main result(s)
Li et al. (2021)	Compare interactive experiment responses between lab and live-streaming settings, also elicit risk preferences	No	Yes	Online university students, in-person university students	Online more risk-averse than in-person on average. Online had a slightly larger proportion who were risk-seeking.
Osberghaus and Reif (2021)	Effect of governmental relief schemes and loss experience on decision to insure	Yes, whether govt disaster relief risky or guaranteed	Yes	Household heads from previous survey, laboratory	Neither group had significant response to treatment. Loss experience has negative effect on insurance demand. No direct comparison of groups.
Ragin et al. (2021)	Effect of disclosure on demand for high-load insurance	Yes, disclosed either loss probability, expected loss, or insurer profit	Yes	MTurk, laboratory	None of the disclosures significantly decreased demand relative to control group. No direct comparison of groups.
Snowberg and Yariv (2021)	Behavioral attributes relevant to economics and psychology (e.g., risk aversion and altruism)	No	Yes	MTurk, undergraduate survey, US survey, laboratory	Most behaviors correlate positively across samples. Average levels tend to be different, with online between students and US representative sample. Nonstudent samples tend to be noisier.