Journal of the Economic Science Association (2023) 9:293–307
https://doi.org/10.1007/s40881-023-00135-5

METHODOLOGY PAPER

Is a non-representative convenience sample of adults good enough? Insights from an economic experiment

Sean F. Ellis[1] · Olesya M. Savchenko[2] · Kent D. Messer[3]

Received: 4 January 2022 / Revised: 1 April 2023 / Accepted: 28 May 2023 /
Published online: 31 August 2023
© The Author(s), under exclusive licence to Economic Science Association 2023

Abstract

Recruitment of representative and generalizable adult samples is a major challenge for researchers conducting economic field experiments. Limited access to representative samples or the high cost of obtaining them often leads to the recruitment of non-representative convenience samples. This research compares the findings from two field experiments involving 860 adults: one from a non-representative in-person convenience sample and one from a representative online counterpart. We find no meaningful differences in the key behaviors of interest between the two samples. These findings contribute to a growing body of literature demonstrating that non-representative convenience samples can be sufficient in certain contexts.

Keywords Non-representative convenience sampling · Field experiments · Online recruitment · Representative sampling

JEL Classification B41 · C83 · C93

✉ Sean F. Ellis
sfellis@wharton.upenn.edu

Kent D. Messer
messer@udel.edu

[1] Behavior Change for Good, The Wharton School, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104, USA

[2] Food and Resource Economics Department, University of Florida, 1183 McCarty Hall A, Gainesville, FL 32611, USA

[3] Department of Applied Economics and Statistics, University of Delaware, 531 S. College Avenue, Newark, DE 19716, USA

1 Introduction

Economic experiments can help inform the design and evaluation of policies in various contexts across the social sciences (Banerjee & Duflo, 2009; Boas et al., 2020; Bol, 2019; List & Price, 2016; Rosch et al., 2021). For causal inferences to be informative, the sample of respondents must be externally valid. In other words, participants in the experiment need to respond in the same way as the target population for the results to be generalizable (Banerjee & Duflo, 2017; Muller, 2014). This is particularly important when testing program interventions that are intended to be scaled up to the broader population. However, the recruitment of sufficiently large samples that are representative of the population of interest is one of the major challenges of conducting economic experiments (Palm-Forster & Messer, 2021; Palm-Forster et al., 2019; Roe & Just, 2009; Weigel et al., 2021). Limited access to representative samples or the high cost of obtaining them often leads to the recruitment of non-representative samples drawn from the student population at a university, the public in a field location, or users of online platforms such as MTurk, Facebook or Qualtrics. Increasing reliance on such samples has generated a significant discussion in the literature, with some researchers highlighting concerns over the generalizability of findings from non-representative samples (Goldberg et al., 2019; Levitt & List, 2007, 2008), and others demonstrating the value of such samples (Camerer, 2011; Frigau et al., 2019; Peth & Mußhoff, 2019; Rosch et al., 2021).

We contribute to this discussion by comparing experimental findings from a non-representative in-person field experiment to its representative online counterpart. Specifically, we conducted two framed field experiments with 860 adult participants to test consumer response to stigma-mitigating strategies for food and drink products produced with recycled water. The in-person field experiment recruited 314 adult participants from the US mid-Atlantic region and was not representative of it. The online study relied on the same experimental design but used a Qualtrics panel to recruit a sample of 546 online participants representative of the US mid-Atlantic (defined in this article as Delaware, Maryland, New Jersey, and Pennsylvania).

Our analysis of the data shows no meaningful differences in the key variables of interest between our representative online mid-Atlantic sample and our non-representative in-person mid-Atlantic sample. This implies that in the context of eliciting consumer preferences for food products, using a non-representative convenience sample can provide insights that closely resemble those found using a representative sample.

This study adds to the growing literature demonstrating that non-representative convenience sampling can be useful in certain contexts. Other examples include, when modeling behavior that is generally influenced by individual characteristics, incentives, or other behavioral interventions (Camerer, 2011), when validating the predictions of a behavioral model of a well-studied target population (Rosch, 2021), or when investigating topics, such as social preferences and business management (Frigau et al., 2019; Peth & Mußhoff, 2020). Our findings also align with the studies that show that non-representative convenience samples from opt-in, online platforms can replicate the public polling of representative samples (Gelman et al., 2016) and that these nonrepresentative samples can be used to accurately forecast election results (Wang et al., 2015).

2 Methods

2.1 Context of original experiments

The two framed filed experiments examined in this paper were originally published in Ellis et al. (2022), which examined the effectiveness of two stigma-mitigating techniques for food products irrigated with recycled water. Recycled water is a cost-effective, dependable, and safe solution to water scarcity; however, it is often stigmatized by consumers (Savchenko et al., 2018). Destigmatizing recycled water is crucial to the success of large-scale potable and non-potable recycled water initiatives. In the field experiments, the first stigma-mitigating technique tested whether passing recycled water through a natural barrier, such as an aquifer, reduces or removes the stigma consumers would otherwise attach to it. The second technique tested in the experiment was whether a food product's trophic level, or in other words, the food product's original place in the food chain, affects consumer concerns about the use of recycled water.

2.2 In-person, non-representative sample collection

Obtaining a sample that is regionally, let alone nationally, representative for an in-person framed field experiment is regarded as difficult and generally infeasible. Convenience sampling that draws a sample from different segments of the population is often viewed as better than recruiting undergraduate students to a university laboratory since it allows one to obtain samples that are more-representative of the general population. However, convenience sampling is constrained by available sampling locations.

To achieve the most representative sample possible, our in-person field experiment was conducted at three locations in the mid-Atlantic region of the United States, with a day of data collection at each location—the state's largest motor vehicle office, a year-round indoor farmer's market[1], and a super-regional shopping mall patronized by close to 20 million consumers per year. These specific locations, all in Delaware, were chosen because they are visited by a diverse population of adult

[1] In the mid-Atlantic region of the U.S., the term 'farmer's market' is often used to refer to large indoor spaces that offer low-cost groceries along with small booths that offer a variety of other low-cost merchandise and food. These locations often host flea-markets during the weekend, drawing diverse groups of consumers, from both an ethnic and income perspective. The farmer's market that this study was conducted at, thus is not like the high-cost farmer's markets that specialize in local and organic food that can be found in other locations of the U.S.

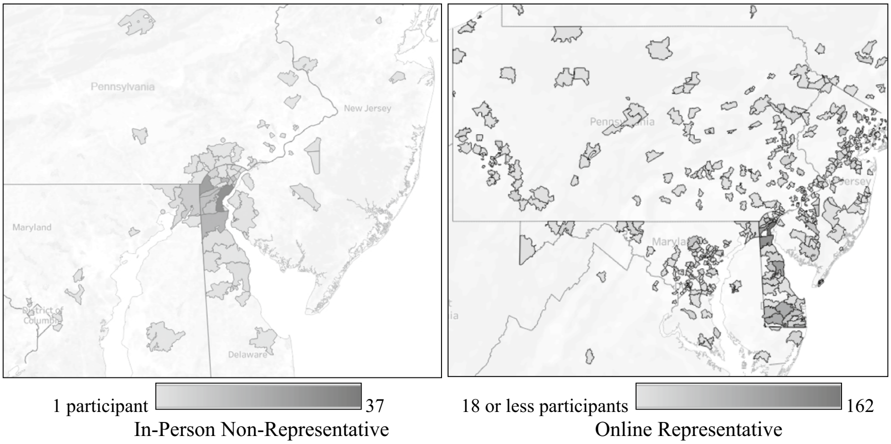

Fig. 1 Regional Distribution of Participants by ZIP code for the In-Person and Online Samples. Source: Ellis et al. (2022)

consumers. The farmer's market and shopping mall are also frequented by consumers from neighboring states, particularly Maryland, New Jersey, and Pennsylvania. Figure 1 shows the density of participants from this study.

In-person participants were recruited by experiment administrators as they walked by the experiment location, incentivized by the opportunity to earn $10. The in-person informed consent form and experiment were presented to participants on tablet computers running a Python-based program. Individuals who were 18 years of age or older and who consented to participate moved from the consent screen to the experiment instructions (see the description of the experimental design below and Appendix A for the experiment instructions).

The products offered to the participants in the in-person experiment were displayed in a central location with all branding information removed. This allowed participants to view and compare the products and reinforced the fact that the participants would be making actual purchasing decisions and paying for the products using real money.

2.3 Online representative sample collection

The online experiment was conducted through Qualtrics to collect a representative sample of Mid-Atlantic consumers. Our power analysis determined that a sample size of 543 participants was needed to ensure adequate power (we ended up collecting data from 546 participants)[2].

[2] We conducted a power analysis using parameter estimates from the in-person sample for the primary variables of interest in Ellis et al. (2022) to determine the online sample size needed to detect changes that would be statistically significant at the 1% level or less. At an 80% power level, the results of 1,000 simulations showed that 543 participants were needed.

Qualtrics conducted recruitment by sending a link to the online experiment to individuals. Individuals who clicked the link were first presented with the experiment consent form. Those who consented were then prompted to provide a valid mailing address. A valid address was important since the experiment required the cash and/or products earned in the experiment be sent to the participant. This helped ensure that the experiment was non-hypothetical. Once participants entered their individual mailing address, they were asked to provide their personal demographic information (age, sex, political affiliation, ethnicity, household income, and highest level of education attained) before proceeding to the experiment instructions and making purchasing decisions.

2.4 Experiment design

The non-representative in-person field experiment and its representative online counterpart were designed to test the effectiveness of several techniques for mitigating the stigma associated with food and drink products produced with recycled water[3]. Participants in both experiments were told they would earn $10 for their participation and that they should think of those funds as a bank account from which they could withdraw money to purchase the offered products[4]. After reviewing the instructions, participants were presented with the following definitions of key terms used in the experiment.

a Recycled water is highly treated wastewater from various sources such as domestic sewage, industrial wastewater, and storm water runoff.
b Groundwater is a source of fresh water that lies in aquifers beneath the land surface.
c An aquifer is an underground body of rock that contains or can transmit groundwater.
d Aquifer recharge is a process that replenishes groundwater stored in aquifers.

Participants were then randomly assigned to a control group or one of three social-marketing treatment groups. The treatments consisted of showing participants a statement making a social comparison and/or a video presenting various public figures promoting products produced with recycled water. The experiment for the in-person non-representative sample was designed so that everyone needed headphones since two of the treatment groups would be watching a video. Thus, the control group watched a ten second video that displayed the following message, "Recycled water purified to drinking water standards is a safe and sustainable water source."

[3] See Ellis, Savchenko, and Messer (2022) for more details and results related to the various treatments.

[4] Online participants received additional compensation from Qualtrics for participating in the experiment. Qualtrics allowed the participants to choose either a gift card or points for online games as additional compensation, but Qualtrics did not disclose the exact amount of that compensation except that it was equivalent to "a few dollars.".

Participants in the first treatment watched a video in which celebrities, including Bill Gates, Jack Black, and Jimmy Fallon; local and state politicians; well-known journalists; and astronauts on the International Space Station, drink potable recycled water. A modified version of the statement from the control group video was displayed during the last ten seconds of the celebrity endorsement video, “Recycled water purified to drinking water standards is a safe and sustainable water source. These people drink it.”

In the second treatment group, participants were presented with a social comparison statement presenting favorable social information about recycled water: “Recycled water purified to drinking water standards is a safe and sustainable water source.”

In the third treatment group, the participants were exposed to both the celebrity endorsement video and the social comparison statement.

In the course of both experiments, participants made a series of purchasing decisions. The in-person participants were presented with fifteen purchase opportunities consisting of five products (bottled water, fresh spinach, frozen lamb chops, cheddar cheese, and hot chocolate mix) produced with three types of water–groundwater, recycled water, and groundwater drawn from an aquifer recharged with recycled water. In addition to receiving the same purchasing options as the participants in the in-person experiment, online participants were presented with three additional purchase opportunities as we added sirloin steak as a sixth product to address concerns that the potential limited appeal of lamb could affect consumer purchase decisions. The following questions were used to present the purchase decisions to participants (using recycled water as an example).

1. Do you want to purchase 16 oz of bottled [recycled water] for $_____?
2. Do you want to purchase approximately 8 ounces of spinach irrigated with [recycled water] for $_____?
3. Do you want to purchase approximately half a pound of lamb chops from lamb that grazed on grass irrigated with [recycled water] for $_____?
4. Do you want to purchase an approximately one-pound block of cheddar cheese made with milk from a cow that grazed on grass irrigated with [recycled water] for $____?
5. Do you want to purchase approximately 16 ounces of hot chocolate mix made with powdered milk from a cow that grazed on grass irrigated with [recycled water] for $___?
6. Do you want to purchase approximately 6 oz of sirloin steak from cattle that grazed on grass irrigated with [recycled water] for $_____?

The products presented to participants were used to test the effect of a product’s trophic level on consumers’ stigmatization of the product. Trophic level refers to an organism’s place in the food chain. Plants, such as spinach, are categorized a trophic level one because they are a primary producer in the food chain (turn light into organic matter). Herbivores, where products such as cheddar cheese, hot chocolate mix, lamb, and sirloin steak come from, are categorized as trophic level two because they consume organisms from trophic level one. Technically trophic levels do not apply to water because it is a chemical substance and not an organism, but for the sake of consistency in this experiment we referred to it as trophic level zero.

Once the purchase opportunities were completed, the screens presented participants with a post-experiment survey that included questions on the frequency at which participants consumed each product. For the in-person sample, the survey included demographic questions that were asked at the beginning of the online sample's experiment to screen participants. After completing the survey, one of the purchasing opportunities was randomly implemented. If a participant chose yes to the randomly selected opportunity, then the participant received the product and the difference between their initial balance of $10 and the price of the product. If the participant chose no, then the participant received only the initial balance of $10.

3 Analysis and results

3.1 Summary statistics

Table 1 presents summary statistics for the in-person and online samples. Among the 371 individuals who participated in the in-person experiment, 314 completed it successfully.[5] The resulting sample is representative regionally for female/male proportion. However, it underrepresents those earning $50,000 or more annually and Hispanic and non-Hispanic white consumers. It oversamples those participants who are 18 to 34 years old at the expense of those who are 55 years and older. Likewise, it oversamples participants who possess some college education (less than a bachelor's degree) at the expense of those who possess a high school diploma or less and those who possess a bachelor's degrees or higher.

In the online sample, 546 participants from states in the U.S. mid-Atlantic region successfully completed the experiment. While the sample is slightly skewed towards females, it is far more representative of the mid-Atlantic region on the basis of educational attainment, ethnicity, income distribution, and age than the in-person sample.

Also presented in Table 1 is the proportion of participants that chose to purchase a product by water type and trophic level treatment. Participants in the online sample chose to purchase all products at a higher rate than participants in the in-person sample. This difference is likely driven by in-person participants being recruited in a motor vehicle office, an indoor farmers' market, and a shopping mall. They may have viewed the possibility of carrying around food and

[5] Participants successfully completed the experiment if they made it through the entire experiment. Participant attrition during the in-person experiment was due to individuals not completing the experiment after they started it.

Table 1 Summary statistics of the in-person and online mid-atlantic (DE, MD, NJ, & PA) Samples

	Population (%)	Online sample (%)	In-person sample (%)
Total Participants		546	314
Female	52	57	51
Educational attainment			
High School or less	41	34	30
Some college	17	21	33
Associate degree	7	8	9
Bachelor's degree	21	22	13
Graduate degree	14	16	14
Ethnicity			
Non-hispanic white	64	70	54
Black	16	17	29
Hispanic	12	6	8
Asian	6	4	7
Other	2	3	10
Income			
\$49,999 or less	35	39	51
\$50,000–\$99,999	29	32	28
\$100,000–\$149,999	17	16	11
\$150,000 and above	19	12	10
Age			
18 – 34	28	31	46
35 – 54	33	37	37
55 and older	38	32	17
Proportion of product's purchased			
Overall		35 (0.50)	23 (0.60)
Trophic level zero		33 (1.20)	30 (1.50)
Trophic level one		38 (1.20)	21 (1.30)
Trophic level two		34 (0.60)	22 (0.80)
Groundwater		38 (0.80)	25 (1.10)
Recharged aquifer		35 (0.80)	25 (1.10)
Recycled water		31 (0.80)	19 (1.00)

The mid-Atlantic (DE, MD, NJ, PA) population summary statistics are from the 2018 American Community Survey. Standard deviations for proportion of product's purchased are reported in parentheses

drink products in these settings as inconvenient. In contrast, participants in the online sample knew that any products they purchased would be mailed directly to them. In both samples, products produced with recycled water were purchased at a lower rate, and products produced with groundwater and groundwater from an aquifer recharged with recycled water were purchased at the same or similar rates. Between the two samples, however, there were no consistent purchase patterns by trophic level.

Table 2 Logistic regression estimated impact of trophic level and water type on likelihood of purchasing, comparing the in-person sample to the online sample.

	Equation 1		Equation 2		Equation 3	
	Coef	S.E	Coef	S.E	Coef	S.E
Price	− 0.363***	0.021	− 0.364***	0.021	− 0.369***	0.022
Trophic level						
One	0.467***	0.074	0.467***	0.074	0.546***	0.082
Two	1.143***	0.087	1.143***	0.087	1.352***	0.095
Water type						
Recharged aquifer	− 0.165	0.090	− 0.209*	0.106	− 0.165	0.091
Recycled	− 0.513***	0.097	− 0.481***	0.117	− 0.515***	0.098
Freq. of consumption						
Trophic level zero	0.158**	0.061	0.159**	0.061	0.159**	0.061
Trophic level one	0.244***	0.066	0.244***	0.066	0.244***	0.066
Trophic level two	1.207***	0.119	1.208***	0.119	1.210***	0.119
Subgroup						
In-person	− 0.371*	0.154	− 0.389*	0.194	0.025	0.180
Interactions						
In-person* Recharged aquifer			0.153	0.202		
In-person*recycled			− 0.118	0.212		
In-person*trophic level one					− 0.178	0.162
In-Person*trophic level two					− 0.598***	0.134
Constant	− 4.910***	0.393	− 4.906***	0.394	− 5.053***	0.394
Total N	14,538		14,538		14,538	
Individuals	860		860		860	

Equation 1 controls for sample, Eq. 2 incorporates interaction terms between water type and sample, and Eq. 3 incorporates interaction terms between trophic level and sample

*$p<0.05$, **$p<0.01$, ***$p<0.001$

3.2 Comparison of in-person and online mid-atlantic subsamples

In Ellis et al. (2022), a logit model with a random effects specification and clustered standard errors was used to analyze (1) whether passing recycled water through an aquifer reduces or removes the stigma consumers would otherwise attach to it and (2) whether a products trophic level affects consumer concerns about recycled water. The study showed that (1) utilizing an aquifer to filter recycled water eliminates the negative perception that consumers have towards it, and (2) consumers' stigmatization of recycled water decreases as the number of levels in the food chain between the organism and the use of the water increases.

To determine whether the results of these hypotheses vary across in-person and online sample, we estimate the same regression, adding in a control for sample:

Table 3 Wald tests for logistic regression estimated impact of water type (Eq. 2) and trophic level (Eq. 3) on likelihood of purchasing, comparing the in-person sample to the online sample

Wald tests		χ^2	P value	BCP
Equation 2 Wald tests (within sample)				
For in-person sample	Ground = Aquifer recharged	0.11	0.743	1.000
For online sample	Ground = Aquifer recharged	3.91	0.048	0.865
For in-person sample	Aquifer recharged = Recycled	14.07	0.000	0.003
For online sample	Aquifer recharged = Recycled	9.79	0.002	0.032
Equation 2 Wald tests (between sample)				
For ground–recharged aquifer	In-person = Online	0.57	0.451	1.000
For aquifer recharged–recycled	In-person = Online	2.56	0.110	1.000
For ground	In-person = Online	4.02	0.045	0.809
For aquifer recharge	In-person = Online	1.61	0.205	1.000
For recycled	In-person = Online	6.81	0.009	0.163
Equation 3 Wald tests (within sample)				
For in-person sample	Trophic zero = Trophic one	6.43	0.011	0.202
For online sample	Trophic zero = Trophic one	44.44	0.000	0.000
For in-person sample	Trophic one = Trophic two	14.74	0.000	0.002
For online sample	Trophic ONE = Trophic two	86.61	0.000	0.000
Equation 3 Wald tests (between sample)				
For trophic zero–trophic one	In-person = Online	1.20	0.273	1.000
For trophic one–trophic two	In-person = Online	10.89	0.001	0.017
For trophic level zero	In-person = Online	0.02	0.890	1.000
For trophic level one	In-person = Online	0.67	0.412	1.000
For trophic level two	In-person = Online	12.41	0.000	0.008

Both the original p values (P value) and the p values adjusted for multiple comparisons using the Bonferroni correction (BCP) are presented

$$\log\left(\frac{D_{ij}}{1-D_{ij}}\right) = \alpha + \beta_1' P_{ij} + \beta_2' W_{ij} + \beta_3' T_{ij} + \beta_4' S_i + \beta_5' X_i + \mu_i + \varepsilon_{ij} \qquad (1)$$

where P_{ij} is the price of participant i's purchase opportunity j, W_{ij} is a vector of dummy variables for irrigation water type, T_{ij} is a vector of dummy variables for trophic levels, S_i is a dummy variable for the sample, X_i is a matrix of control variables representing the frequency of participant i's consumption of the product offered in the purchase opportunities, and $\mu_i \sim N(0, \sigma^2_\mu)$, and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The regression results from Eq. 1, presented in columns 2 and 3 of Table 2, show what the summary statistics in Table 1 suggested, those in the in-person sample were less likely to purchase products than those in the online sample ($p = 0.016$). To determine what was driving this difference and to see if there are differences between sample in the main hypotheses of interest in Ellis et al. (2022), we estimated two modified versions of Eq. 1 that we refer to as Eqs. 2 and 3. Equation 2 incorporates interaction terms between water type and sample

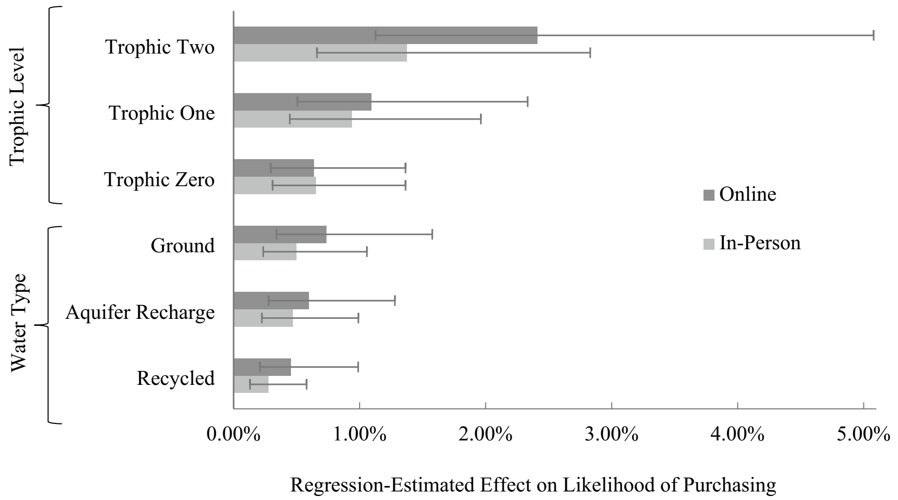

Fig. 2 Regression-Estimated Probability of Purchasing by Trophic Level and Water Type for the In-Person and Online Samples. This figure presents the regression estimated probability of purchasing by trophic level and water type, from least stigmatized to most stigmatized, respectively. See Table 2, Eq. 3 for the complete regression results for the trophic level estimates and Table 2, Eq. 2 for the complete regression results for the water type estimates

($W_{ij}S_i$), while Eq. 3 incorporates interaction terms between trophic level and sample ($T_{ij}S_i$).

The regression results for Eq. 2 are reported in columns 4 and 5 of Table 2 and the Wald test results are presented in Table 3. All reported p-values are adjusted for multiple comparisons using the Bonferroni correction. Wald tests 1 and 2 ($p \geq 0.865$) and Wald tests 3 and 4 ($p \leq 0.032$) show that the main water type results from Ellis et al. (2022) hold for both the in-person and online samples (see Fig. 2 for a visualization of this result and subsequent results). Specifically, passing recycled water through an aquifer removes the stigma consumers would otherwise attach to it. These results are supported by Wald tests 5 and 6 ($p = 1.000$) which find no statistical difference in the effect of passing recycled water through an aquifer between the two samples and Wald tests 7 through 8 ($p \geq 0.163$) which find no statistical difference between the two samples in the regression-estimated log odds ratio of purchasing for each type of water.

Equation 3's regression results are reported in columns 6 and 7 of Table 2, with the corresponding Wald test results presented in Table 3. Wald tests 10 ($p = 0.202$) and 11 ($p = 0.000$) present conflicting evidence between the two samples on whether trophic level one (plants) products inherit all the stigma attached to trophic level zero (water) products. However, Wald tests 16, and 17 ($p = 1.000$) suggest there is no significant difference between the regression-estimated effects of these two trophic levels and Wald test 14 ($p = 1.000$) indicates that there is no significant difference between the trophic level one and zero's effects. This suggests that the stigma mitigating effect on recycled water of going from trophic level zero to trophic level one may be small and possibly tenuous.

In Table 3, Wald tests 12 and 13 ($p \leq 0.002$) show a consistent stigma mitigating effect on recycled water of going from trophic level one to trophic level two (products produced from herbivores). Wald tests 15 and 18 ($p \leq 0.017$) suggest this effect was stronger in the online sample than in the in-person sample. This difference could be due to the in-person logistical challenges involved with a couple of the trophic level two products (cheddar cheese and lamb). As stated earlier, participants in the in-person sample may have found it inconvenient to carry around these products since they were recruited in a motor vehicle office, an indoors farmers' market, and a large shopping mall. Regardless, these findings are consistent with those of Ellis et al. (2022), the greater number of steps between a product and the recycled water used in its production, the less it is stigmatized.

4 Conclusion

While representative samples are considered more externally valid than nonrepresentative samples, our findings suggest that purchasing decisions made by adult participants in our in-person non-representative sample are largely the same as those in our representative online sample. There were only two differences found. First, the stigma mitigating effect on recycled water of going from trophic level zero (water) to trophic level one (plants) was lower or nonexistent for the in-person sample than online sample. This likely indicates that the stigma mitigating effect on recycled water of going from trophic level zero to trophic level one is small and possibly weak. Second participants in the in-person sample were less likely to purchase two products that had the logistical challenge of needing immediate refrigeration compared to the online sample where participants were assured that the product would be delivered on ice. This type of logistical issue should not be a problem for experiments not involving perishable goods and could also be addressed in experiments involving perishable goods by providing participants with a free thermal bag and packaging with an ice pack or offering to have the product delivered on ice to their homes for free.

These findings contribute to a growing body of literature that non-representative convenience samples can be sufficient in certain contexts (Camerer, 2011; Frigau et al., 2019; Gelman et al., 2016; Peth & Mußhoff, 2020; Rosch, 2021; Wang et al., 2015). Representative samples can be difficult and expensive to obtain. Recruitment of our online representative sample was 14% more expensive per participant than the in-person non-representative sample and involved logistical challenges, such as packaging and mailing the payments and products to participants.[6] Requiring every study to have a representative sample could create unnecessary barriers to research that stifles scientific progress. These tradeoffs should be kept in mind when considering sample selection.

[6] It is important to note that our in-person sample was collected near our home institution and thus did not require significant expenses, such as travel, food, and lodging. In-person field experiments can become much more expensive if these types of expenses are required.

If a representative or quasi-representative sample is needed, one could be recruited through the careful selection of in-person experiment locations and quota targeting. While we chose our three in-person experiment locations to achieve a more representative sample of the general adult population than the traditional sample of undergraduate students in a university laboratory, we did not set specific quotas to target. If we were to do this, then we likely would have had to monitor the demographic makeup of our sample during data collection so that we could end data collection at certain locations when we reached certain targets. We consequently would have then needed to add locations to achieve other targets. For example, to recruit a larger proportion of individuals in the 55 and older bracket, we could have recruited participants at the local adult learning center that provides educational programs to retirees.

Testing ideas at scale with representative populations is crucial to ensuring that the research which policy is based on is generalizable to the target population. However, when facing limited resources, researchers often need to make tradeoffs between the cost and ease of collecting a non-representative sample and the importance of external validity. Our study demonstrates that the results from a non-representative convenience sample closely resemble those from a representative sample. This suggests that non-representative convenience sampling can generate useful results and should be used in economic experiments in certain contexts that we know of, and perhaps others, to move science forward.

Appendix: experiment instructions

Printed instructions

Please read these instructions carefully and do not communicate with anyone while you are making your decisions.

- You will earn $10 by participating in this research that you may keep and/or use to purchase food or drink products. You may think of this money as a bank account from which you can withdraw money.
- Depending on the decisions you make, you may receive a combination of cash and food or drink products.
- Your decisions are just like the ones you make in a store: you either buy the product at the listed price or you do not.
- There are no greater physical risks from participating in this study than those you would face in a store. Please remember that all decisions are real purchasing decisions, but only one of your purchasing decisions will be randomly selected and implemented.

Steps

You will face a series of “options” to purchase a product. For each option, decide if you want to buy the product at the listed price by selecting “Yes” or “No.”

Complete a short survey.
Roll a digital die to determine which purchasing option will be implemented (only one will be implemented).
Receive cash and/or product.

Example 1: If you selected Yes for an option that cost $3 and this option is randomly implemented, you will receive the product and $7 cash ($10–$3 = $7).

Example 2: If you selected No for an option and this option is randomly implemented, you will receive $10 cash and will not receive any product.

Funding Funding support for this research was provided by the USDA National Institute for Food and Agriculture (grant number: 20166800725064), which established CONSERVE: A Center of Excellence at the Nexus of Sustainable Water Reuse, Food, and Health and the National Science Foundation's EPSCoR program entitled "Water in the Changing Coastal Environmental of Delaware". Neither funding source had any role in the study design, collection, analysis, and interpretation of data, writing of the manuscript, or decision to submit the manuscript for publication.

Data availability All data and analysis scripts have been deposited in the Open Science Framework (https://osf.io/hduym/).

References

Banerjee, A. V., & Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics, 1*(1), 151–178. https://doi.org/10.1146/annurev.economics.050708.143235
Boas, T. C., Christenson, D. P., & Glick, D. M. (2020). Recruiting large online samples in the United States and India: facebook, mechanical turk, and qualtrics. *Political Science Research and Methods, 8*(2), 232–250. https://doi.org/10.1017/psrm.2018.28
Bol, D. (2019). Putting politics in the lab: a review of lab experiments in political science. *Government and Opposition, 54*(1), 167–190. https://doi.org/10.1017/gov.2018.14
Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1977749
Ellis, S. F., Savchenko, O. M., & Messer, K. D. (2022). Mitigating stigma associated with recycled water: Aquifer recharge and trophic levels. *American Journal of Agricultural Economics, 104*(3), 1077–1099.
Frigau, L., Medda, T., & Pelligra, V. (2019). From the field to the lab. an experiment on the representativeness of standard laboratory subjects. *Journal of Behavioral and Experimental Economics, 78*, 160–169. https://doi.org/10.1016/j.socec.2018.06.003
Gelman, A., Goal, S., Rothschild, D., & Wang, W. (2016). High-frequency polling with non-representative data. In D. Schill (Ed.), *Political communication in real time*. New York: Routledge.
Goldberg, M., Van Der Linden, S., Ballew, M. T., Rosenthal, S. A., & Leiserowitz, A. (2019). Convenient but biased? the reliability of convenience samples in research about attitudes toward climate change. *Osfpreprints., 11*, 4453.
Levitt, S. D., & List, J. A. (2007). Viewpoint: on the generalizability of lab behaviour to the field. *Canadian Journal of Economics/revue Canadienne D'économique, 40*(2), 347–370. https://doi.org/10.1111/j.1365-2966.2007.00412.x
Levitt, S. D., & List, J. A. (2008). Homo economicus evolves. *Science, 319*(5865), 909–910.
List, J. A., & Price, M. K. (2016). The use of field experiments in environmental and resource economics. *Review of Environmental Economics and Policy, 10*(2), 206–225.

Muller, S. (2014). *Randomised Trials for Policy: A Review of the External Validity of Treatment Effects* (Working Paper). Accessed 14 Dec 2020

Palm-Forster, L. H., Ferraro, P. J., Janusch, N., Vossler, C. A., & Messer, K. D. (2019). Behavioral and experimental agri-environmental research: Methodological challenges, literature gaps, and recommendations. *Environmental and Resource Economics, 73*(3), 719–742. https://doi.org/10.1007/s10640-019-00342-x

Palm-Forster, L. H., & Messer, K. D. (2021). Experimental and Behavioral Economics to Inform Agri-Environmental Programs and Policies. In L. H. Palm-Forster (Ed.), *Handbook of Agricultural Economics.* USA: David Just and Chris Barrett.

Peth, D., & Mußhoff, O. (2020). Comparing compliance behaviour of students and farmers an extra-laboratory experiment in the context of agri-environmental nudges in Germany. *Journal of Agricultural Economics., 71*(2), 601–615. https://doi.org/10.1111/1477-9552.12352

Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics, 91*(5), 1266–1271.

Rosch, S., Raszap Skorbiansky, S., Weigel, C., Messer, K. D., & Hellerstein, D. (2021). Barriers to using economic experiments in evidence-based agricultural policymaking. *Applied Economic Perspectives and Policy, 43*(2), 531–555.

Savchenko, O., Kecinski, M., Li, T., & Messer, K. D. (2018). reclaimed water and food production: cautionary tales from consumer research. *Environmental Research., 170*, 320–331.

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting, 31*(3), 980–991.

Weigel, C., Paul, L. A., Ferraro, P. J., & Messer, K. D. (2021). Challenges in recruiting US farmers for policy-relevant economic field experiments. *Applied Economic Perspectives and Policy, 43*(2), 556–572.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.