Original Scholarship

Who Would Pay Higher Taxes for Better Mental Health? Results of a Large-Sample National Choice Experiment

F. REED JOHNSON,* JUAN MARCOS GONZALEZ,* JUI-CHEN YANG,[†] SEMRA OZDEMIR,[‡] and STEVEN KYMES[§]

Duke University; [†]Pacific Economic Research LLC; [‡]Duke-NUS Medical School; [§]H. Lundbeck A/S

Policy Points:

- Public funding for mental health programs must compete with other funding priorities in limited state budgets.
- Valuing state-funded mental health programs in a policy-relevant context requires consideration of how much benefit from other programs the public is willing to forgo to increase mental health program benefits and how much the public is willing to be taxed for such program benefits.
- Taxpayer resistance to increased taxes to pay for publicly funded mental health programs and perceived benefits of such programs vary with state population size.
- In all states, taxpayers seem to support increased public funding for mental health programs such as state Medicaid services, suggesting such programs are underfunded from the perspective of the average taxpayer.

Context: The direct and indirect impacts of serious mental illness (SMI) on health care systems and communities represents a significant burden. However, the value that community members place on alleviating this burden is not known, and SMI treatment must compete with a long list of other publicly

funded priorities. This study defines the value of public mental health interventions as what the public would accept, either in the form of higher taxes or in reductions in nonhealth programs, in return for increases in the number of mental health program beneficiaries.

Methods: We developed and fielded a best-practice discrete-choice experiment survey to quantify respondents' willingness to be taxed for increased spending among several competing programs, including a program for treating severe mental health conditions. A realistic decision frame was used to elicit respondents' willingness to support expanded state budgets for mental health programs if that expansion required either cuts in the competing publicly financed programs or tax increases. The survey was administered to a general population national sample of 10,000 respondents.

Findings: Nearly half the respondents in our sample either chose "no budget increase" for all budget scenarios or had preferences that were too disordered to estimate trade-off values. Including zero values for those respondents, we found that the mean (median) amount that all respondents were willing to be taxed annually for public mental health programs ranged between $156 ($99) per year for large-population states and $343 ($181) per year for small-population states. Respondents would accept reductions of between 1.6 and 3.4 beneficiaries in other programs in return for 1 additional mental health program beneficiary.

Conclusions: Our results are consistent with findings that a substantial portion of the US public is unwilling to pay higher taxes. Nevertheless, even including the substantial number of respondents who opposed any tax increase, the willingness of both the mean and median respondent to be taxed for mental health program expansions implies that programs providing mental health services such as state Medicaid are underfunded.

Keywords: mental health, social values, willingness to pay, discrete-choice experiment.

I T IS WELL KNOWN THAT SERIOUS MENTAL ILLNESS (SMI) IMPOSES a significant social burden on communities. Roehrig reported that in the United States in 2013, direct health care spending by individuals with mental illness exceeded $200 billion (with 40% going to care of institutionalized individuals), accounting for almost 10% of all US health care spending.[1] In the United States, all-cause health care costs are estimated as $51,085 for the first year after diagnosis in patients with schizophrenia and type I bipolar disorder.[2] However, the direct medical cost of SMI represents but a fraction of the total burden that

the condition places on society. Focusing narrowly on direct health outcomes understates the value to patients, their families, and communities of successfully treating SMI. Accounting for the societal costs of unemployment, increased disease prevalence, and other sources of community burden, estimates of the annual global economic burden of mental disorders are between $2.5 trillion and $8.5 trillion.[3]

In 2018, SMI affected 11.4 million (4.6%) of adults in the United States. The prevalence of SMI in the United States has been rising over the last decade, with the greatest increases in prevalence in young adults between the ages of 18 and 25 years.[4] The increased prevalence of SMI has broad ramifications. Economically, individuals with SMI earn a third less than their counterparts who do not live with mental illness,[5] and disease severity is positively correlated with likelihood of long-term unemployment.[6] Individuals with SMI also are more likely than those without SMI to contract HIV[7] and to experience increased mortality from respiratory, digestive, or genitourinary diseases[8] and other conditions.[9] Furthermore, SMI is prevalent in homeless populations, which have higher than average rates of alcohol and drug dependence.[10] Homelessness has been linked particularly to SMI such as schizophrenia or bipolar disorder.[11]

People living with SMI often depend on publicly funded health programs to obtain access to treatment. As consequential as SMI is for individuals, their families, and their communities, effective treatment of mental illness inevitably must compete with a long list of other publicly funded priorities such as education, national defense, poverty, and other health care programs. Allocation of funds raised by taxation among such competing priorities in a complex political system may or may not accurately reflect the concerns and preferences of segments of the general population. Thus, it is difficult to derive principled measures of relative societal values from observed federal and state budget allocations among programs.

This study defines the value of public mental health interventions as what the public would accept, in the form either of higher taxes or of spending reductions for nonhealth programs, in return for increases in the number of mental health program beneficiaries. The research objective was to apply accepted best-practice stated-preference methods[12] to quantify general population state-level estimates of the public's willingness to accept tradeoffs when improved mental health program benefits compete with nonhealth social program benefits or require increased

taxes. Valid and reliable estimates of such values can inform benefit-cost analyses of new programs or expansions of existing programs. In addition, understanding how such values vary by characteristics of individual citizens could be helpful when evaluating potential inequities in the distribution of the perceived benefits of public programs relative to the distribution of the tax burden required to support such programs.

Methods

Conceptual Framework

In this study, we used discrete-choice experiments to quantify the general population's willingness to be taxed for social programs. Choice experiments elicit stated preferences in a survey instrument by simulating decision-making under controlled hypothetical conditions.[13,14] The resulting pattern of choices provides sufficient statistical information to estimate the implicit relative importance preference weights respondents used to evaluate tradeoffs among constructed alternatives.[15]

Benefit-cost studies of government programs conventionally measure households' willingness to pay for the benefits of such programs.[16] Economists generally assume that willingness to pay does not depend on institutional arrangements for obtaining payments. However, this assumption is not consistent with data indicating that many people object to paying higher taxes, even when the value to them of increased services is greater than the value of the associated tax payments.[17,18] Policy-relevant allocation of tax revenues among competing spending priorities requires public-sector decision makers to be concerned with how much taxpayers actually would be willing to be taxed for changes in publicly funded programs. We used a choice-experiment survey to quantify respondents' willingness to be taxed for increased spending among several competing programs, including a program for treating severe mental health conditions.

Survey Development

Choice experiments require identifying the features or attributes used to construct choice-alternative profiles for eliciting respondent preferences. Following good research practices for health-preference studies,[19-22] we identified and defined the attributes and levels in our experiment

in consultation with an advisory panel of subject-matter experts and with mental health patient-organization representatives; additionally, 20 cognitive interviews of taxpayers helped inform the study design. Programs and beneficiary levels required for budget trade-off questions were refined iteratively to focus on a tractable number of publicly funded programs that plausibly could compete with mental-health programs.

After obtaining ethics review from Duke University's institutional review board, we recruited a convenience sample of 20 adults from a registry of patients in a regional medical center who had agreed to participate in research studies. We conducted face-to-face pretest interviews using a "think aloud" protocol. Respondents read the survey text aloud and were encouraged to say whatever came to mind as they obtained information and used the information to answer questions. These interviews guided numerous revisions of the survey instrument to improve readability, clarify ambiguities, and improve the choice-task layout. The interviews also verified that the budget trade-off items were plausible and that respondents readily accepted tradeoffs over the specified range of program benefits and costs. The Online Supplementary Materials contain the final version of the survey instrument.

Figure 1 presents how the survey describes the mental health program without specifying a particular program. Descriptions of the other programs had similar structures and levels of detail. Table 1 summarizes the attributes and levels used in the study, and Figure 2 is an example of the choice-question format.

The following text defined a realistic, policy-relevant decision frame for the choice questions.

"Suppose the [state] legislature were considering expanding an existing mental health program, [one of 4 randomly selected non-health programs], and [a second of 4 randomly selected non-health programs]. Suppose also that your representative sent you a letter asking whether he or she should vote for or against a budget increase. Which option would you want your representative to vote for?"

In addition to a "no budget change" option, the answers for each choice question were a mental health program and two programs randomly selected for each respondent from four options: food safety, disaster relief, unemployment, and motor vehicle safety. The descriptions of the programs used a common structure and level of detail.

Figure 1. Survey Description of Mental Health Program

Severe mental illness often can be treated safely and effectively with available medicines. However, these medicines do not work for about 1/3 of people with the most severe mental illness. Suppose the [STATE] legislature were considering expanding an existing mental-health program to make a new effective medicine available to people whose current treatment is not working for them.

The expanded program would:

- Use prescription information to identify and contact people with mental illness whose treatment is not working;

- Provide free access to the new medicine at community health centers; and

- Help treated patients get access to health-education and medical-care services.

Survey Design

The survey instrument included the following:

- Informed consent
- Attribute descriptions
- Comprehension and reflection questions
- Practice choice questions
- Extensive use of color and graphics
- Preference elicitation choice questions
- Pop-up attribute definition reminders throughout the survey
- Demographic and background questions

An experimental design determined how program benefits and tax levels were combined to describe budget profiles and profile pairings in each choice question. To optimize D-efficiency and maximize the statistical power available to estimate preference weights for program benefits and tax levels, SAS software version 9.4 was used to generate the experimental design. The design was generated with no priors and one restriction that excluded the budget profiles with a tax increase but

Table 1. Budget Features and Feature Levels[a]

Type of Program	Size of Program Benefit for {Small/Medium/Large} States[b]
Mental health	{800/1500/3000} fewer people per year
	{400/500/1500} fewer people per year
	{150/150/500} fewer people per year
	No Budget Change
Two additional programs from:	{800/1500/3000} fewer people per year
	{400/500/1500} fewer people per year
• Disaster relief	{150/150/500} fewer people per year
• Motor vehicle safety	No Budget Change
• Unemployment	
• Food safety	

Increased taxes, per year (per month)	Small and Medium States[b]	Large States[b]	
		Low-Cost Arm	High-Cost Arm
	$60 ($5)	$60 ($5)	$60 ($5)
	$120 ($10)	$120 ($10)	$120 ($10)
	$240 ($20)	$240 ($20)	$240 ($20)
	$480 ($40)	$480 ($40)	$720 ($60)

[a] For the No Budget Change option, benefit and tax levels were set at 0.
[b] States are classified by population size. Small states are Alaska, Arkansas, Delaware, District of Columbia, Hawaii, Idaho, Iowa, Kansas, Maine, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, Utah, Vermont, West Virginia, and Wyoming. Medium states are Alabama, Arizona, Colorado, Connecticut, Georgia, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Virginia, Washington, and Wisconsin. Large states are California, Florida, New York, and Texas.

no program benefits. The final experimental design consisted of 64 unique choice questions. Because answering all 64 choice questions would present an unacceptable burden for a single respondent, the full design was divided into 16 survey versions, each containing five choice questions. Each respondent evaluated one of the versions. The subset of programs shown varied across respondents, but it did not vary within respondents. Both the order of choice questions and the order of programs were randomized within versions.

Figure 2. Example Choice Question Format

Program	No Budget Change	Budget A	Budget B
Food-safety program	No change	400 fewer foodborne illnesses per year	150 fewer foodborne illnesses per year
Motor-vehicle-safety program	No change	800 fewer serious injuries per year	150 fewer serious injuries per year
Mental-health program	No change	No change	400 more people helped per year
Increase in your taxes	No increase	$60 per year ($5 per month)	$480 per year ($40 per month)
Which option would you want your representative to vote for?	☐ No Budget Change	☐ Budget A	☐ Budget B

Data Collection

The online survey was administered to 10,000 US adults from a national consumer panel in November and December 2018. We obtained a minimum of 100 respondents from each state, with the remaining 5,000 respondents allocated among states in proportion to their share of the total US population. Benefits shown in Table 1 varied roughly in proportion to three state population sizes.

Internal Validity Tests

Stated-preference surveys are vulnerable to possible hypothetical bias. While it would be ideal to compare behavioral data against stated-preference data, such comparisons often are not possible. In

this case, the relationships between taxpayers' preferences and actual budget decisions are highly confounded by institutional and ideological factors. Moreover, political processes do not reveal taxpayer preferences at the level of granularity that could be useful in guiding decisions regarding marginal tradeoffs among competing expenditures. Nevertheless, credibility of stated-preference data should be evaluated using tests of whether respondents were attentive to the choice tasks and whether their answers were logical and consistent.[23]

The study design included four tests of internal validity:

- The respondent always chose the same budget scenario (No Budget Change, Budget A, or Budget B) regardless of the tradeoffs involved.
- Logical inconsistency among a participant's responses to choice questions.
- The respondent always chose the budget scenario with the better level of one program or with the lower taxes, regardless of the tradeoffs involved.
- Insensitivity to absolute levels of the tax attribute. Tax levels for the four large states included a split-sample test of sensitivity. Half the sample saw a highest tax level of $480 per year ($40 per month), and half the sample saw a highest tax level of $720 per year ($60 per month).

Estimation

Choice models derive the set of implicit preference weights that is consistent with the observed pattern of choices. Because respondents who picked No Budget Change in all choice questions provided no trade-off information for analysis, they were not included in the statistical modeling. However, we assumed zero willingness to be taxed for those observations and included them in calculating mean and median willingness-to-be-taxed values. While we did not delete any other observations from analysis, we also could not estimate willingness to be taxed for another group of respondents because of disordered, imprecise estimates and thus assigned a zero value also to them. These assumptions ensured that we obtained lower-bound estimates for the actual values.

The latent-class analysis employed categorical variables for all at-
tributes to avoid functional-form assumptions and obtains separate
weights for all attribute levels. While latent-class analysis identifies
preferences that are more similar within classes than between classes,
there still can be taste variability within classes. Assuming taste unifor-
mity can bias estimates.[12] Random-parameters logit accounts for taste
variability in each attribute-level parameter by estimating means and
standard deviations for normally distributed tastes. Statistically signif-
icant standard deviation estimates indicate lack of preference consensus
within latent classes and confirm the importance of controlling for taste
variability.

Aggregate, state-population, size-specific models yielded significant
but disordered categorical coefficients and implausible value estimates.
This result is consistent with previous studies that aggregated groups
of respondents with quite dissimilar preferences.[24] Based on fit, parsi-
mony, and covariate significance, the final specification reported here
is a latent-class specification with three classes, uncorrelated random
parameters, and covariates. Separate models were estimated for small-,
medium-, and large-population states using state-size-specific numbers
of beneficiaries.

Person-Tradeoff Equivalents and Willingness to Be Taxed

A set of estimated relative preference utilities makes it possible to cal-
culate equivalent values. We can calculate "person-tradeoff" decreases in
beneficiaries of any other program that would exactly offset the value
of adding one beneficiary to a given program. In effect, we can mea-
sure the relative value of each kind of program beneficiary denominated
in the base "currency" of equivalent numbers of beneficiaries in some
another program. For each preference class, we calculated the equiva-
lent numbers of beneficiaries in other programs corresponding to a one-
beneficiary increase in a mental-health program. (See the Technical Ap-
pendix for details.) Class-weighted person-tradeoff equivalents were es-
timated by state size, and 95% confidence intervals were estimated based
on 10,000 draws using the Krinsky and Robb procedure.[25]

Willingness to be taxed is directly analogous to calculating person-
tradeoff equivalences among program beneficiaries. In this case, we used
increased taxes as the base "currency." For any change in preference

weight from a given beneficiary increase, we calculated how many tax dollars in a given class would yield the same offsetting change in preference weight. (See the Technical Appendix for details.)

Findings

Demographic Characteristics

Table S1 in the Online Supplementary Materials provides demographic comparisons between the survey sample and the general US population. Our target population was active voters. Registered voters represented 90% of our sample, and 86% of the sample voted in the last four years. In contrast, only 70% and 61%, respectively, of adults in the general US population were registered and voted in the last four years. The sample age and income distributions closely resembled the national distributions, whereas white respondents and respondents with higher educational attainment were overrepresented.

Internal Validity Tests

Of the 10,000 respondents, 5% always chose "Budget A" or always chose "Budget B," which is on par with other studies. (Johnson et al. reported an average of 7% for a sample of 30 published studies.[23]) In addition, 20% of the overall sample in our study always chose "No Budget Change," which we assume is a valid expression of opposition to any of the offered program expansions and increased taxes. About 10% of the sample failed at least one test of logical consistency between questions, which is greater than the average of 6% reported by Johnson and coauthors.[23] About 10% of respondents in our survey always chose the better level of the mental health program. Similar choice patterns for other programs were 5% or less. Results of the test of sensitivity to differences in the largest tax level shown in large states were mixed. Differences in parameter estimates between $480 and $720 for large states were correctly ordered; however, because of wide confidence intervals, the differences were not statistically significant at conventional levels.

Choice Model Estimates

Figure 3 compares relative-importance-weight estimates for medium-population states. Qualitative results were similar for all three state sizes.

Figure 3. Preference Weight, Medium-Population States (95% Confidence Intervals)
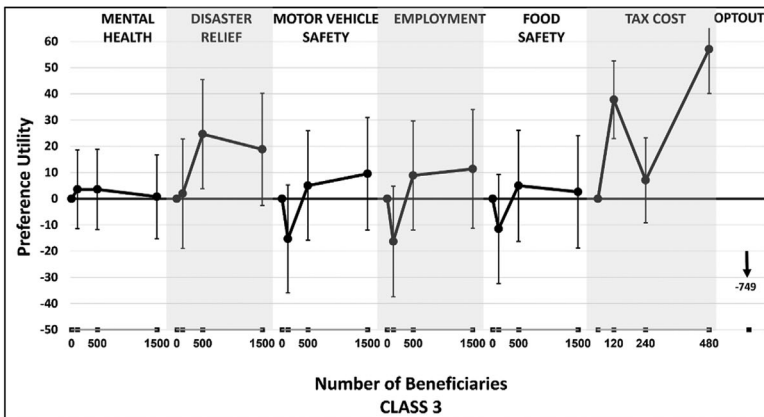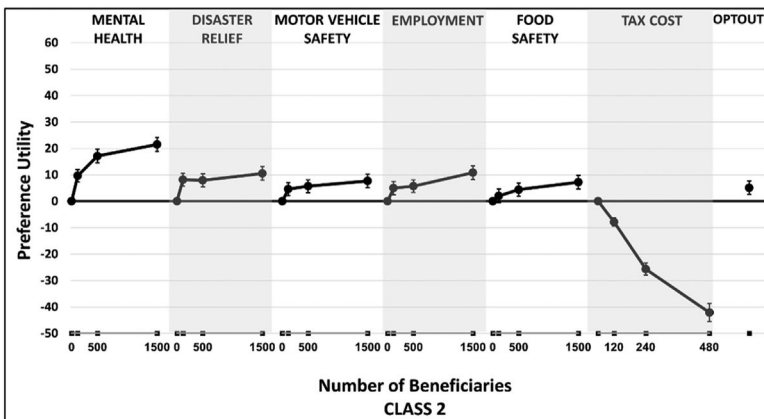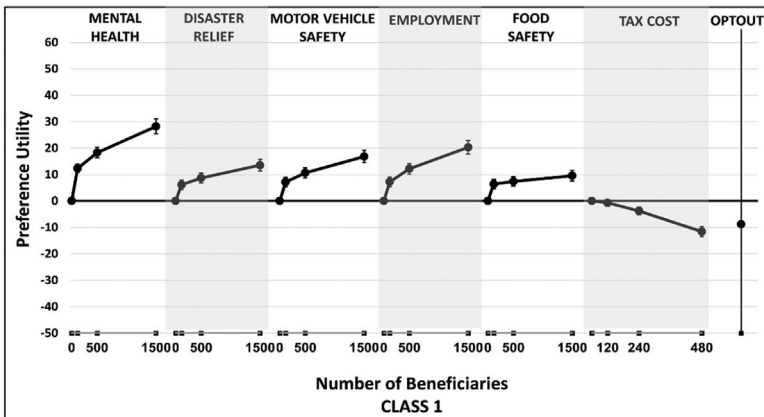
In choice models, only relative preference utility differences matter. Because the absolute scales vary among classes, we normalized the relative scale so the sum of the values of the best level of each attribute equaled 100. Table S2 in the Online Supplementary Materials contains detailed estimates for all three state sizes.

Latent class analysis estimates probabilities of class assignment rather than sorting individual respondents into specific classes. The average probability of assignment to class 1 was 43%. The data points in Figure 2 indicate the relative importance of each level of each budget feature. The best level for each attribute indicates overall relative importance. For respondents with class 1 preferences, a mental health program with 1,500 beneficiaries has the preference weight of about 28. The order of importance of other programs and specified range of tax costs is unemployment (20), motor vehicle safety (17), disaster relief (14), tax cost (12), and food safety (10). Thus, the mental health preference weight of 28 indicates that helping 1,500 people with SMI is nearly three times more important than the least important program, food safety.

The positive preference weights for all programs except food safety are larger than the −1.5 preference weight for the tax cost of $480 per year ($40 per month). This result indicates that subtracting the value of the utility loss of $480 from each program results in net positive values. Hence, the average respondent with class 1 preferences would be willing to be taxed at this level for each program except food safety.

The preference weight for No Budget Change is not plotted, but its value is shown as −8.8. The large negative value indicates that the probability that respondents with class 1 preferences chose No Budget Change effectively is zero; they all chose one of the budget alternatives in every question.

Class 2 represented about 31% of the expected class assignments. In strong contrast to respondents in class 1, respondents with class 2 preferences had a positive preference weight for No Budget Change (0.3), while the preference weight for tax cost was −42. Among the non–mental-health programs, the largest 1,500-beneficiary weights were for disaster relief and employment (both 11). A medium-population state respondent with class 2 preferences would prefer No Budget Change to a budget with the highest mental health benefit level (22) and the highest levels for disaster relief and employment if the tax cost also were set at the highest level. The total benefit utility equals 44 (22 + 11 + 11), and

the tax cost utility equals –42. Thus, the net benefit equals 0.2, which is less than the No Budget Change utility of 0.3.

Class 1 and class 2 estimates show strong nonlinearities, with particularly diminishing marginal values beyond 150 beneficiaries in most programs. This finding seems to indicate a strong preference for some increase in existing spending, but a greater level of comfort with or trust in smaller government-led programs. In class 2, little or no additional value was found for higher levels of disaster relief and motor vehicle safety. Respondents with class 1 preferences showed no difference in taxes between $60 and $120 per year ($5 and $10 per month), whereas class 2 preferences were approximately linear through the full range of tax costs.

Detailed covariate estimates can be found in Table S3 in the Online Supplementary Materials. The differences in covariates for probability of class membership in classes 1 and 2 are logical. Respondents who said they could afford higher taxes and were not opposed to paying taxes had a higher than average probability of being in class 1. In comparison, respondents who said they could not afford higher taxes or indicated they were opposed to paying taxes had a higher than average probability of being in class 2. Respondents had a higher likelihood of being in class 1 or class 2 if they perceived that they paid less than average taxes. Class 1 membership also correlated with respondents' experience with mental illness in their family or friends, and with identifying with a liberal political philosophy. Class 2 membership negatively correlated with experience with mental illness or experience with the outcomes of the other programs, and also negatively correlated with identifying with a liberal political philosophy. Finally, class 2 members were more likely than those in class 1 to be older, spend more time taking the survey, say that they needed more information about how programs would work, doubt whether mental illness programs would work, and be white.

Class 3 represented about 26% of the expected class assignments. Preferences in this class were disordered, with very wide confidence intervals. On average, there were few statistically different preference weights, and respondents in this class were most consistent about not caring about mental health. Their positive value for higher taxes was illogical. Willingness-to-be-taxed calculations require that better levels of each attribute logically be preferred to worse levels. Disordered preferences such as those in class 3 thus provided unusable trade-off information. We assumed that respondents with such preferences had zero

willingness to be taxed. Among the statistically significant covariates for class 3 membership, the largest covariate coefficients were related to respondents being male and to saying they did not need more information about programs, could afford to pay more taxes, were not opposed to paying taxes, and perceived that they paid more than average taxes.

Taste Variability Within Classes

Statistically significant standard-deviation estimates in the random-parameters logit estimates indicate less consensus in within-class preferences. (See Table S4 in the Online Supplementary Materials for parameter estimates.) More than half of the attribute-level standard deviations were statistically significant within several attributes among state sizes and classes. Classes 1 and 2 for all state sizes had statistically significant within-class taste variability for mental-health programs and tax cost. No Budget Change was statistically heterogeneous for class 2 in medium states and class 1 in large states. The greatest preference consensus was for disaster-relief programs, with only one significant standard deviation for all classes in small and medium states. Interestingly, although class 3 preferences for all three state sizes were badly disordered, within-class preferences were quite homogeneous.

Table 2 contains person-tradeoff estimates between mental-health and other programs for all three state sizes and classes. Values indicate how many beneficiaries in another program it would take to produce the same value as one mental health program beneficiary. Values range from about 1 beneficiary in all programs for class 2 preferences in large states to about 5.4 food safety beneficiaries for class 2 preferences in small states. On average, it would take 2 beneficiaries in other programs to produce the same value as 1 mental health beneficiary. Equivalents generally are smallest for large states, with only food safety being greater than 2. Equivalents in small states are greater than those in medium and large states for disaster relief and food safety.

In standard economic theory, optimal resource allocation requires that the ratios of the marginal benefits of two programs be equal to the ratio of their corresponding marginal costs. This result implies that if the ratio between mental health beneficiaries and another program's beneficiaries is 2, the cost of treating mental health patients could be up to twice as large as that of the competing program and still be cost effective.

Table 2. Number of Program Beneficiaries of Equal Value to One Mental Health Beneficiary by State Size and Latent Class

State Size[a]	Class 1	Class 2	Class 3
		Disaster relief	
Small	2.46	3.16	NA
Medium	2.06	1.60	NA
Large	1.94	1.05	NA
		Motor vehicle safety	
Small	1.75	2.40	NA
Medium	1.71	2.53	NA
Large	1.81	1.08	NA
		Unemployment	
Small	1.77	1.74	NA
Medium	1.55	2.23	NA
Large	1.94	1.01	NA
		Food safety	
Small	2.42	5.37	NA
Medium	2.29	3.89	NA
Large	1.88	2.87	NA

Abbreviation: NA, not applicable.
[a] States are classified by population size. Small states are Alaska, Arkansas, Delaware, District of Columbia, Hawaii, Idaho, Iowa, Kansas, Maine, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, Utah, Vermont, West Virginia, and Wyoming. Medium states are Alabama, Arizona, Colorado, Connecticut, Georgia, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Virginia, Washington, and Wisconsin. Large states are California, Florida, New York, and Texas.

Table 3 contains mean and median willingness-to-be-taxed estimates for the mental health beneficiary levels shown in each of the three state-size survey versions. The 95% confidence intervals were estimated by bootstrapping with 10,000 draws. Program-valuation studies typically report mean values. However, the relevant metric for allocating tax revenues among competing programs could be the preferences of the median voter, not the average voter. The proportion of zeros in the sample has a greater impact on medians than on means, and there is a large fraction of zero willingness-to-be-taxed values for program expansions. Respondents located in medium and large states were asked to evaluate program increases of 1,500 beneficiaries. The mean willingness to

Table 3. Mean and Median Willingness to Be Taxed for Mental Health Programs by State Size and Number of Program Beneficiaries

State Size[a]	No. of Beneficiaries	Willingness to Be Taxed	
		Mean, $ (95% CI)	Median, $
Small	800	343 (257-530)	181
	400	274 (206-274)	146
	150	177 (131-274)	94
Medium	1500	404 (290-679)	145
	500	267 (193-447)	115
	150	177 (127-297)	65
Large	3000	156 (115-245)	99
	1500	113 (85-176)	100
	500	65 (43-108)	24

[a] States are classified by population size. Small states are Alaska, Arkansas, Delaware, District of Columbia, Hawaii, Idaho, Iowa, Kansas, Maine, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, Utah, Vermont, West Virginia, and Wyoming. Medium states are Alabama, Arizona, Colorado, Connecticut, Georgia, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Virginia, Washington, and Wisconsin. Large states are California, Florida, New York, and Texas.

be taxed is 3.6 times larger for medium states compared to large states. The mean and median willingness-to-be-taxed values for large states are similar ($113 and $100, respectively). However, there is a large difference between the mean and median values for medium states ($404 and $145, respectively).

The largest program expansion shown to respondents living in small states was 800 beneficiaries—about half of the 1,500-beneficiary level shown to respondents in medium and large states. However, for half the number of beneficiaries, willingness to be taxed in small states is three times the mean and about twice the median of large states.

Discussion

To our knowledge, this study is the largest discrete-choice experiment ever conducted in the social sciences. The study employed a best-practice

stated-preference survey to understand the US public's willingness to be taxed to provide improved care for those living with SMI relative to other publicly funded programs. Our results are consistent with findings that a substantial portion of the US public is unwilling to accept higher taxes.[17,18] Nevertheless, after accounting for respondents who would oppose tax increases, we still found significant support for raising taxes to expand existing public mental illness treatment programs.

State Medicaid programs are an important source of treatment access and payment for SMI patients. The 2017 National Survey on Drug Use and Health reports that over one in four patients with SMI receive coverage through a state Medicaid program, and 48% of the national Medicaid budget goes to pay for the care of individuals with SMI.[4] Our results indicate public support for these programs in budget tradeoffs involving disaster-relief, motor-vehicle-safety, unemployment, and food-safety programs. Overall, respondents were willing to accept reductions of two beneficiaries in other programs for a one-beneficiary increase in mental health programs.

It may seem implausible that willingness to be taxed is lower in more heavily taxed large states with substantial public programs than in lighter-taxed small states with much smaller public programs. However, we framed the preference elicitation in the context of expansions of existing programs. It seems that even if the total value of these programs is large in large states, the marginal value of expanding such programs is less valuable than expanding corresponding programs in smaller states.

Even including the substantial number of respondents who opposed any tax increase, the willingness of both the mean and median respondent to be taxed for program expansions implies that programs providing SMI services, such as state Medicaid programs, are underfunded from the perspective of taxpayers. This finding should be useful to mental health advocates in documenting public support for these programs and to state policymakers considering funding to alleviate the burden of mental illness relative to competing social programs.

Hypothetical choices do not have the same emotional and financial consequences as real choices. Thus, there always is potential for hypothetical bias in choice experiment studies. This study adhered to best practices for limiting hypothetical bias by framing the preference elicitation in a realistic context, defining program processes and benefits carefully, and minimizing the cognitive effort required to evaluate budget alternatives. We mitigated the well-known skepticism many voters

have about the effectiveness of federal programs by framing the preference elicitation in the context of existing state programs. Nevertheless, some respondents remained skeptical of the described program benefits, and this attitude was significant for some class-membership probabilities.

Internal validity generally was comparable with previously published studies, with the exception of the sensitivity to differences in the maximum tax level. Estimates were correctly ordered, but we were unable to reject the null hypothesis because of relatively wide confidence intervals for the four states used for the test.

Because this study was conducted prior to the COVID-19 pandemic, we must acknowledge the possibility that people's attitudes toward health care and government-funded programs could be quite different from those represented during the time our survey was conducted. Moreover, it is not clear what impact pandemic-related population-wide increases in stress and anxiety alongside greater awareness of public health policies and programs might have on the public's assessment of the relative value of mental health programs in comparison to nonhealth programs.

Although our unusually large sample matched several general population characteristics and oversampled our target population of registered voters, members of large consumer panels are not necessarily representative of the general US voting population. Nevertheless, the large proportion of respondents who resisted tax increases is consistent with observed voting patterns. Assuming all those respondents had no value for mental health programs in our calculations of means and medians could suggest that our estimates are lower bounds on actual values.

Conclusion

In the largest choice experiment ever conducted in the social sciences, this study documented significant values for publicly funded mental health programs, both in terms of willingness to be taxed for program expansions and in terms of willingness to accept reductions in the number of beneficiaries in other publicly funded programs. Our results account for a large proportion of respondents who refused to accept tax increases for any purpose. The willingness of both the mean and median respondent to be taxed for program expansions implies that programs

providing SMI services, such as state Medicaid programs, are under-funded from the perspective of taxpayers.

TECHNICAL APPENDIX

Marginal Rates of Substitution (Person-Tradeoffs)

We calculated the utility-equivalent units of program j beneficiaries corresponding to a 1-beneficiary increase in program i for each class as follows:

$$MRS(B_i, B_j) \equiv \frac{\partial V/\partial B_i}{\partial V/\partial B_j} \cong \frac{\left(\beta_{im} - \beta_{in}\right)/\left(B_{im} - B_{in}\right)}{\left(\beta_{j_{max}} - \beta_{j_{min}}\right)/\left(B_{j_{max}} - B_{j_{min}}\right)} \quad (1)$$

where MRS is marginal rate of substitution, B_i is the number of program i beneficiaries, B_j is the number of program j beneficiaries, and V represents the value (utility) function for each program described by the estimated β coefficients from the choice model. We approximated marginal changes for our categorical models as the mean value changes per beneficiary. The numerator is the mean change in utility per beneficiary between any two levels m and n. We calibrated the program j scaling factor in the denominator using the average slope over the full-range difference between the maximum and minimum beneficiary levels.

Willingness to Be Taxed

Willingness to be taxed is directly analogous to calculating MRS equivalences among program beneficiaries. In this case, we use increased taxes as the numeraire. For any change in preference weight, a beneficiary increase from B_{in} to B_{im} for any program i, we calculated how many tax dollars in a given class would yield the same offsetting change in preference weight as follows:

$$WTT(B_{im}, B_{in}) \equiv \frac{\partial V/\partial B_i}{\partial V/\partial Tax} \cong \frac{\left(\beta_{im} - \beta_{in}\right)}{-\left(\beta_{Tax_{max}} - \beta_{Tax_{min}}\right)/(Tax_{max} - Tax_{min})} \quad (2)$$

We otherwise have just substituted the tax slope in the denominator of Equation (1). Because taxes have negative value, there is a minus sign in the denominator to make the ratio positive. This

calculation converts the program-benefit-denominated utility into the equivalent dollar-denominated utility, taking into account any resistance people have to financing the benefits via taxation. We use the estimated slope between $120 and $240 per year to standardize the approximate marginal value of $1 in taxes for a realistic range of tax increases across models. Class-weighted MRS and willingness to be taxed for mental health benefits were estimated by state size.

References

1. Roehrig C. Mental disorders top the list of the most costly conditions in the United States: $201 billion. *Health Aff (Millwood)*. 2016;35(6):1130-1135.

2. Broder MS, Greene M, Chang E, et al. Health-care resource use, costs, and diagnosis patterns in patients with schizophrenia and bipolar disorder: real-world evidence from US claims databases. *Clin Ther*. 2018;40(10):1670-1682.

3. Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? *EMBO Rep*. 2016;17(9):1245-1249.

4. Bose J, Hedden SL, Lipari RN, Park-Lee E. *Key substance use and mental health indicators in the United States: results from the 2017 National Survey on Drug Use and Health*. US Department of Health and Human Services Substance Abuse and Mental Health Services Administration. https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.pdf. Published September 2018. Accessed February 20, 2021.

5. Levinson D, Lakoma MD, Petukhova M, et al. Associations of serious mental illness with earnings: results from the WHO World Mental Health surveys. *Br J Psychiatry*. 2010;197(2):114-121.

6. Goldberg RW, Lucksted A, McNary S, Gold JM, Dixon L, Lehman A. Correlates of long-term unemployment among inner-city adults with serious and persistent mental illness. *Psychiatr Serv*. 2001;52(1):101-103.

7. Cournos F, McKinnon K, Sullivan G. Schizophrenia and comorbid human immunodeficiency virus or hepatitis C virus. *J Clin Psychiatry*. 2005;66(suppl 6):27-33.

8. Newman SC, Bland RC. Mortality in a cohort of patients with schizophrenia: a record linkage study. *Can J Psychiatry*. 1991;36(4):239-245.

9. Dixon L, Postrado L, Delahanty J, Fischer PJ, Lehman A. The association of medical comorbidity in schizophrenia with poor

physical and mental health. *J Nerv Ment Dis*. 1999;187(8):496-502.

10. Fazel S, Khosla V, Doll H, Geddes J. The prevalence of mental disorders among the homeless in Western countries: systematic review and meta-regression analysis. *PLoS Med*. 2009;5:e225.

11. Ayano G, Tesfaw G, Shumet S. The prevalence of schizophrenia and other psychotic disorders among homeless people: a systematic review and meta-analysis. *BMC Psychiatry*. 2019;19(1):370.

12. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CG, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value Health*. 2016;19(4):300-315.

13. Champ P, Boyle K, Brown T, eds. *A Primer on Nonmarket Valuation*, 2nd ed. New York: Springer; 2017..

14. McFadden D. Economic choices. *Am Econ Rev*. 2001;91:351-378.

15. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3-13.

16. Circular A-4: regulatory analysis. *Fed Regist*. 2003;68(106): 58366. https://www.federalregister.gov/documents/2003/10/09/03-25606/circular-a-4-regulatory-analysis.

17. Foucault M, Seki K, Whitten GD. Good times, bad times: taxation and electoral accountability. *Elect Stud*. 2017;45:191-200.

18. Enachescu J, Olsen J, Kogler C, Zeelenberg M, Breugelmans SM, Kirchler E. The role of emotions in tax compliance behavior: a mixed-methods approach. *J Econ Psychol*. 2019;74:102194.

19. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health*. 2011;14(4):403-413.

20. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics*. 2019;37(2):201-226.

21. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*. 2014;32(9):883-902.

22. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145-172.

23. Johnson FR, Yang J-C, Reed SD. The internal validity of discrete choice experiment data: a testing tool for quantitative assessments. *Value Health*. 2019;22(2):157-160.

24. Zhou M, Thayer WM, Bridges JFP. Using latent class analysis to model preference heterogeneity in health: a systematic review. *Pharmacoeconomics*. 2018;36(2):175-187.

25. Krinsky I, Robb A. On approximating the statistical properties of elasticities. *Rev Econ Stat*. 1986;68(4):715-719.

Address correspondence to: F. Reed Johnson, Duke Clinical Research Institute, Duke University, 200 Morris St, Durham, NC 27701 (email: reed.johnson@duke.edu).

Supplementary Material

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1468-0009:

Table S1. Demographic Comparisons Between Survey Sample and General US Population
Table S2. Choice-Model Parameter Estimates by State Size and Latent Class
Table S3. Statistically Significant Covariates for Probability of Latent Class Membership by State Size and Class
Table S4. Statistically Significant Random-Parameter Logit Standard Deviations for Taste-Parameter Distributions