Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

ScienceDirect

Methodology

Affordability and Value in Decision Rules for Cost-Effectiveness: A Survey of Health Economists

Alyssa Bilinski, PhD, Evan MacKay, AM, Joshua A. Salomon, PhD, Ankur Pandya, PhD

ABSTRACT

Objectives: New health technologies are often expensive, but may nevertheless meet standard thresholds for cost effectiveness, a situation exemplified by recent hepatitis C cures. Currently, cost-effectiveness analysis (CEA) does not supply practical means of weighing trade-offs between cost-effectiveness and affordability, particularly when costs and benefits are temporally separated and in health systems with multiple payers, such as the United States. We formally characterized disagreements in CEA theory and identified how these trade-offs are presently addressed in practice.

Methods: We surveyed 170 health economics researchers.

Results: When presented with a hypothetical cost-effective drug therapy in the United States that would require 20% of a state's Medicaid budget over 5 years, 34% of survey respondents recommended that policy makers fund the drug for all patients and 26% for a subset. By contrast, 26% recommended against funding the drug. We found additional disagreement regarding whether the willingness-to-pay threshold should be based on the budget (42%) or societal preferences (41%) and identified 4 approaches to weighing cost-effectiveness and affordability. A total of 61% of respondents did not believe that the threshold used in their last article (most often 1×-3× per capita gross domestic product) represented either the budget or societal willingness-to-pay threshold.

Conclusions: We use these findings to recommend metrics that can inform translation of CEA theory into practice. By contextualizing cost and value, researchers can provide more actionable policy recommendations.

Keywords: allocative efficiency, budget impact, cost-benefit analysis, cost-effectiveness, government expenditures and health, health.

VALUE HEALTH. 2022; 25(7):1141–1147

Introduction

When hepatitis C drug Sovaldi was released in the United States in 2013, its sticker price was $84 000 for a 12-week course of treatment. Even at this price, researchers agreed that the drug was "cost-effective" when judged against typical thresholds for the United States, costing < $100 000 to produce 1 additional quality-adjusted life-year (QALY).[1] Nevertheless, governments and insurers have struggled to pay for it.[2-4] Today, many patients in the United States still lack access to the drug.[5]

Sovaldi is one example of a growing phenomenon: the cost of drugs and medical devices has been increasing, driven in part by new health technologies.[2] Nevertheless, it is common for pricey technologies to have sizable health benefits and meet benchmarks for cost-effectiveness. A systematic review of published cost-effectiveness analyses (CEAs) found many authors characterize interventions as both cost-effective and unaffordable.[6] Health payers in the United States then face the dilemma of whether to strain limited budgets to pay for these services.

CEA can inform these decisions. CEA theory generally suggests lowering the cost-effectiveness threshold after funding interventions with nonmarginal budget impacts.[7] Nevertheless, this assumes there exists a single payer seeking to maximize aggregate health in the population. In fragmented health systems, such as the United States, it is more complex to identify opportunity cost-based thresholds. Moreover, some argue implementing more stringent cost-effectiveness thresholds for high-cost services unfairly penalizes high-value interventions with a large target population.[8]

In this article, we help bridge the gap between CEA theory and policy recommendations from academic researchers by characterizing challenges and disagreements in applying CEA. We report results of a survey of researchers in which we assessed opinions about trade-offs between cost-effectiveness and affordability. We use these findings to propose simple metrics to aid decision making for expensive, high-value programs and discuss implications for evaluating and adopting health interventions.

1098-3015/$36.00 - see front matter Copyright © 2022, International Society for Pharmacoeconomics and Outcomes Research, Inc. Published by Elsevier Inc.

Rationale for Survey Design

CEA assumes that allocation decisions are made in the following framework. Suppose a policy maker has a fixed budget and a set of possible interventions, each of which has an incremental cost-effectiveness ratio (ICER). Each ICER represents the ratio of incremental health gained by the intervention, divided by its incremental cost. To maximize health, decision makers rank interventions according to ICER, smallest to largest. They then purchase interventions in order of ICER. The ICER of the most expensive purchased intervention is the willingness-to-pay (WTP) threshold and represents the opportunity cost of displacing the lowest value (highest ICER) intervention to add an alternative.

Therefore, if an expensive but high-value intervention becomes available, it is added to the ranking to determine whether it should be adopted. If the new intervention has a lower ICER than existing interventions but takes up a large portion of the budget, population health may be maximized by lowering the threshold and excluding lower-value services. This would mean adding the new, higher-value intervention and eliminating existing, less efficient interventions. Nevertheless, in practice, stopping funding for existing interventions may be impractical or politically unpopular. An alternative is increasing the budget, which may allow policy makers to afford all of the previous interventions as well as the new intervention. In some cases, because expensive but cost-effective interventions often have short-term costs and long-term benefits,[6] payers may be able to distribute costs over time, reducing the strain on the budget.

The best course of action depends on societal preferences, the flexibility of the budget, pricing options, and expected future technologies. However, formal application of this optimization approach would constitute a complex multiperiod optimization problem. In lieu of this, most health systems rely on simple heuristics to select WTP thresholds and make binary classifications of cost-effectiveness for each new intervention.

There has been considerable controversy over how to select practical WTP thresholds. Some argue for a fixed budget–based WTP threshold that roughly approximates the opportunity cost of displacing current technologies to adopt new ones. For example, with a simplified empirical version of the above optimization problem, researchers estimated that funding interventions with an ICER above of £13 000 is inefficient in the United Kingdom, below the standard benchmark of £20 000 to £30 000.[7,9,10] In this view, labeling an intervention with a low ICER as “cost-effective,” regardless of price, is inappropriate.

Others argue for a higher societal threshold. They believe that societal preferences generally differ from the existing budget and that the government should increase funding to pay for expensive interventions that improve health.[11-15] These thresholds might also reflect the value of common interventions in the existing system, with the argument that payers implicitly endorse this threshold.[16] Both a “budget threshold” and “societal threshold” could be deemed a “WTP threshold.”

Furthermore, determining appropriate thresholds is even more challenging in decentralized health systems, like the United States, where there many different healthcare payers, including the government, private insurers, and private citizens. Traditionally, interventions are deemed cost-effective in the United States if they fall below $50 000 to $150 000 per QALY, approximately 1 to 3 times per capita gross domestic product (GDP), a measure of "societal threshold."[16] A recent estimate of the empirical threshold for individuals with private insurance was similar at $100 000, but was sensitive to estimates of health insurance on mortality and premium increases on uninsurance.[17]

In our survey, we sought to understand how researchers providing cost-effectiveness-based policy recommendations interpret tension between cost-effectiveness and affordability in an applied US context: with multiple payers and CEAs that only evaluate 1 intervention.

Methods

Survey Design

Questions were designed to reflect standard terminology in academic CEA. After drafting, the questionnaire was first reviewed by experts in cost-effectiveness and survey design. It was then piloted with several PhD students and professors who were asked to explain their reaction to each survey question and the thought process entailed in making answer selections. These responses were used to revise questions, after which the survey was sent to approximately 10 additional PhD students and professors to fill out and provide written feedback. We also performed item analysis to assess correspondence between sections (i.e., case study and grid questions).

Sampling

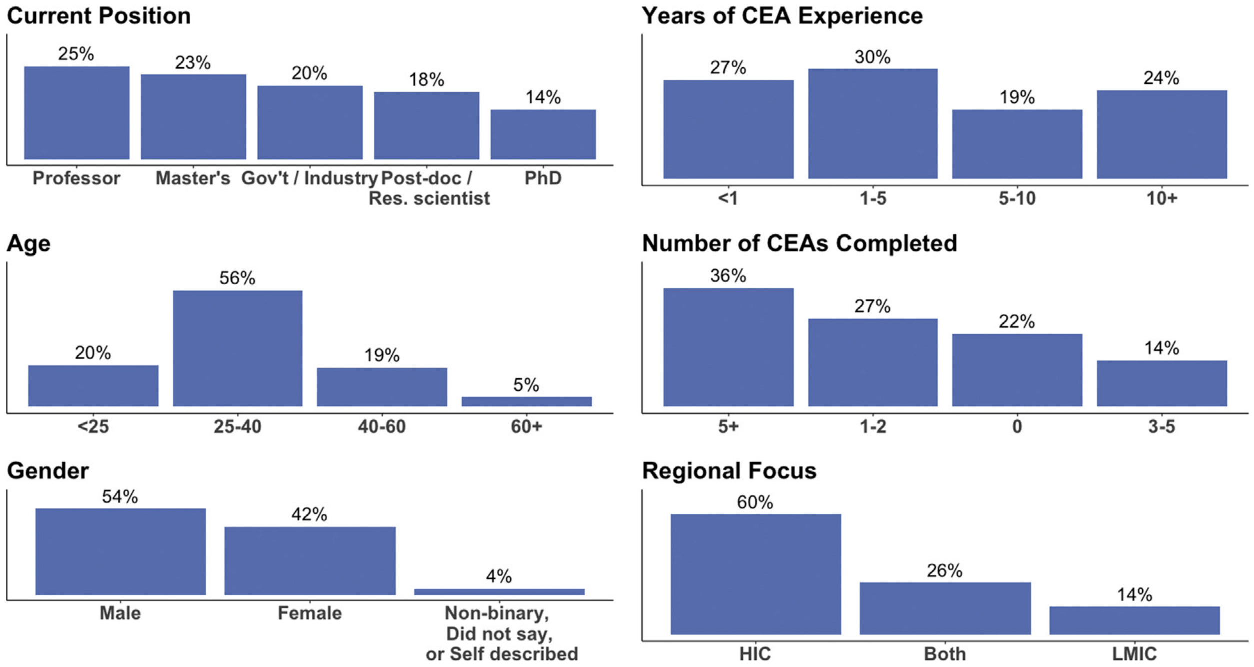
To obtain a purposive sample, we sent invitations to a Qualtrics survey to a convenience sample of approximately 60 professors, PhD students, and industry contacts at 35 medium to large research institutions, all but 2 of which were in the United States and the United Kingdom. Most invitations were sent to individuals studying or working in academic institutions, with approximately 20% sent to professionals in industry. We adopted a “snowball sampling” approach, encouraging individuals to pass the survey on to potential respondents to reach a broader pool of the medical decision making research community. We also placed an invitation in the Society for Medical Decision Making newsletter. Respondents were required, at a minimum, to have taken at least 1 course in CEA or to have contributed to at least 1 cost-effectiveness analysis. Because we wanted to highlight ongoing debate in the health economics community regarding how CEA should be interpreted, we did not survey individuals who lacked experience with CEA. The survey was open for 2 months during the spring of 2018. The study was deemed exempt from review by the Harvard T.H. Chan School of Public Health (Boston, Massachusetts) Institutional Review Board (IRB18-0259).

Survey Questions

The survey asked participants to respond to a case study of a hypothetical newly approved medication, in the format of traditional academic CEA.[18,19] “Drug X” had an ICER of $40 000/QALY. (Traditionally ICERs below $50 000 are “very cost-effective” in the United States.[16]) Nevertheless, to provide this drug to the entire eligible population would cost approximately 20% of the current Medicaid budget over the next 3 years. The survey asked how researchers would describe this finding in the discussion section of an academic article. It also asked how they would advise a state health commission on whether to adopt drug X in its budget.

We then asked respondents about the cost-effectiveness theory relevant to these decisions: how the WTP threshold (ie, the highest ICER considered cost-effective) should be selected and whether it should reflect the budget constraint or a societal measure of WTP. Participants were asked to fill in a grid of policy recommendations for interventions with different population-level costs and ICERs. In the survey, we defined costs as “total costs for the eligible patient population (‘budget impact’),”

Figure 1. Demographic and professional characteristics of the sample (N = 170). The bars and labels indicate the percentage of respondents in each category. Regional focus had 2 nonrespondents (n = 168).



CEA indicates cost-effectiveness analysis; Gov, Government; HIC, high-income country; LMIC, low- to middle-income country; Res., research.

referring to liquid capital required to provide drug X over a short time horizon. We chose the categories for cost-effectiveness to align with GDP per capita, at $< 1\times$, $1\times$ to $2\times$, $2\times$ to $3\times$, and $> 3\times$. Categories for affordability ranged from low to high cost, benchmarked against the "alarm bell threshold." In the survey, we explained that the "alarm bell threshold," defined by the Institute for Clinical and Economic Review, is the "amount of net cost increase per individual new intervention that would contribute to growth in overall healthcare spending greater than the anticipated growth in national GDP + 1%."[20] For each combination of cost and ICER, researchers could mark the intervention as green ("should fund"), white ("unclear"), or red ("should not fund"). Finally, the survey asked participants the threshold used in their most recent analysis, the geographic location for their usual CEA, and demographic information. The full survey can be found in Supplemental Materials.

Analysis

We calculated descriptive statistics (means, counts, and percentages) of responses to the case study questions and demographic data. For grid responses, we identified clusters of individuals with different preferences for trading off cost-effectiveness and affordability. We first considered prespecified clusters of "ICER hawks" (making judgments primarily based on ICER), "budget hawks" (making judgments primary based on cost), and "moderates" (using a mixture of the 2). Based on preliminary findings, we further subdivided ICER hawks into "hard" (recommending all programs with an ICER $<$ \$50 000/QALY) and "soft" (not recommending against any programs with an ICER of $<$ \$50 000/QALY). We also explored differences in grid responses by demographic characteristics. We quantified these, as well as concordance between clusters and Drug X recommendations, using 2-sample z-tests of proportion where expected counts were $>$ 10 and Fisher's exact tests otherwise.

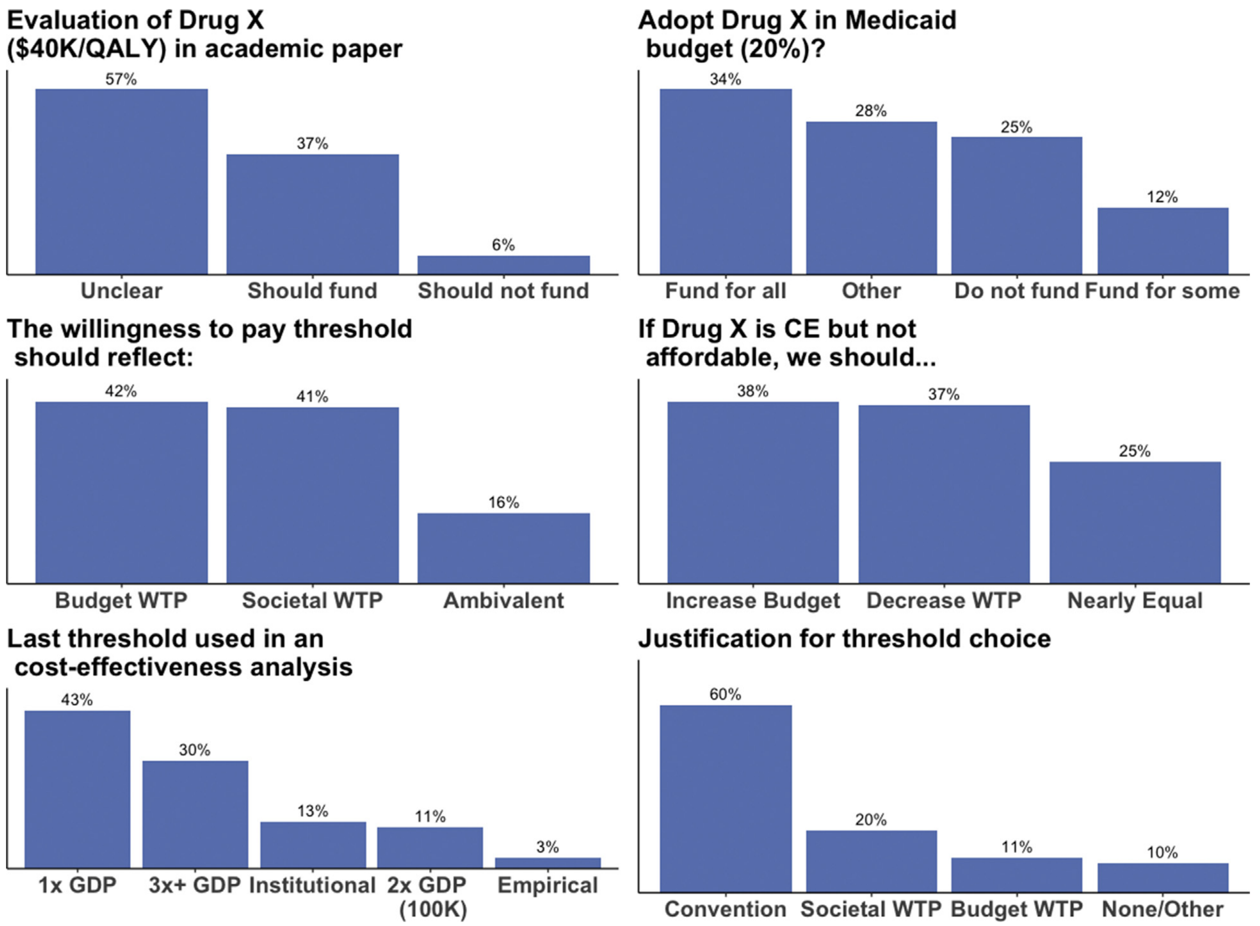
We used logistic regression to understand how theory and policy opinions varied by demographic characteristics. For these regressions, outcomes were binary response variables: (1) whether to recommend funding an expensive but cost-effective drug, Drug X, in an academic article and (2) whether to recommend funding Drug X to a policy maker. Our prespecified predictors, selected based on substantive relevance, included position, gender, years of experience, and country focus. We reported average marginal effects and pseudo-R^2. For statistical tests, we determined statistical significance with a threshold of .05 but also noted *P* values between .05 and .1 ("weak evidence"). In addition, we provide Bonferroni-corrected exploratory subgroup analyses, which were not prespecified, in Appendix Table S2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375.

Results

Demographics

There were 170 respondents, of whom 136 (80%) were from academia, 10 (6%) were from industry, and 13 (8%) were from government (Fig. 1). Of the sample from academia (n = 136), 32% (n = 43) were professors, 23% (n = 31) were in other post-PhD positions (e.g., research scientist, postdoctoral fellow), 17% (n = 23) were PhD students, and 29% (n = 39) were Master's students. Overall, 141 respondents (83%) had participated in at least 1 CEA, with 86 (51%) completing $\geq$ 3. The sample was 54% (n = 91) male. Most respondents were between 25 and 40 years of age (n = 96, 56%), with 20% (n = 34) younger than 25 years and 24% (n = 40) older than 40 years. In addition, most (n = 100, 59%) focused on high-income countries while 14% (n = 23) focused on low-income countries, and 25% (n = 43) a mixture of the 2. More demographic information is shown in Appendix Table S1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375.

Figure 2. Responses to survey questions (n = 170, 170, and 113 for top, middle, and bottom rows, respectively). The bars and labels indicate the percentage of respondents in each category. The top row contains questions about Drug X, a cost-effective drug ($40 000/QALY) that would require 20% of the budget over 3 years to provide to all individuals for whom it is clinically indicated. The second row asks about interpretation of the WTP threshold. A budget WTP means researchers should select the set of CE interventions based on the current budget. A societal WTP means that researchers should derive the threshold based on societal health preferences, even if this would require increasing the budget. A higher WTP threshold indicates that a greater number of interventions are cost-effective. The third row asks about the threshold used most recently by the respondent (left) and the reason for this choice (right). In the third row, respondents could select multiple options and thus percentage sum above 100%. For example, 3 people selected both societal and budget WTP in their justification for threshold choice and are included in both columns.



CE indicates cost-effective; K, thousand; QALY, quality-adjusted life-year; WTP, willingness-to-pay.

Case Study

When asked to evaluate a high-cost, high-value drug, "Drug X," 37% of individuals (n = 63) said they would recommend in an academic article that Drug X be funded, whereas 57% (n = 97) were unsure (Fig. 2, first row). Only 6% of the sample (n = 10) reported that it would recommend against funding the drug. We found similar results when respondents were asked how they would advise a policy maker: 34% would recommend funding drug X for all patients and 26% for a subset of patients, whereas 26% would recommend against funding drug X until the price decreased. (See Appendix Tables S2 and S3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375 for additional breakdowns.)

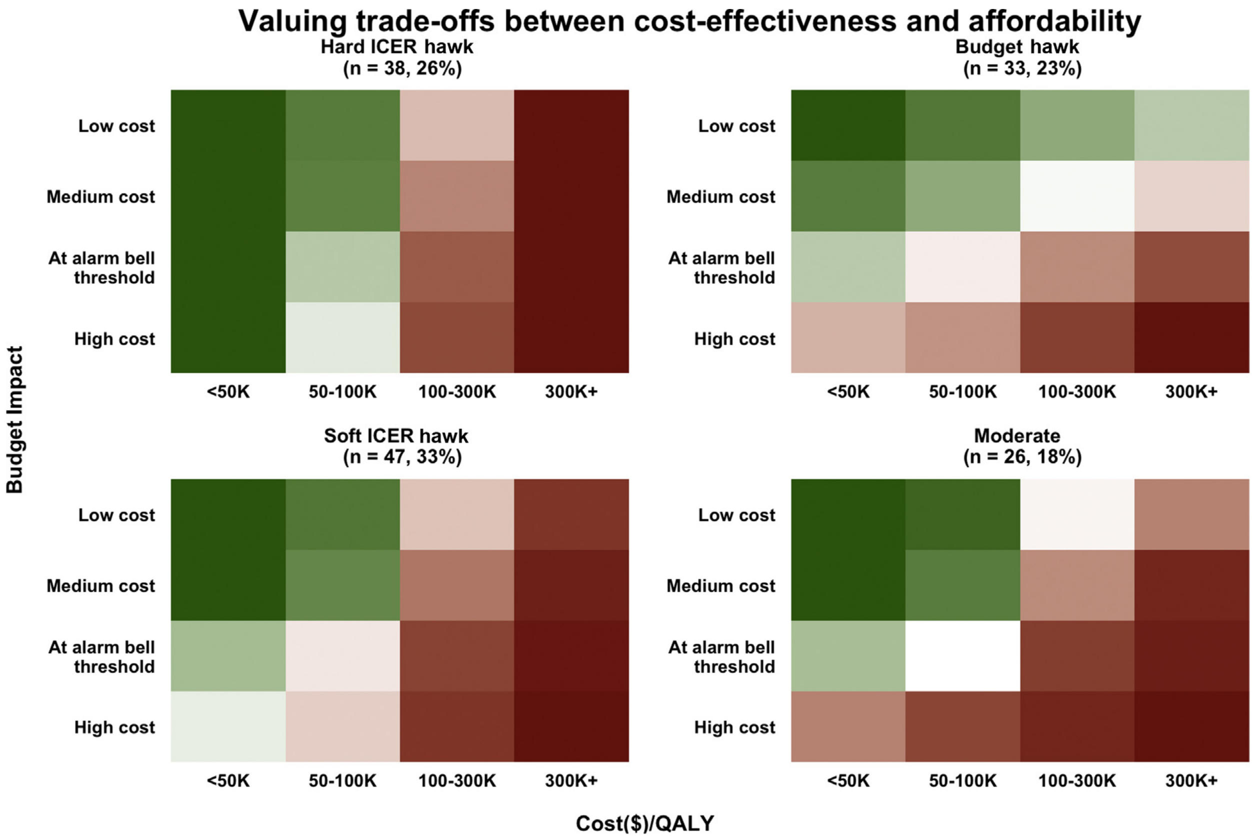
Cost-Effectiveness Theory and Practice

Forty-two percent of researchers (n/N = 72/170) thought that the cost-effectiveness threshold in the academic literature should be derived from the budget, meaning policy makers should delineate cost-effective interventions based on current spending (Fig. 2, second row). By contrast, 41% (n = 70) thought this should be based on societal WTP, even if it would require increasing the budget to implement cost-effective interventions. The remaining 16% (n = 28) felt nearly equal about both types of thresholds.

If a program is cost-effective but not affordable, policy makers could increase the budget to pay for it or choose a lower threshold (rendering the intervention not cost-effective). Our sample was equally split regarding which option is generally appropriate, with 38% (n = 64) saying that policy makers should generally increase the budget, 37% (n = 63) saying that they should usually lower the threshold, and 25% (n = 43) nearly equal about both.

In practice, 75 of 123 researchers (61%) who had completed a CEA selected the threshold for their most recent article based solely on convention; 22 (18%) believed it represented the true societal threshold and 11 (9%) the true budget threshold; and 3 (2%) believed it represented both the budget and societal threshold. The most common choice of threshold was 1- or

Figure 3. Trade-offs between cost-effectiveness and affordability (n = 144). Respondents determined whether each square in the grid should be green (would recommend intervention), red (would not recommend intervention), or white (neutral) in the US context. The x-axis includes cost-effectiveness ($/QALY), with larger values indicating less cost-effective services. The y-axis shows cost, representing total costs for the eligible patient population ("budget impact"). The alarm bell threshold is the "amount of net cost increase per individual new intervention that would contribute to growth in overall healthcare spending greater than the anticipated growth in national GDP + 1%."[19] The color of the square indicate the average response.



GDP indicates gross domestic product; ICER, incremental cost-effectiveness ratio; K, thousand; QALY, quality-adjusted life-year.

3-times GDP, accounting for 73% of the thresholds in our sample (n = 96). Institutional thresholds (such as that set by the National Institute for Clinical Excellence in the United Kingdom) accounted for 13% of respondents (n = 17), and only 3% (n = 4) reported using an empirical threshold based on the budget.

Clusters

Using a grid of policies with varying levels of cost-effectiveness and affordability, we grouped respondents (n = 144) into 4 categories based on how they weighed cost-effectiveness and affordability (Fig. 3,[20] full sample grid and additional breakdowns in Appendix Figs. S1 and S2A-E in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375). "Hard ICER hawks" (26%, n = 38) primarily considered ICERs, recommending all programs with an ICER < $50 000/QALY, regardless of cost. "Soft ICER hawks" (33%, n = 47) gave more weight to cost, but did not recommend against funding any interventions with an ICER < $50 000/QALY.

By contrast, "budget hawks" (23%, n = 33) weighed cost above value, typically approving all low-cost interventions regardless of value and disapproving of all high-cost interventions regardless of value. Finally, "moderates" (18%, n = 26) did not approve of either high-cost or low-value interventions, tending to recommend against programs both along the high-cost and the low-value axes.

A notable indicator for these groups is their response in the high-cost, high-value corner of Figure 3,[20] evaluation of a high-cost program with a very cost-effective ICER: both hard and soft ICER hawks favored funding this type of program, whereas budget hawks and moderates did not.

Comparing these categories with other survey responses as shown in Appendix Figure S3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375, 71% of hard ICER hawks (n = 27/38) reported they would recommend Drug X in an academic article compared with 38% of moderates (n = 10/26), 15% of soft ICER hawks (n = 7/47), and 18% of budget hawks (n = 6/33). Differences between Hard ICER haws and other groups were statistically significant ($P = .02$, $P < .001$, $P < .001$, respectively).

There was weak evidence that respondents who focused primarily on LMIC were more likely to be budget hawks (21% vs 6%, $P = .05$) and that those from non-US developed countries were more likely to be ICER hawks (76% vs 53%, $P = .07$) (breakdowns by demographic characteristics in Appendix Fig. S1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375).

Regression Analysis

We report regression results in Table 1. A model with gender, position (i.e., academic vs industry), years of experience, and geographic focus had a pseudo-R^2 of 3% when the outcome was

Table 1. Logistic regression results.

Predictor	(1) Recommend in an academic article?	(2) Tell policy maker to fund?
	Adjusted marginal effect (95% CI)	Adjusted marginal effect (95% CI)
Years of experience		
<1	(Ref)	(Ref)
1-5	−0.03 (−0.26 to 0.20)	−0.16 (−0.07 to 0.38)
6+	−0.04 (−0.25 to 0.18)	−0.02 (−0.22 to 0.18)
Country focus		
Low- and middle-income countries	(Ref)	(Ref)
A combination of both	0.22 (−0.03 to 0.46)	0.01 (−0.24 to 0.27)
High-income countries	0.14 (−0.06 to 0.34)	0.03 (−0.19 to 0.25)
Gender identity		
Male/other	(Ref)	(Ref)
Female	0.10 (−0.06 to 0.26)	0.13 (−0.02 to 0.29)
Position		
Academic	(Ref)	(Ref)
Government	0.13 (−0.17 to 0.43)	−0.04 (−0.32 to 0.23)
Industry	−0.15 (−0.40 to 0.10)	−0.06 (−0.34 to 0.22)
Other	0.07 (−0.25 to 0.38)	0.05 (−0.25 to 0.35)

Note. For outcomes, 1 = "yes." Average marginal effects are presented with CIs in parentheses. Academia includes professors, postdoctoral/research scientists, PhD students, and Master's students (breakdowns in text). No coefficients were statistically significant at $P < .05$.
CI indicates confidence interval; Ref, referent.

whether respondents would recommend drug X in an academic article and 4% with the outcome of whether respondents would recommend it to a policy maker. Point estimates for these effects were generally small, although wide confidence intervals (CIs) meant that we could not rule out meaningful effects. The only coefficient that was statistically significant at the 10% level was gender in the policy maker regression, with females 13% more likely to recommend the intervention to policy makers (95% CI −0.02 to 0.29) conditional on other covariates. (See more control variables in Appendix Tables S4 and S5 and alternative specifications in Appendix Tables S6 and S7 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.11.1375.)

Discussion

As high-cost technologies proliferate, countries must make difficult trade-offs among healthcare services. We found disagreement regarding when and how health economists would recommend adoption of cost-effective but expensive interventions, which persisted across demographic and professional characteristics. First, when presented with a cost-effective but expensive intervention, researchers disagreed on whether they should characterize this intervention as worth funding. Overall, a substantial minority of researchers (37%) said that they would recommend funding Drug X in an academic article conducted for the US setting, rather than considering it unclear whether the drug should be funded or recommending against funding altogether. However, a similar proportion would recommend that policy makers find funding for Drug X in the budget, despite high cost, with remaining respondents recommended refusing to fund Drug X until prices fall or funding for a subset of the population.

In our stylized scenario, researchers lacked contextual information that would normally inform decisions, such as how effectively limited drug purchases could be targeted to sicker patients. Nevertheless, our survey suggests that the observed disagreement in recommendations reflects different underlying perspectives. We found that 42% of researchers believed that the WTP threshold should be derived from the budget, whereas 41% believed it should be derived from a societal measure of WTP. We also identified 4 distinct ways in which health economists traded off cost-effectiveness against affordability. In particular, 71% of hard ICER hawks recommended funding Drug X in an academic article, more than twice the overall average and 3.5 times the percentage of budget hawks or moderates. Still, few researchers believed that commonly used thresholds represent either the fixed budget or societal WTP threshold. Approximately half of respondents had chosen the threshold for their last article based solely on convention.

Given these disagreements, it may be prudent to avoid binary judgments of cost-effectiveness, especially in articles written in a US setting where there are multiple decision makers with distinct budgets with different levels of flexibility. Instead, researchers might state whether an intervention is high value for a specific payer's budget or only high value if a specific budget could be increased to fund expensive but cost-effective options. It may also be useful to be explicit about threshold choice and opportunity cost, delineating the extent to which a budget would have to increase to implement an intervention or which specific programs might be eliminated to afford it, and to consider multiple thresholds. In the absence of consensus, it may be useful to increase efforts to standardize a set of reported metrics across journals.

Current guidelines for CEA recommend a sensitivity analyses over different input parameters and the Second Panel on Cost-Effectiveness in Health and Medicine noted that a range of thresholds "may be appropriate, depending on how new technologies are funded."[21] We argue that future CEA reporting guidelines, such as the Consolidated Health Economic Evaluation Reporting Standards checklist, might also consider recommending the use of multiple types of cost-effectiveness thresholds or justifying threshold choice based on relevant context for the decision maker (such as whether or not the budget is fixed).[22]

In addition, authors should provide budget information necessary for implementation in CEA. As a ratio, an ICER is typically invariant to coverage and timing. In the Tufts Cost-Effectiveness Registry, which include all English-language CEAs, 93% of ICERs do not refer to a specific population size for the intervention.[23] Authors should estimate and report population-level health and economic outcomes, including budget impact, and be explicit about the timing of health and economic consequences. Distilling the time component of CEA to exponential discounting, as is current practice, may not adequately capture all policy-relevant time considerations, given unresolved debates around discounting.[24,25]

Some authors have proposed expansions to CEA theory to consider implementation factors such as equity, nonmarginal impacts, or how budget impacts manifest over time.[10,26] These allow CEA to more flexibly inform a decision problem. Nevertheless, they require more information than is traditionally required for CEA, may be difficult to estimate in many contexts, and increase the difficulty of comparing across different CEAs. This invites a larger future debate: whether it is more useful for cost-effectiveness to be a simple but incomplete measure presented alongside other measures, a view discussed by the Second Panel on Cost-Effectiveness in Health and Medicine,[19] or whether it should attempt to integrate additional relevant considerations like budget impact.[10]

There are several other limitations to our analysis. Our survey had limited details about the scenarios proposed and focused specifically on individuals with experience in health economics. Furthermore, because of our convenience sample, our results may have been affected by sampling bias, the results of which are difficult to quantify, and we had limited data and power to explore differences by demographic subgroups. Nevertheless, we obtained a diverse sample, and our findings indicate that experienced CEA researchers with varied demographic and professional characteristics may disagree on whether expensive interventions are worth funding. In an era of high-cost health technologies, US policy makers should move beyond simple CEA heuristics for interventions with large budgetary impacts, contextualizing cost-effectiveness within both current budgets and the potential for budget growth.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2021.11.1375.

Article and Author Information

Accepted for Publication: November 30, 2021

Published Online: February 24, 2022

doi: https://doi.org/10.1016/j.jval.2021.11.1375

Author Affiliations: Brown School of Public Health, Providence, RI, USA (Bilinski); Harvard Graduate School of Arts and Sciences, Cambridge, MA, USA (MacKay); Stanford University School of Medicine, Stanford, CA, USA (Salomon); Center for Health Decision Science, Harvard T.H. Chan School of Public Health, Boston, MA, USA (Pandya).

Correspondence: Alyssa Bilinski, PhD, Brown School of Public Health, 121 S Main St, Providence, RI 02903, USA. Email: alyssa_bilinski@brown.edu

Author Contributions: *Concept and design:* Bilinski, MacKay, Salomon, Pandya
Acquisition of data: Bilinski, MacKay
Analysis and interpretation of data: Bilinski, MacKay, Salomon, Pandya
Drafting of the manuscript: Bilinski, Pandya
Critical revision of the paper for important intellectual content: Bilinski, Salomon, Pandya
Statistical analysis: Bilinski, MacKay
Administrative, technical, or logistic support: Pandya
Supervision: Pandya, Salomon

Conflict of Interest Disclosures: Dr Bilinski reported receiving funding from a National Institutes of Health T32. No other conflicts of interests were reported. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funder.

Funding/Support: Financial support for this study was provided in part by a grant from the National Institutes of Health (Dr Bilinski, T32 AI007433).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors thank Alan Zaslavsky and Peter Neumann for their assistance with survey development. The authors also acknowledge Katherine McDaniel and Sam Doernberg for helpful comments.

Data Sharing: Data and code will be made publicly available on GitHub. Dr Bilinski and Evan MacKay had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

1. Chahal HS, Marseille EA, Tice JA, et al. Cost-effectiveness of early treatment of hepatitis C virus genotype 1 by stage of liver fibrosis in a US treatment-naive population. *JAMA Intern Med*. 2016;176(1):65–73.
2. Henry B. Drug pricing & challenges to hepatitis C treatment access. *J Health Biomed Law*. 2018;(14):265–283.
3. Sanger-Katz M. *Why the price of Sovaldi is a shock to the system*. The New York Times; August 6, 2014. https://www.nytimes.com/2014/08/07/upshot/why-the-price-of-sovaldi-is-a-shock-to-the-system.html. Accessed September 12, 2019.
4. Khazan O. *The true cost of an expensive medication*. The Atlantic; September 25, 2015. http://www.theatlantic.com/health/archive/2015/09/an-expensive-medications-human-cost/407299/. Accessed December 11, 2015.
5. Mera J, Reilley B, Leston J, Stephens D. In a critical state: ongoing barriers to treatment for hepatitis C virus (HCV). *Am J Med*. 2019;132(5):547–549.
6. Bilinski A, Neumann P, Cohen J, Thorat T, McDaniel K, Salomon JA. When cost-effective interventions are unaffordable: integrating cost-effectiveness and budget impact in priority setting for global health programs. *PLoS Med*. 2017;14(10):e1002397.
7. Lomas J, Claxton K, Martin S, Soares M. Resolving the "cost-effective but unaffordable" paradox: estimating the health opportunity costs of non-marginal budget impacts. *Value Health*. 2018;21(3):266–275.
8. Lakdawalla D, Neumann P. *Budget criteria and drug value assessments: a case of apples and oranges?* Health Affairs; September 22, 2016. http://www.healthaffairs.org/do/10.1377/hblog20160922.056735/full/. Accessed November 10, 2017.
9. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14), 1-503, v-vi.
10. Lomas JRS. Incorporating affordability concerns within cost-effectiveness analysis for health technology assessment. *Value Health*. 2019;22(8):898–905.
11. Paulden M. Recent amendments to NICE's value-based assessment of health technologies: implicitly inequitable? *Expert Rev Pharmacoecon Outcomes Res*. 2017;17(3):239–242.
12. Phelps CE. A new method to determine the optimal willingness to pay in cost-effectiveness analysis. *Value Health*. 2019;22(7):785–791.
13. Phelps CE. When opportunity knocks, what does it say? *Value Health*. 2019;22(8):851–853.
14. Topol EJ. *Malcolm Gladwell on fixing the US healthcare mess*. Medscape; July 14, 2015. http://www.medscape.com/viewarticle/847495. Accessed September 12, 2019.
15. Frank RG, Nichols LM. Medicare drug-price negotiation - why now… and how. *N Engl J Med*. 2019;381(15):1404–1406.
16. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness – the curious resilience of the $50,000-per-QALY threshold. *N Engl J Med*. 2014;371(9):796–797.
17. Vanness DJ, Lomas J, Ahn H. A health opportunity cost threshold for cost-effectiveness analysis in the United States. *Ann Intern Med*. 2021;174(1):25–32.
18. Chhatwal J, Kanwal F, Roberts MS, Dunn MA. Cost-effectiveness and budget impact of hepatitis C virus treatment with sofosbuvir and ledipasvir in the United States. *Ann Intern Med*. 2015;162(6):397–406.
19. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine [published correction appears in *JAMA*. 2016;316(18):1924]. *JAMA*. 2016;316(10):1093–1103.
20. Review. Evaluating the value of new drugs and devices. Institute for Clinical and Economic. https://icer-review.org/wp-content/uploads/2014/01/Slides-on-value-framework-for-national-webinar-new-slide-19.pdf; 2015. Accessed May 3, 2016.
21. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, eds. *Cost Effectiveness in Health and Medicine*. 2nd ed. Oxford, United Kingdom: Oxford University Press; 2017.
22. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ*. 2013;346:f1049.
23. CEA Registry. Center for the evaluation of value and risk in health. https://cevr.tuftsmedicalcenter.org/databases/cea-registry. Accessed August 2, 2019.
24. Brouwer WB, Niessen LW, Postma MJ, Rutten FF. Need for differential discounting of costs and health effects in cost effectiveness analyses. *BMJ*. 2005;331(7514):446–448.
25. Disability weights, discounting and age weighting of DALYs. World Health Organization. http://www.who.int/healthinfo/global_burden_disease/daly_disability_weight/en/. Accessed April 3, 2014.
26. Verguet S, Laxminarayan R, Jamison DT. Universal public finance of tuberculosis treatment in India: an extended cost-effectiveness analysis. *Health Econ*. 2015;24(3):318–332.

Costly Disagreements: A Survey of Researchers on Cost-Effectiveness and Affordability: Supplement 1

Alyssa Bilinski, Evan MacKay, Ankur Pandya

Contents

Sampling

To obtain our sample, we sent invitations to approximately 60 professors, PhD students, and industry contacts at 35 medium-to-large research institutions. Individuals were encouraged to pass the survey onto potential respondents. We also placed an invitation in the Society for Medical Decision-Making newsletter. Respondents were required to have taken at least one course in cost-effectiveness analysis (CEA) and/or contributed to at least one analysis.

Survey Participant Demographics

Table 1: Participant demographics

Current position		Number of people
	Government	13 (8%)
	Industry	10 (6%)
	Master's Student	39 (23%)
	PhD Student	23 (14%)
	Post-doc/Research scientist	31 (18%)
	Professor	43 (25%)
	Other	11 (6%)
Age (years)		
	<25	34 (20%)
	25-40	96 (56%)
	40-60	32 (19%)
	60+	8 (5%)
Gender identity		
	Female	72 (42%)
	Male	91 (54%)
	Non-binary	2 (1%)
	Prefer not to say	3 (2%)
	Prefer to self describe	1 (1%)
Years of CEA experience		
	<1	46 (27%)
	1-5	51 (30%)
	5-10	32 (19%)
	10+	41 (24%)
Number of cost-effectiveness analyses (approx)		
	0	38 (22%)
	1-2	46 (27%)
	3-5	24 (14%)
	5+	62 (36%)
Setting of last cost-effectiveness analysis		
	Low- and middle-income countries	29 (17%)
	High-income countries	94 (55%)
	A combination of both	8 (5%)
Regional focus		
	Low- and middle-income countries	23 (14%)
	High-income countries	100 (59%)
	A combination of both	43 (25%)

Demographic information about survey respondents (n = 170). Years of CEA experience refers to coursework, research, or teaching

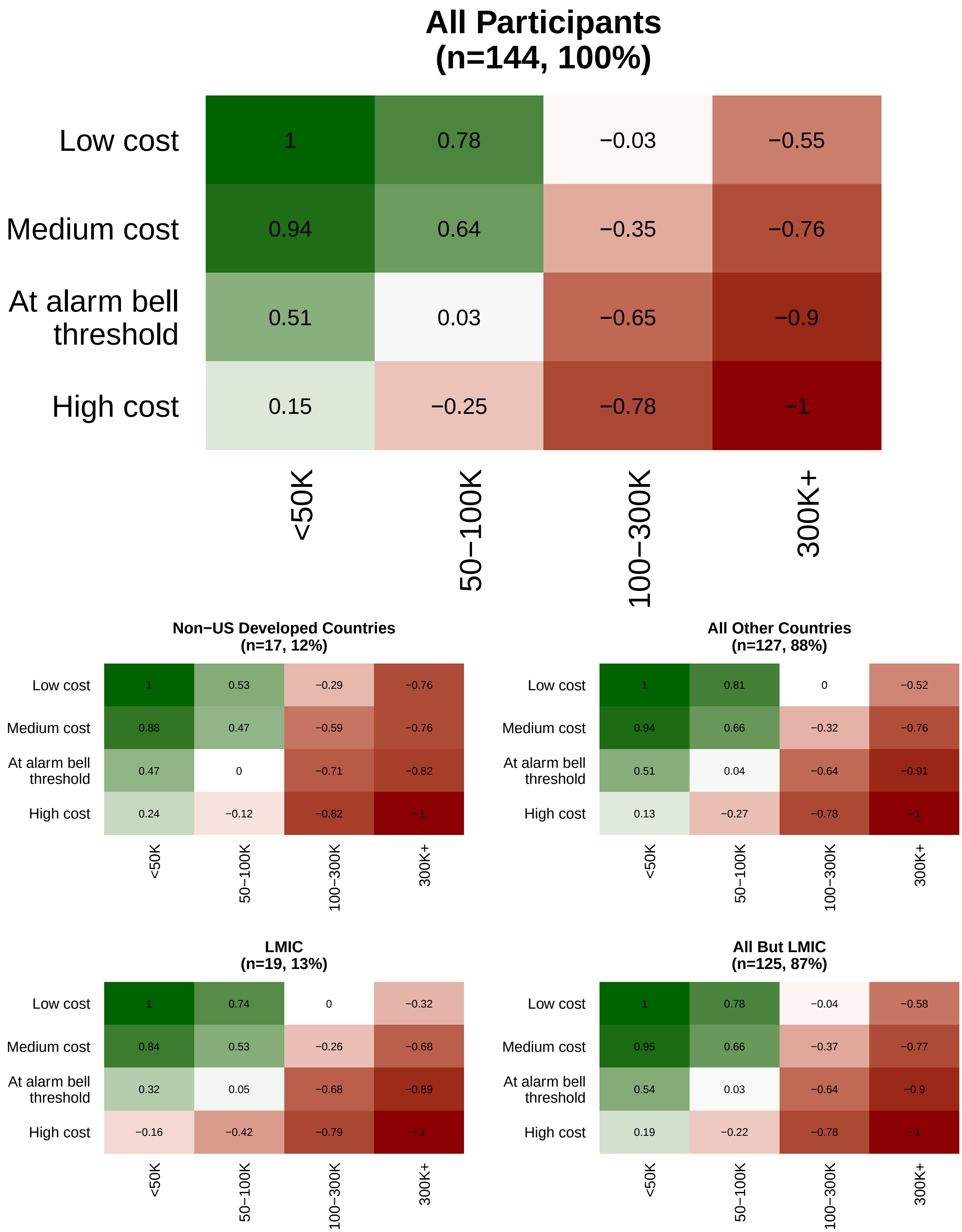
Subgroup heatmaps

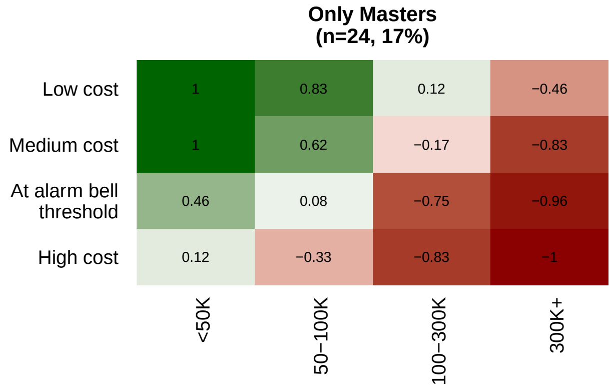
We show heatmaps (Figure 2) for different groups: all respondents as well as respondents 1) studying non-US developed countries, 2) studying low- and middle-income countries (LMICs), 3) excluding Master's students, and 4) who had conducted ≥ 1 cost-effectiveness analysis.

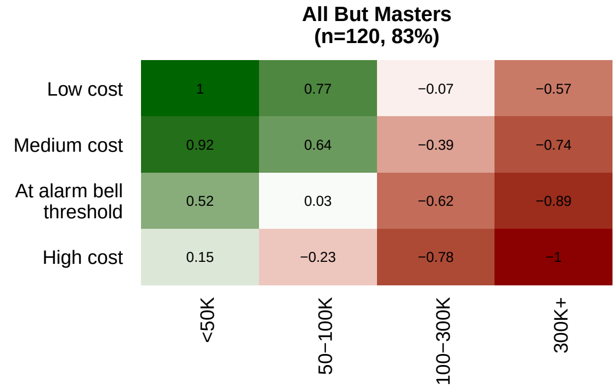
Respondents determined whether each square in the grid should be green (would recommend intervention), red (would not recommend intervention), or white (neutral) in the United States context. The x-axis includes cost-effectiveness ($/QALY), with larger values indicating less cost-effective services. The y-axis shows cost, representing total costs for the eligible patient population ("budget impact"). The alarm bell threshold is the "amount of net cost increase per individual new intervention that would contribute to growth in overall health care spending greater than the anticipated growth in national GDP + 1%."6 The color of the square is the average response over all participants with value responses to this question (n = 144, 85% of total). Non-US developed countries included respondents who noted being in such settings in the free response section, including the United Kingdom (n = 14) as well as Spain and Canada (n = 17), provided thresholds in pounds/Euros, or reported using the NICE threshold for cost-effectiveness. LMIC included respondents who reported low- and middle-income countries as their main focus in cost-effectiveness analysis (n = 23). All but Master's excludes Master's students, leaving n = 122. At least 1 CEA reference to people who self-reported having worked on more than 1 CEA or who had provided a threshold in their most recent CEA (n = 119).

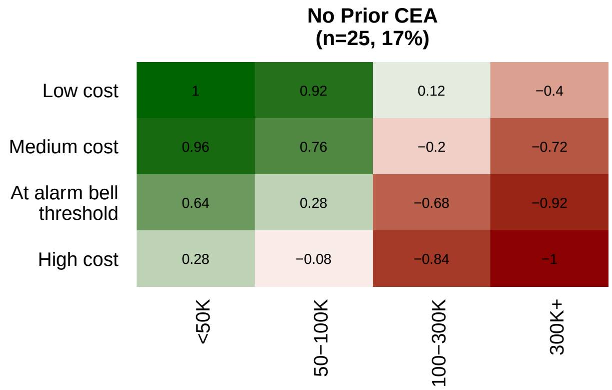
Trends were similar across professors, other academic positions, and non-academic positions. Professors were slightly less likely than other academic respondents to recommend high-cost/high value interventions (lower left) or low-cost/low-value interventions (upper right). By contrast, they were more likely than non-academic respondents to recommend both high-cost/high value and low cost/low value interventions.

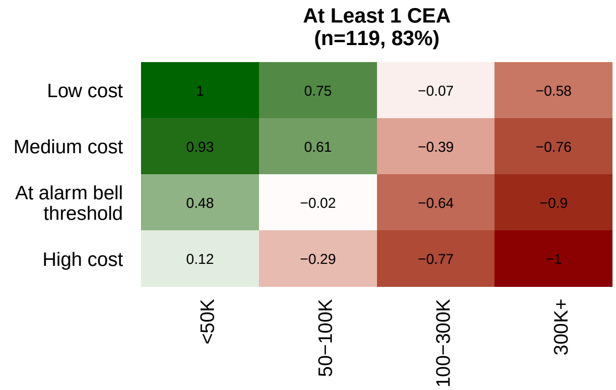
Figure S1. Tradeoffs between cost-effectiveness and affordability by subgroup.

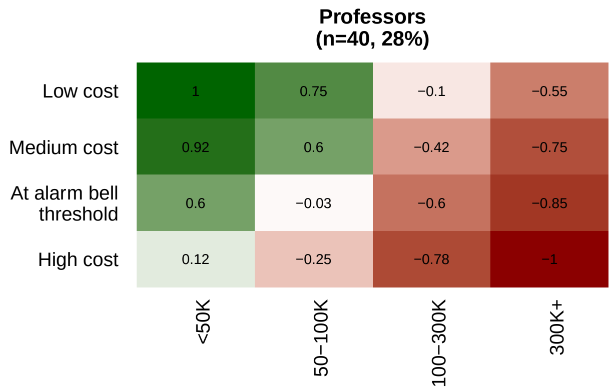


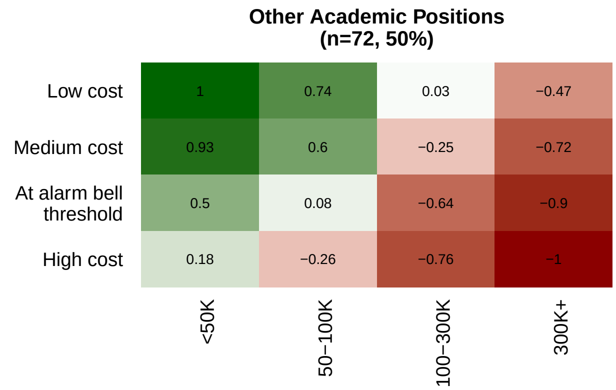

Only Masters
(n=24, 17%)
Low cost
1
0.83
0.12
−0.46
Medium cost
1
0.62
−0.17
−0.83
At alarm bell threshold
0.46
0.08
−0.75
−0.96
High cost
0.12
−0.33
−0.83
−1
<50K
50−100K
100−300K
300K+

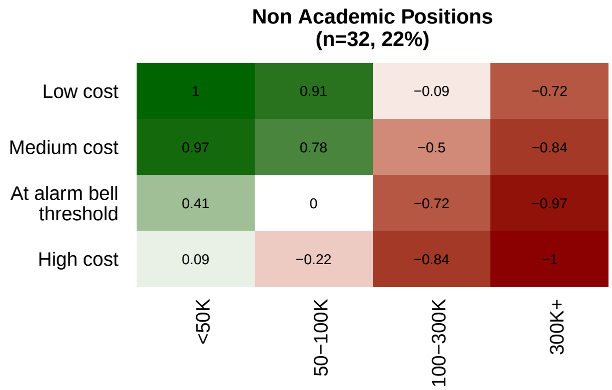

All But Masters
(n=120, 83%)
Low cost
1
0.77
−0.07
−0.57
Medium cost
0.92
0.64
−0.39
−0.74
At alarm bell threshold
0.52
0.03
−0.62
−0.89
High cost
0.15
−0.23
−0.78
−1
<50K
50−100K
100−300K
300K+


No Prior CEA
(n=25, 17%)
Low cost
1
0.92
0.12
−0.4
Medium cost
0.96
0.76
−0.2
−0.72
At alarm bell threshold
0.64
0.28
−0.68
−0.92
High cost
0.28
−0.08
−0.84
−1
<50K
50−100K
100−300K
300K+


At Least 1 CEA
(n=119, 83%)
Low cost
1
0.75
−0.07
−0.58
Medium cost
0.93
0.61
−0.39
−0.76
At alarm bell threshold
0.48
−0.02
−0.64
−0.9
High cost
0.12
−0.29
−0.77
−1
<50K
50−100K
100−300K
300K+


Professors
(n=40, 28%)
Low cost
1
0.75
−0.1
−0.55
Medium cost
0.92
0.6
−0.42
−0.75
At alarm bell threshold
0.6
−0.03
−0.6
−0.85
High cost
0.12
−0.25
−0.78
−1
<50K
50−100K
100−300K
300K+


Other Academic Positions
(n=72, 50%)
Low cost
1
0.74
0.03
−0.47
Medium cost
0.93
0.6
−0.25
−0.72
At alarm bell threshold
0.5
0.08
−0.64
−0.9
High cost
0.18
−0.26
−0.76
−1
<50K
50−100K
100−300K
300K+


Non Academic Positions
(n=32, 22%)
Low cost
1
0.91
−0.09
−0.72
Medium cost
0.97
0.78
−0.5
−0.84
At alarm bell threshold
0.41
0
−0.72
−0.97
High cost
0.09
−0.22
−0.84
−1
<50K
50−100K
100−300K
300K+

Enhanced heatmaps

These heatmaps show the breakdown of individuals that selected each option for the subgroups listed above.

Respondents determined whether each square in the grid should be green (would recommend intervention), red (would not recommend intervention), or white (neutral) in the United States context. The x-axis includes cost-effectiveness ($/QALY), with larger values indicating less cost-effective services. The y-axis shows cost, representing total costs for the eligible patient population ("budget impact"). The alarm bell threshold is the "amount of net cost increase per individual new intervention that would contribute to growth in overall health care spending greater than the anticipated growth in national GDP + 1%."6 The colors in each square represent the breakdown of participant responses for each square (n = 144, 85% of total). Non-US developed countries included respondents who noted being in such settings in the free response section, including the United Kingdom (n = 14) as well as Spain and Canada (n = 17), provided thresholds in pounds/Euros, or reported using the NICE threshold for cost-effectiveness. LMIC included respondents who reported low- and middle-income countries as their main focus in cost-effectiveness analysis (n = 23). All but Master's excludes Master's students, leaving n = 122. At least 1 CEA reference to people who self-reported having worked on more than 1 CEA or who had provided a threshold in their most recent CEA (n = 119).

Figure S2-A

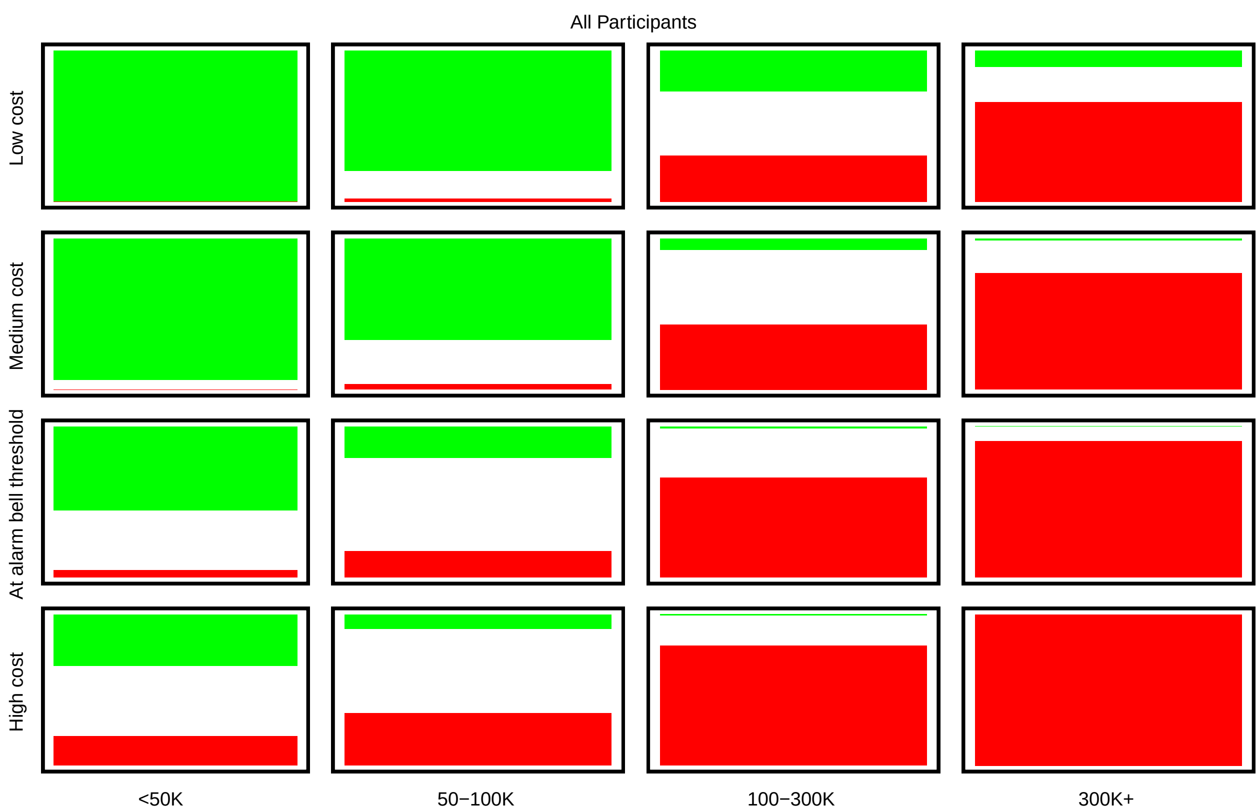


Figure S2-B

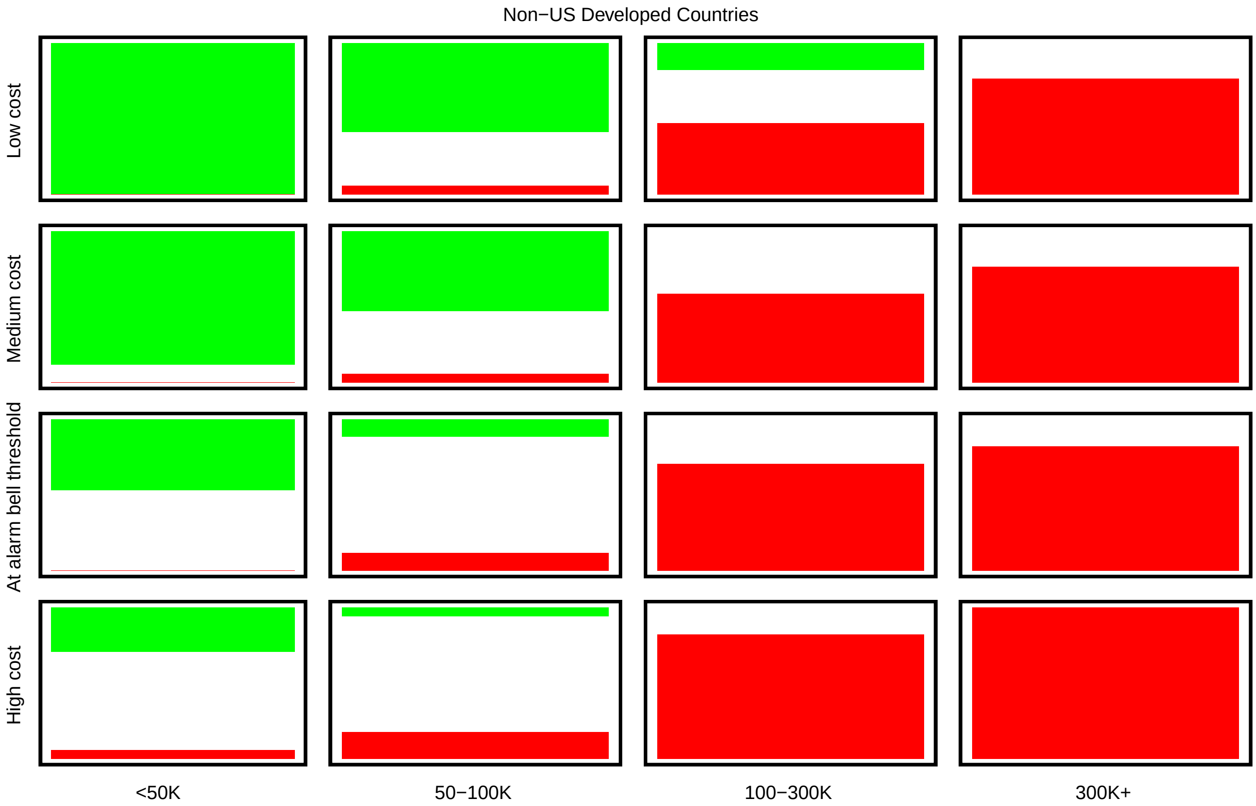

Non−US Developed Countries
Low cost
Medium cost
At alarm bell threshold
High cost
<50K
50−100K
100−300K
300K+

Figure S2-C

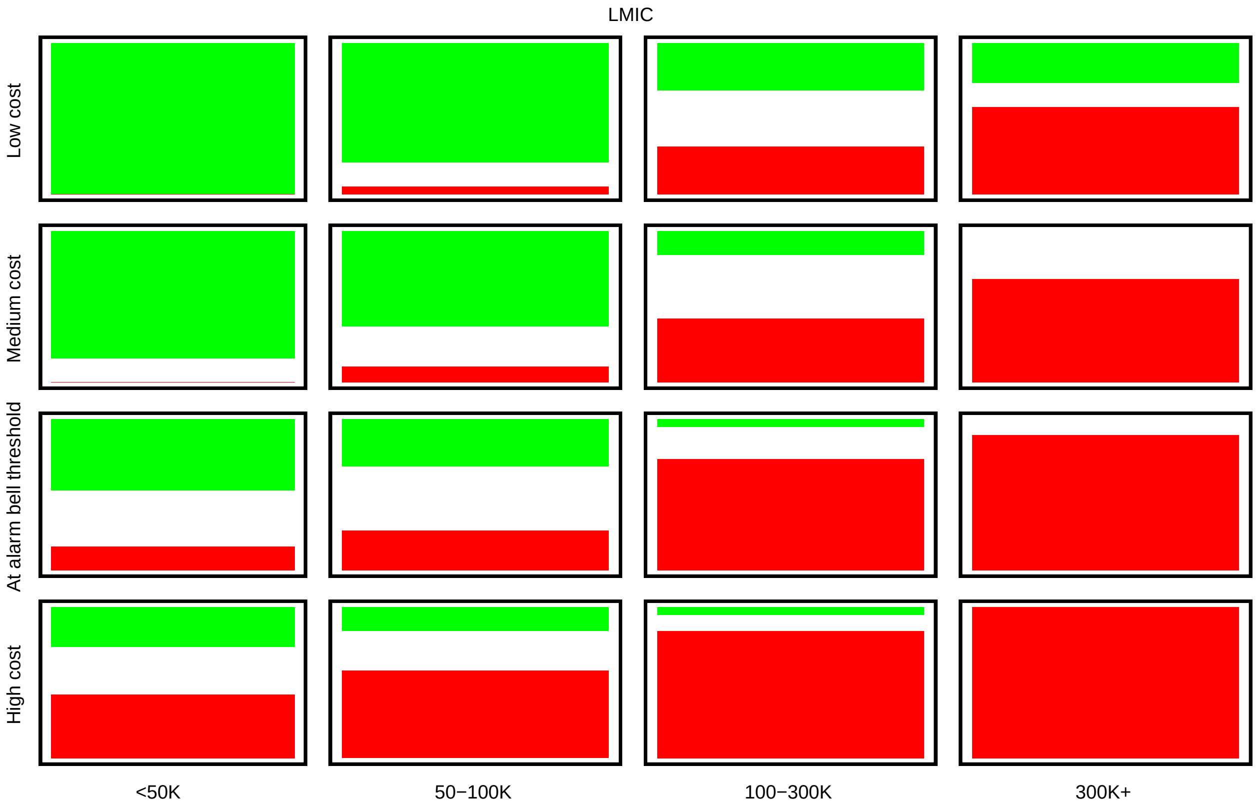

LMIC
Low cost
Medium cost
At alarm bell threshold
High cost
<50K
50−100K
100−300K
300K+

Figure S2-D

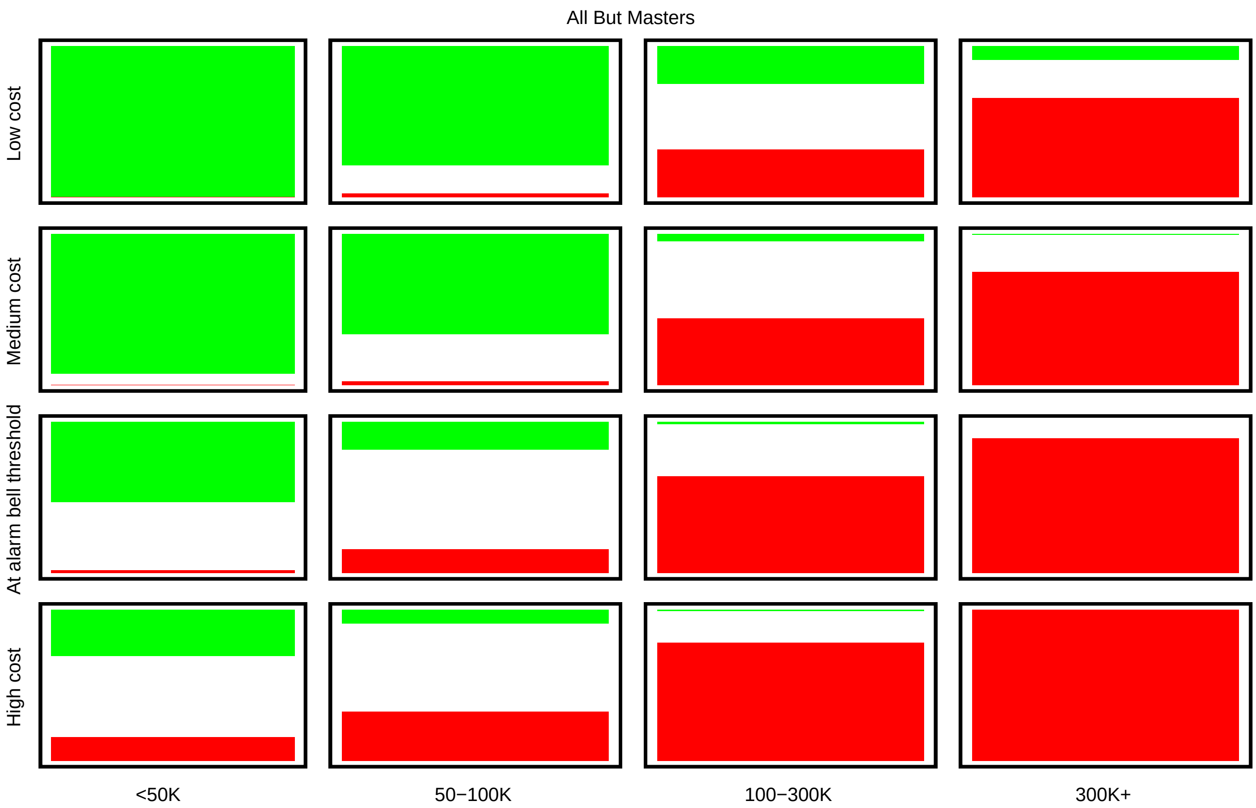
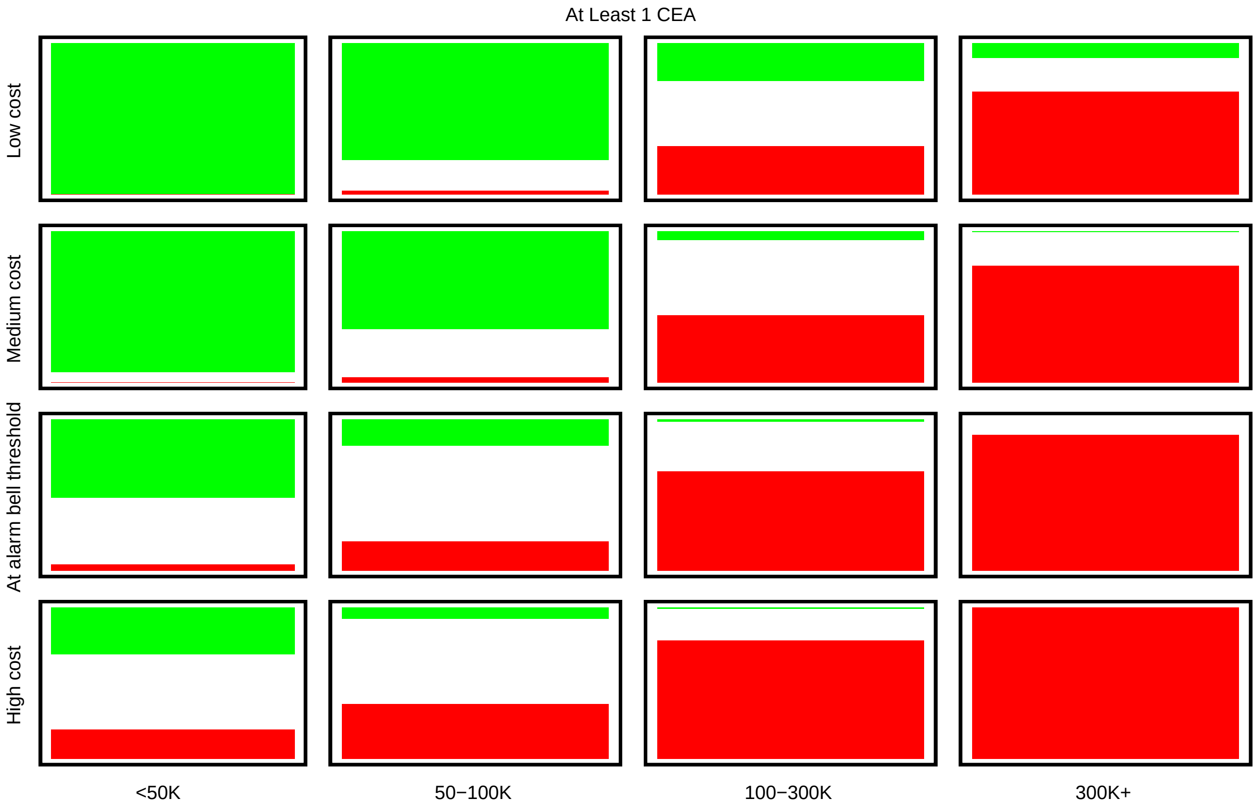

All But Masters
Low cost
Medium cost
At alarm bell threshold
High cost
<50K
50−100K
100−300K
300K+

Figure S2-E


At Least 1 CEA
Low cost
Medium cost
At alarm bell threshold
High cost
<50K
50−100K
100−300K
300K+

Summaries of Free Response Questions

Our example scenarios attempted to provide as options high-level recommendations commonly provided in applied papers. We allowed authors to select "Other" as a response to this question and fill in alternative recommendations. Below are summaries to the fill-in responses.

Question: Which of the following best explains your response in the previous question?
A) Drug X is cost-effective, but policymakers would have to to reallocate funding from existing programs to pay for the drug.
B) Drug X is cost-effective, but policymakers should wait until more competitors or generic options enter the market.
C) Other

Results for "Other" are summarized below.

Table 2:

Cost-effectiveness is incomplete information	4
Increase budget or reallocate (add the reallocation to main category)	1
Need special decision process given high budget impact	10
Negotiate prices or wait until generics (just add the waiting to the main categoriy)	1

Question: You are asked to advise the Massachusetts health commision on whether and how to adopt Drug X in its Medicaid budget. Which of the following best summarizes how you would advise them?
A) Fund Drug X for all eligible patients.
B) Refuse to fund Drug X at its current price. Wait for a lower price or competitors.
C) Only fund Drug X for half of the eligible population (chosen at random) to reduce budget impact concerns.
D) Other

Results for "Other" are summarized below.

Table 3:

Identify subpopulations that would benefit most	23
Adjust budget	2
Negotiate price/Novel payment mechanisms	6
Use shopping spree on Medicaid budget	7
Other	6

Survey Outcomes by Demographic Subgroups

In Figure 2 below, we present main survey outcomes by demographic subgroups. After adjusting the significance threshold (Bonferroni corrected threshold: 0.05/32=0.0016) for multiple testing, we did not identify significant differences by demographic characteristics. (The smallest p-values observed were across academic positions in terms of their Evaluation of Drug X ($p = 0.052$) and across years of experience in terms of whether the willingness-to-pay threshold should reflect budget or societal willingness-to-pay ($p = 0.03$), both of which may merit further exploration in a higher-powered sample.) For reference, the category label was significant in the evaluation of Drug X ($p < 0.00001$) and whether to Adopt Drug X in the Medicaid Budget ($p = 0.0008$).

Figure 2, Part 1

			Evaluation of Drug X % (N)			Adopt Drug X in Medicaid Budget			
		Total	Should fund	Should not fund	Unclear	Fund for all	Fund for some	Do not fund	Other
Overall			37% (63)	6% (10)	57% (97)	34% (58)	12% (21)	25% (43)	28% (48)
Category	Hard ICER Hawk	26% (38)	71% (27)	3% (1)	26% (10)	61% (23)	8% (3)	13% (5)	18% (7)
	Soft ICER Hawk	33% (47)	15% (7)	0% (0)	85% (40)	19% (9)	13% (6)	34% (16)	34% (16)
	Budget Hawk	23% (33)	18% (6)	6% (2)	76% (25)	15% (5)	15% (5)	27% (9)	42% (14)
	Moderate	18% (26)	38% (10)	4% (1)	58% (15)	35% (9)	8% (2)	23% (6)	35% (9)
Country Focus	LMIC	26% (23)	30% (7)	4% (1)	65% (15)	35% (8)	9% (2)	26% (6)	30% (7)
	A combination of both	48% (43)	44% (19)	9% (4)	47% (20)	35% (15)	12% (5)	28% (12)	26% (11)
	High-income countries	26% (23)	30% (7)	4% (1)	65% (15)	35% (8)	9% (2)	26% (6)	30% (7)
Position	Academic	80% (136)	38% (51)	6% (8)	57% (77)	35% (47)	14% (19)	25% (34)	26% (36)
	Government	8% (13)	46% (6)	8% (1)	46% (6)	31% (4)	8% (1)	23% (3)	38% (5)
	Industry	6% (10)	20% (2)	10% (1)	70% (7)	30% (3)	10% (1)	30% (3)	30% (3)
	Other	6% (11)	36% (4)	0% (0)	64% (7)	36% (4)	0% (0)	27% (3)	36% (4)
Within academia	Professors	32% (43)	37% (16)	0% (0)	63% (27)	33% (14)	2% (1)	23% (10)	42% (18)
	Masters students	29% (39)	54% (21)	13% (5)	33% (13)	46% (18)	28% (11)	23% (9)	3% (1)
	PhD Students	17% (23)	30% (7)	4% (1)	65% (15)	22% (5)	26% (6)	35% (8)	17% (4)
	Post-docs / Research scientists	23% (31)	23% (7)	6% (2)	71% (22)	32% (10)	3% (1)	23% (7)	42% (13)
Gender Identity	Female	42% (72)	40% (29)	8% (6)	51% (37)	38% (27)	15% (11)	24% (17)	24% (17)
	Male and all others	58% (98)	35% (34)	4% (4)	61% (60)	32% (31)	10% (10)	27% (26)	32% (31)
Years Experience	<1	27% (46)	50% (23)	11% (5)	39% (18)	41% (19)	28% (13)	28% (13)	2% (1)
	1-5	30% (51)	31% (16)	4% (2)	65% (33)	41% (21)	10% (5)	25% (13)	24% (12)
	6+	43% (73)	33% (24)	4% (3)	63% (46)	25% (18)	4% (3)	23% (17)	48% (35)
Age	<25	20% (34)	59% (20)	12% (4)	29% (10)	41% (14)	29% (10)	26% (9)	3% (1)
	25-40	56% (96)	29% (28)	5% (5)	66% (63)	31% (30)	10% (10)	27% (26)	31% (30)
	40-60	19% (32)	38% (12)	3% (1)	59% (19)	31% (10)	3% (1)	25% (8)	41% (13)
	60+	5% (8)	38% (3)	0% (0)	62% (5)	50% (4)	0% (0)	0% (0)	50% (4)
# of CEAs Completed	1+	78% (132)	34% (45)	4% (5)	62% (82)	32% (42)	7% (9)	27% (36)	34% (45)
	0	22% (38)	47% (18)	13% (5)	39% (15)	42% (16)	32% (12)	18% (7)	8% (3)

Figure 2, Part 2

			WTP threshold should reflect:			If Drug X is CE but not affordable, we should:		
		Total	Budget WTP	Societal WTP	Ambivalent	Increase budget	Decrease WTP	Felt Nearly Equal
Overall			42% (72)	41% (70)	16% (28)	38% (64)	37% (63)	25% (43)
Category	Hard ICER Hawk	26% (38)	26% (10)	55% (21)	18% (7)	32% (12)	39% (15)	29% (11)
	Soft ICER Hawk	33% (47)	40% (19)	36% (17)	23% (11)	32% (15)	38% (18)	30% (14)
	Budget Hawk	23% (33)	42% (14)	42% (14)	15% (5)	45% (15)	27% (9)	27% (9)
	Moderate	18% (26)	58% (15)	31% (8)	12% (3)	42% (11)	35% (9)	23% (6)
Country Focus	LMIC	26% (23)	35% (8)	39% (9)	26% (6)	43% (10)	35% (8)	22% (5)
	A combination of both	48% (43)	51% (22)	40% (17)	9% (4)	23% (10)	58% (25)	19% (8)
	High-income countries	26% (23)	35% (8)	39% (9)	26% (6)	43% (10)	35% (8)	22% (5)
Position	Academic	80% (136)	43% (59)	38% (52)	18% (25)	36% (49)	37% (50)	27% (37)
	Government	8% (13)	46% (6)	38% (5)	15% (2)	46% (6)	54% (7)	0% (0)
	Industry	6% (10)	30% (3)	60% (6)	10% (1)	40% (4)	30% (3)	30% (3)
	Other	6% (11)	36% (4)	64% (7)	0% (0)	45% (5)	27% (3)	27% (3)
Within academia	Professors	32% (43)	56% (24)	26% (11)	19% (8)	28% (12)	37% (16)	35% (15)
	Masters students	29% (39)	44% (17)	44% (17)	13% (5)	38% (15)	41% (16)	21% (8)
	PhD Students	17% (23)	39% (9)	43% (10)	17% (4)	35% (8)	35% (8)	30% (7)
	Post-docs / Research scientists	23% (31)	29% (9)	45% (14)	26% (8)	45% (14)	32% (10)	23% (7)
Gender Identity	Female	42% (72)	49% (35)	35% (25)	17% (12)	32% (23)	42% (30)	26% (19)
	Male and all others	58% (98)	38% (37)	46% (45)	16% (16)	42% (41)	34% (33)	24% (24)
Years Experience	<1	27% (46)	46% (21)	48% (22)	7% (3)	39% (18)	35% (16)	26% (12)
	1-5	30% (51)	31% (16)	39% (20)	29% (15)	45% (23)	27% (14)	27% (14)
	6+	43% (73)	48% (35)	38% (28)	14% (10)	32% (23)	45% (33)	23% (17)
Age	<25	20% (34)	50% (17)	44% (15)	6% (2)	38% (13)	44% (15)	18% (6)
	25-40	56% (96)	38% (36)	45% (43)	18% (17)	41% (39)	31% (30)	28% (27)
	40-60	19% (32)	53% (17)	25% (8)	22% (7)	28% (9)	47% (15)	25% (8)
	60+	5% (8)	25% (2)	50% (4)	25% (2)	38% (3)	38% (3)	25% (2)
# of CEAs Completed	1+	78% (132)	43% (57)	39% (51)	18% (24)	39% (51)	35% (46)	27% (35)
	0	22% (38)	39% (15)	50% (19)	11% (4)	34% (13)	45% (17)	21% (8)

Logistic Regression Analysis

We performed regression analysis on two questions about survey participants' recommendations. The first used an outcome variable of whether a researcher would recommend **in an academic paper** that decision-makers fund drug X at current prices, where drug X is a hypothetical cost-effective medication with an ICER of $40,000/QALY but would cost 20% of the Medicaid budget over the next 5 years. The second had outcome variable is whether a researcher would recommend **to a policymaker** to fund drug X at current prices, where drug X is a hypothetical cost-effective medication with an ICER of $40,000/QALY but would cost 20% of the Medicaid budget over the next 5 years. Both used demographic variables as covariates, including current position, whether the respondent was a Master's student, age, gender, years of CEA experience, regional focus, and number of CEAs performed. We found some evidence that Master's students were more likely to advise funding an expensive but cost-effective intervention in an academic paper ($p<0.05$, adjusted and unadjusted), but little variation in other variables. Master's students were not any more or less likely to advise funding an expensive but cost-effective intervention to a policymaker.

Table 4: Logistic regression of whether to recommend drug X in an academic paper vs demographic characteristics

		Unadjusted ORs (95% CI)	Adjusted ORs (95% CI)
Current position			
	Academia	1.00 (Ref)	1.00 (Ref)
	Government	0.4 (-0.8, 1.5)	0.8 (-0.4, 2.1)
	Industry	-0.9 (-2.5, 0.7)	-0.5 (-2.2, 1.1)
	Other	0 (-1.3, 1.2)	0.3 (-1.1, 1.6)
Master's			
	No	1.00 (Ref)	1.00 (Ref)
	Yes	0.9 (0.2, 1.6)*	1.1 (0, 2.1)*
Age (years)			
	<40	1.00 (Ref)	1.00 (Ref)
	40 +	0 (-0.7, 0.8)	0.5 (-0.5, 1.4)
Gender identity			
	Female	1.00 (Ref)	1.00 (Ref)
	Male or other	-0.2 (-0.9, 0.4)	-0.2 (-0.9, 0.5)
Years of CEA experience			
	<1	1.00 (Ref)	1.00 (Ref)
	1-5	-0.8 (-1.6, 0)	-0.2 (-1.3, 0.9)
	6 +	-0.7 (-1.5, 0)	-0.3 (-1.5, 1)
Number of CEAs (approx)			
	0	1.00 (Ref)	1.00 (Ref)
	1+	-0.6 (-1.3, 0.2)	0.1 (-0.9, 1.2)
Regional focus			
	Low- and middle-income countries	1.00 (Ref)	1.00 (Ref)
	High-income countries	0.6 (-0.5, 1.7)	0.4 (-0.7, 1.6)
	A combination of both	0.2 (-0.8, 1.2)	0.2 (-0.8, 1.2)

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

Logistic regression (n = 170). The outcome variable is whether a researcher would recommend **in an academic paper** that decision-makers fund drug X at current prices, where drug X is a hypothetical cost-effective medication with an ICER of $40,000/QALY but would cost 20% of the Medicaid budget over the next 5 years. Higher OR estimates indicate more favorable options toward drug X.

Adjusted ORs are adjusted for all other variables in the table.

Years of CEA experience refers to coursework, research, or teaching

Master's = 'Yes' indicates that the respondent reported currently being a Master's student (n = 39).

Gender identity 'Other' refers to participants who self reported 'Male' (n = 91), 'Non-binary' (n = 2), 'Prefer not to say' (n = 3), or 'Prefer to self describe' (n = 1).

Abbreviations: Cost-effectiveness analysis (CEA)

Table 5: Logistic regression of whether to recommend drug X to a policymaker vs demographic characteristics

		Unadjusted ORs (95% CI)	Adjusted ORs (95% CI)
Current position	Academia	1.00 (Ref)	1.00 (Ref)
	Government	-0.2 (-1.4, 1.1)	0.3 (-1.1, 1.6)
	Industry	-0.2 (-1.6, 1.2)	0 (-1.5, 1.5)
	Other	0.1 (-1.2, 1.4)	0.2 (-1.2, 1.5)
Master's			
	No	1.00 (Ref)	1.00 (Ref)
	Yes	0.7 (-0.1, 1.4)	0.9 (-0.1, 1.9)
Age (years)			
	<40	1.00 (Ref)	1.00 (Ref)
	40 +	0.1 (-0.7, 0.8)	0.9 (-0.1, 1.9)
Gender identity			
	Female	1.00 (Ref)	1.00 (Ref)
	Male or other	-0.3 (-0.9, 0.4)	-0.2 (-0.9, 0.5)
Years of CEA experience			
	<1	1.00 (Ref)	1.00 (Ref)
	1-5	0 (-0.8, 0.8)	0.5 (-0.6, 1.6)
	6 +	-0.8 (-1.6, 0)	-0.6 (-1.9, 0.7)
Number of CEAs (approx)			
	0	1.00 (Ref)	1.00 (Ref)
	1+	-0.4 (-1.2, 0.3)	0 (-1.1, 1.1)
Regional focus			
	Low- and middle-income countries	1.00 (Ref)	1.00 (Ref)
	High-income countries	0 (-1.1, 1.1)	-0.1 (-1.2, 1.1)
	A combination of both	-0.1 (-1, 0.9)	-0.1 (-1.1, 0.9)

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

Logistic regression (n = 170). The outcome variable is whether a researcher would recommend **to a policymaker** that decision-makers fund drug X at current prices, where drug X is a hypothetical cost-effective medication with an ICER of $40,000/QALY but would cost 20% of the Medicaid budget over the next 5 years. Higher OR estimates indicate more favorable options toward drug X.

Adjusted ORs are adjusted for all other variables in the table.

Years of CEA experience refers to coursework, research, or teaching

Master's = 'Yes' indicates that the respondent reported currently being a Master's student (n = 39).

Gender identity 'Other' refers to participants who self reported 'Male' (n = 91), 'Non-binary' (n = 2), 'Prefer not to say' (n = 3), or 'Prefer to self describe' (n = 1).

Abbreviations: Cost-effectiveness analysis (CEA)

Linear Regression Analysis

We considered a linear probability model to assess sensitivity to functional form assumptions in logistic regression. Both regressions have similar results.

Table 6: Linear regression of whether to recommend drug X in an academic paper vs demographic characteristics

Current position		Unadjusted Coefficients (95% CI)	Adjusted Coefficients (95% CI)
	Academia	0.00 (Ref)	0.00 (Ref)
	Government	0.1 (-0.2, 0.4)	0.2 (-0.1, 0.5)
	Industry	-0.2 (-0.5, 0.1)	-0.1 (-0.4, 0.2)
	Other	0 (-0.3, 0.3)	0.1 (-0.2, 0.4)
Master's			
	No	0.00 (Ref)	0.00 (Ref)
	Yes	0.2 (0, 0.4)*	0.2 (0, 0.5)*
Age (years)			
	<40	0.00 (Ref)	0.00 (Ref)
	40 +	0 (-0.2, 0.2)	0.1 (-0.1, 0.3)
Gender identity			
	Female	0.00 (Ref)	0.00 (Ref)
	Male or other	-0.1 (-0.2, 0.1)	0 (-0.2, 0.1)
Years of CEA experience			
	<1	0.00 (Ref)	0.00 (Ref)
	1-5	-0.2 (-0.4, 0)	-0.1 (-0.3, 0.2)
	6 +	-0.2 (-0.3, 0)	-0.1 (-0.3, 0.2)
Number of CEAs (approx)			
	0	0.00 (Ref)	0.00 (Ref)
	1+	-0.1 (-0.3, 0)	0 (-0.2, 0.3)
Regional focus			
	Low- and middle-income countries	0.00 (Ref)	0.00 (Ref)
	High-income countries	0.1 (-0.1, 0.4)	0.1 (-0.2, 0.3)
	A combination of both	0 (-0.2, 0.3)	0 (-0.2, 0.3)

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

Linear regression (n = 170). The outcome variable is whether a researcher would recommend **in an academic paper** that decision-makers fund drug X at current prices, where drug X is a hypothetical cost-effective medication with an ICER of $40,000/QALY but would cost 20% of the Medicaid budget over the next 5 years. Higher coefficient estimates indicate more favorable options toward drug X.

Adjusted coefficients are adjusted for all other variables in the table.

Years of CEA experience refers to coursework, research, or teaching

Master's = 'Yes' indicates that the respondent reported currently being a Master's student (n = 39).

Gender identity 'Other' refers to participants who self reported 'Male' (n = 91), 'Non-binary' (n = 2), 'Prefer not to say' (n = 3), or 'Prefer to self describe' (n = 1).

Abbreviations: Cost-effectiveness analysis (CEA)

Table 7: Linear regression of whether to recommend drug X to a policymaker vs demographic characteristics

		Unadjusted Coefficients (95% CI)	Adjusted Coefficients (95% CI)
Current position			
	Academia	0.00 (Ref)	0.00 (Ref)
	Government	0 (-0.3, 0.2)	0.1 (-0.2, 0.3)
	Industry	0 (-0.4, 0.3)	0 (-0.3, 0.3)
	Other	0 (-0.3, 0.3)	0 (-0.3, 0.3)
Master's			
	No	0.00 (Ref)	0.00 (Ref)
	Yes	0.2 (0, 0.3)	0.2 (0, 0.4)
Age (years)			
	<40	0.00 (Ref)	0.00 (Ref)
	40 +	0 (-0.2, 0.2)	0.2 (0, 0.4)
Gender identity			
	Female	0.00 (Ref)	0.00 (Ref)
	Male or other	-0.1 (-0.2, 0.1)	0 (-0.2, 0.1)
Years of CEA experience			
	<1	0.00 (Ref)	0.00 (Ref)
	1-5	0 (-0.2, 0.2)	0.1 (-0.1, 0.4)
	6 +	-0.2 (-0.3, 0)	-0.1 (-0.4, 0.2)
Number of CEAs (approx)			
	0	0.00 (Ref)	0.00 (Ref)
	1+	-0.1 (-0.3, 0.1)	0 (-0.2, 0.2)
Regional focus			
	Low- and middle-income countries	0.00 (Ref)	0.00 (Ref)
	High-income countries	0 (-0.2, 0.2)	0 (-0.3, 0.2)
	A combination of both	0 (-0.2, 0.2)	0 (-0.2, 0.2)

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

Linear regression (n = 170). The outcome variable is whether a researcher would recommend **to a policymaker** that decision-makers fund drug X at current prices, where drug X is a hypothetical cost-effective medication with an ICER of $40,000/QALY but would cost 20% of the Medicaid budget over the next 5 years. Higher coefficient estimates indicate more favorable options toward drug X.

Adjusted coefficients are adjusted for all other variables in the table.

Years of CEA experience refers to coursework, research, or teaching

Master's = 'Yes' indicates that the respondent reported currently being a Master's student (n = 39).

Gender identity 'Other' refers to participants who self reported 'Male' (n = 91), 'Non-binary' (n = 2), 'Prefer not to say' (n = 3), or 'Prefer to self describe' (n = 1).

Abbreviations: Cost-effectiveness analysis (CEA)

Cost-effectiveness survey - v2

Directions:

You are being asked to take part in a research study. This research is being conducted to learn about best practices for cost-effectiveness analysis. Specifically, we are interested in learning about how researchers would recommend using cost-effectiveness analysis in decision-making. You are being asked to participate in this research because you have completed at least one course about cost-effectiveness analysis and/or written in at least one cost-effectiveness analysis.

Your participation in this study is voluntary and you may withdraw your participation at any time for any reason. If you take part in this study, you will be asked to complete a survey, which will take about 5-7 minutes. Specifically, you will be asked a number of questions about whether you would implement certain programs and how you conduct and interpret cost-effectiveness analysis. You can terminate your participation at any time. If you have any questions about this study, you can contact Alyssa Bilinski (abilinski@g.harvard.edu).

At the end of the survey, you will be given an opportunity to enter your email for a chance to win a $25 gift card; we will give out 5 gift cards in total. Thank you again for your time and participation.

To participate in the survey, click “I consent to participate in this survey.” To decline, close this tab.

Q1 You are a researcher conducting a cost-effectiveness analysis on a new drug, Drug X in Massachusetts. You find that the drug has an incremental cost-effectiveness ratio of $40,000/QALY (compared to a typical threshold of $50,000-$100,000 QALY) over a lifetime time horizon. To provide the drug to the entire eligible population would cost about 20% of the current Medicaid budget over the next 3 years. Which of the following best summarizes how you write up this result in the discussion section in an academic paper?

- ◯ In general, decision-makers should fund Drug X at current prices. (1)
- ◯ In general, decision-makers should not fund Drug X at current prices. (2)
- ◯ It's unclear whether the decision-maker should fund Drug X current prices. (3)

Display This Question:

If You are a researcher conducting a cost-effectiveness analysis on a new drug, Drug X in Massachuse... = In general, decision-makers should not fund Drug X at current prices.

Or You are a researcher conducting a cost-effectiveness analysis on a new drug, Drug X in Massachuse... = It's unclear whether the decision-maker should fund Drug X current prices.

Q2 Which of the following best explains your response in the previous question?

◯ Drug X is cost-effective, but policymakers would have to to reallocate funding from existing programs to pay for the drug. (3)

◯ Drug X is cost-effective, but policymakers should wait until more competitors or generic options enter the market. (4)

◯ Other (5) ____

Page Break

Q3 You are asked to advise the Massachusetts health commission on whether and how to adopt Drug X in its Medicaid budget. Which of the following best summarizes how you would advise them?

○ Recommend funding Drug X for all eligible patients (1)

○ Refuse to fund Drug X at its current price. Wait for a lower price or competitors. (2)

○ Only fund Drug X for half of the eligible population (chosen at random) to reduce budget impact concerns (3)

○ Other (4) ____

Q4 Each box in the grid represents an intervention, with its ICER across the top and total cost across the left side. For example, the top left box represents an intervention of low cost with an ICER of less than $50K/year.

Click the boxes to indicate an opinion on these interventions in a United States setting:

- One click (green): decision-maker should fund intervention
- Two clicks (red): decision-maker should not fund the intervention
- No click (white): unclear whether the decision-maker should fund the intervention

ICERs are given in units of $/QALY. Costs represent total costs for the eligible patient population ("budget impact"). The alarm bell threshold is the “amount of net cost increase per individual new intervention that would contribute to growth in overall health care spending greater than the anticipated growth in national GDP + 1%” (Institute for Clinical and Economic Review).

	< 50K	50-100K	100-300K	300K +
Low cost				
Medium cost				
At alarm bell threshold				
High cost				

Q5

	In the academic literature, the willingness to pay threshold should reflect the current budget, i.e. be explicitly derived from funding interventions in order of cost-effectiveness until budget is exhausted. (1)	In the academic literature, the willingness to pay threshold should reflect societal willingness to pay, even if this would result in substantial increases or decreases to the current health budget (2)
I more strongly agree with the following: (1)	○	○

Q6 Rate the strength of your agreement on the previous question.

- ○ Completely agree with the option I chose (1)
- ○ Felt conflicted but had a clear preference (2)
- ○ Felt nearly equal about both answer choices (3)

Q7

	In most cases, if a program is cost-effective but not affordable, this typically means that policymakers should increase the budget. (1)	If a program is cost-effective but not affordable, this typically means that policymakers should decrease the willingness to pay threshold. (2)
I more strongly agree with the following: (1)	○	○

Q8 Rate the strength of your agreement on the previous question.

- ◯ Completely agree with the option I chose (1)
- ◯ Felt conflicted but had a clear preference (2)
- ◯ Felt nearly equal about both answer choices (3)

Q9 In the last cost-effectiveness analysis to which you contributed, what standard did you use for the WTP threshold? (Check all that apply.)

- ▢ 1x GDP (1)
- ▢ 3x GDP (2)
- ▢ Other (3) ____

Q10 Why did you use this threshold? (Check all that apply.)

- ◯ It is the convention in the literature. (1)
- ◯ It is approximately the true societal WTP. (2)
- ◯ It is approximately the true WTP according to the current budget. (3)
- ◯ Other (4) ____

Q4c What was the setting of your last cost-effectiveness analysis?

- ◯ High-income countries (1)
- ◯ Low- and middle-income countries (2)
- ◯ A combination of both (3)

Demographic Information

Q11 Current position

- ○ Government (6)
- ○ Industry (5)
- ○ Master's Student (1)
- ○ PhD Student (2)
- ○ Post-doc/Research scientist (3)
- ○ Professor (4)
- ○ Other (7) ____

Q14 Gender identity

- ○ Male (1)
- ○ Female (2)
- ○ Non-binary (3)
- ○ Prefer to self describe (4) ____
- ○ Prefer not to say (5)

Q15 Age

____

Q16 Years of experience (coursework, research, OR teaching) with cost-effectiveness analysis

____

Q17 Number of about cost-effectiveness analyses to which you have contributed

- ○ 0 (1)
- ○ 1-2 (2)
- ○ 3-5 (3)
- ○ 5+ (4)

Q5f Which of the following are your main focus in cost-effectiveness analysis?

- ○ High-income countries (1)
- ○ Low- and middle-income countries (2)
- ○ A combination of both (3)

Q18 Any further comments?

____