Circulation: Heart Failure

ORIGINAL ARTICLE

Quantifying Benefit-Risk Preferences for Heart Failure Devices: A Stated-Preference Study

Shelby D. Reed, PhD; Jui-Chen Yang, MEM; Timothy Rickert, BS; F. Reed Johnson, PhD; Juan Marcos Gonzalez, PhD; Robert J. Mentz, MD; Mitchell W. Krucoff, MD; Sreekanth Vemulapalli, MD; Philip B. Adamson, MD, MSc; David J. Gebben, PhD; Liliana Rincon-Gonzalez, PhD; Anindita Saha, BS; Daniel Schaber, PharmD; Kenneth M. Stein, MD; Michelle E. Tarver, MD, PhD; Dean Bruhn-Ding, BS

BACKGROUND: Regulatory and clinical decisions involving health technologies require judgements about relative importance of their expected benefits and risks. We sought to quantify heart-failure patients' acceptance of therapeutic risks in exchange for improved effectiveness with implantable devices.

METHODS: Individuals with heart failure recruited from a national web panel or academic medical center completed a web-based discrete-choice experiment survey in which they were randomized to one of 40 blocks of 8 experimentally controlled choice questions comprised of 2 device scenarios and a no-device scenario. Device scenarios offered an additional year of physical functioning equivalent to New York Heart Association class III or a year with improved (ie, class II) symptoms, or both, with 30-day mortality risks ranging from 0% to 15%, in-hospital complication risks ranging from 0% to 40%, and a remote adjustment device feature. Logit-based regression models fit participants' choices as a function of health outcomes, risks and remote adjustment.

RESULTS: Latent-class analysis of 613 participants (mean age, 65; 49% female) revealed that two-thirds were best represented by a pro-device, more risk-tolerant class, accepting up to 9% (95% CI, 7%–11%) absolute risk of device-associated mortality for a one-year gain in improved functioning (New York Heart Association class II). Approximately 20% were best represented by a less risk-tolerant class, accepting a maximum device-associated mortality risk of 3% (95% CI, 1%–4%) for the same benefit. The remaining class had strong antidevice preferences, thus maximum-acceptable risk was not calculated.

CONCLUSIONS: Quantitative evidence on benefit-risk tradeoffs for implantable heart-failure device profiles may facilitate incorporating patients' views during product development, regulatory decision-making, and clinical practice.

Key Words: decision making ■ heart failure ■ patient preference ■ risk assessment ■ surveys and questionnaires

Several types of medical devices currently are being studied to treat heart failure, including interatrial shunts and percutaneous devices.[1,2,3] In addition, incremental improvements to existing types of devices like those used for transcatheter mitral valve repair, cardiac resynchronization therapy, and left-ventricular assist devices require ongoing benefit-risk assessments. This concept also applies to existing and new devices like those providing baroreflex neuromodulation and cardiac contractility modulation for patients with heart failure.[4]

Acceptability of benefit-risk tradeoffs requires value judgments by patients and clinicians. Quantifying patients' views on the relative importance of specified benefits and risks as well as other desirable or undesirable device features can assist medical-device developers in evaluating and ranking which prototype device profiles to pursue in clinical studies. Evidence about patients' acceptance

Circulation: Heart Failure is available at www.ahajournals.org/journal/circheartfailure

WHAT IS NEW?

- Our findings demonstrate that patients' views on benefit-risk tradeoffs can be elicited and quantified using a systematic, repeatable, theoretically sound approach to inform multiple decisions pertaining to the development, testing, and clinical use of new medical devices for heart failure.
- Patients with heart failure vary in their acceptance of device-associated adverse-event risks for specified durations of improvements in physical functioning and survival.

WHAT ARE THE CLINICAL IMPLICATIONS?

- Shared decision making pertaining to use of current and future heart failure devices could be improved with a structured, theoretically sound approach to eliciting patients' benefit-risk tradeoffs.
- Research is underway to determine how to best adapt and use stated-preference methods to elicit individual patient preferences to support shared decision making and its impact on satisfaction, adherence, and health outcomes.

Nonstandard Abbreviations and Acronyms

DUHS	Duke University Health System
MAR	maximum-acceptable risk
NYHA	New York Heart Association

of benefit-risk tradeoffs also may influence the design of clinical trials evaluating new medical devices. The United States Food and Drug Administration Center for Devices and Radiological Health recognizes the value of patient preference data in the evaluation of medical devices.[5] Patient preference data also can help to elucidate differences in stakeholders' judgments, improve transparency in decision making, and potentially resolve disagreements among stakeholders.

Decades of marketing and economics research have yielded a theoretically sound methodological framework to value nonmarket goods by eliciting stated preferences for sets of characteristics that vary with regard to their magnitude, quality, or type.[6] These stated-preference methods are increasingly being used to quantify the relative importance of medical intervention features to inform product-development decisions, regulatory decision-making, and individualized treatment decisions.

The Medical Device Innovation Consortium is a nonprofit public-private partnership with a mission to advance regulatory science pertaining to the medical device industry.[7] With funding from several medical device manufacturers and grant support from the Food and Drug Administration, Medical Device Innovation Consortium coordinated a study to quantify patient preferences for

features relevant to various heart failure devices, including safety and effectiveness outcomes.

METHODS

The data that support the study findings are available from the corresponding author upon reasonable request.

Best-practice methods were used to design a discrete-choice experiment survey to quantify patients' willingness to accept risks of adverse events in exchange for improved effectiveness for hypothetical implantable devices for heart failure.[8,9] Discrete-choice experiments employ hypothetical treatment alternatives described using characteristics or attributes, each of which has multiple possible levels. Choice questions present alternative treatment profiles representing combinations of attribute levels governed by an experimental design with known statistical properties. Participants are asked which alternative they would choose among sets of 2 or more alternatives.

Designing the study required group-level stakeholder discussion and agreement at several points. For this study, the team agreed to focus on patient-centered benefits, risks, and other features that could be important in regulatory decisions for future medical devices for patients with Stage C heart failure (Table 1). When selecting the attributes, the study team considered the decision context, respondent burden, potential co-dependencies between attributes, clinical and regulatory relevance, and importance to patients in making treatment decisions. The deliberative process was informed by a previously completed stated-preference study which found that treatment choices primarily were driven by improvements in physical functioning rather than the number of heart-failure hospitalizations.[10] The team built on the prior study by adding a time dimension to the attribute representing physical functioning. In this study, a no-device scenario was represented by a health trajectory representing the possibility of living three years with physical functioning and symptoms equivalent to New York Heart Association (NYHA) class III, followed by 2 years with functioning and symptoms equivalent to NYHA class IV. Levels representing improvements were framed in one of 2 ways: (1) as a 1-year improvement from NYHA class III to NYHA class II, and (2) an additional year with functioning equivalent to NYHA class III and associated extended survival times. Two risk attributes were included. As an unequivocally severe adverse event upon which to evaluate risk tolerance, the group chose risk of device-related 30-day mortality. The second risk attribute represented a collection of in-hospital potential complications that could occur to account for various adverse events. The study team discussed several potential device-associated features for inclusion, but ultimately agreed on an attribute representing whether the device could be adjusted remotely because cybersecurity was noted as a concern.[11]

Pretest Interviews

To test the draft survey instrument and guide revisions, the study team conducted 10 one-on-one interviews with a diverse set of patients recruited from cardiology practices of the coauthors (RJM, SV). Interviewers employed a think-aloud protocol in which participants were asked to read aloud and explain their reasoning as they completed the survey items, allowing

Table 1. Study Attributes and Levels

Attributes	Levels			
Benefits				
Disease trajectory Years in NYHA class II, III, and IV representing potential: Improvements in functioning Extended duration before worse functioning Longer survival duration		Years in		
		NYHA II	NYHA III	NYHA IV
	No-device scenario	0	3	2
	1-y gain in NYHA II	1	3	2
	1-y gain in NYHA III	0	4	2
	1-y gains in NYHA II and III	1	4	2
Device-associated risks				
Mortality risk All-cause death within 30 d of device implantation	0%* 2% 5% 10% or 15%†			
Complications risk An adverse event that would require a 2-d hospital stay with no long-term health consequences	0%* 5% 15% 40%			
Device features				
Remote device programming Device that can or cannot be adjusted via internet connection	No* Yes			

NYHA indicates New York Heart Association.

*This level also represents the no-device scenario.

†Patients were randomized to choice questions with 10% or 15% as the highest mortality risk levels for a scope test.

interviewers to gain insights into their interpretive strategies, detect unfamiliar terms, and identify where participants' understanding of information diverged from the intentions of the authors. The survey instrument included 10 quiz questions to assess participants' understanding of survey content, including icon arrays depicting risks and information portrayed in the choice questions. Participants responded to bidding-game questions to determine the size of benefit-risk contrasts necessary to induce switching between alternatives. The full survey instrument is available upon request. An example choice question is shown in Figure 1.

Experimental Design

Combinations of attribute levels shown for alternative device profiles in a discrete-choice experiment are based on an experimental design. The D-efficient design used in the study was generated using SAS, version 9.4 (SAS Institute, Inc, Cary, NC) and comprised of 40 blocks of 8 choice questions. Each participant was randomly assigned to answer one block of questions. A dominant-choice question was positioned as the initial question for each participant to assess decision-making logic for a total of nine unique choice questions. In the dominant-choice question, all three alternatives (ie, no device, device A, and device B) provided the same disease trajectories, but each device option included a risk level >0%. Thus, the logical choice was no device.

Recruitment

Participants were recruited from 2 sources. The first source was a national web panel (Kantar Health) of individuals in the

United States who self-reported a diagnosis of heart failure. The second source was patients with a physician-verified diagnosis of heart failure treated at Duke University Health System (DUHS). Individuals <18 years of age or those with congenital heart disease were excluded. The study protocol was reviewed and approved by the DUHS Institutional Review Board. Participants provided informed consent.

STATISTICAL ANALYSES

Data Quality

The study team developed a statistical analysis plan to predefine data-quality checks, the approach for model specification, and subgroups of interest. Quality checks included estimating time to survey completion, performance on comprehension questions, nonvariation in responses (choosing device A or device B for every choice question), responses to the dominant-choice question, and performance on three internal-validity tests. These included attribute dominance, across-set monotonicity (Table S1), and scope tests (see Supplemental Material).

Preference Weights

Conditional logit models were used for initial evaluations of model specification. Box-Cox specification tests were applied to determine whether risk attribute levels could be modeled using a continuous linear, log, or inverse
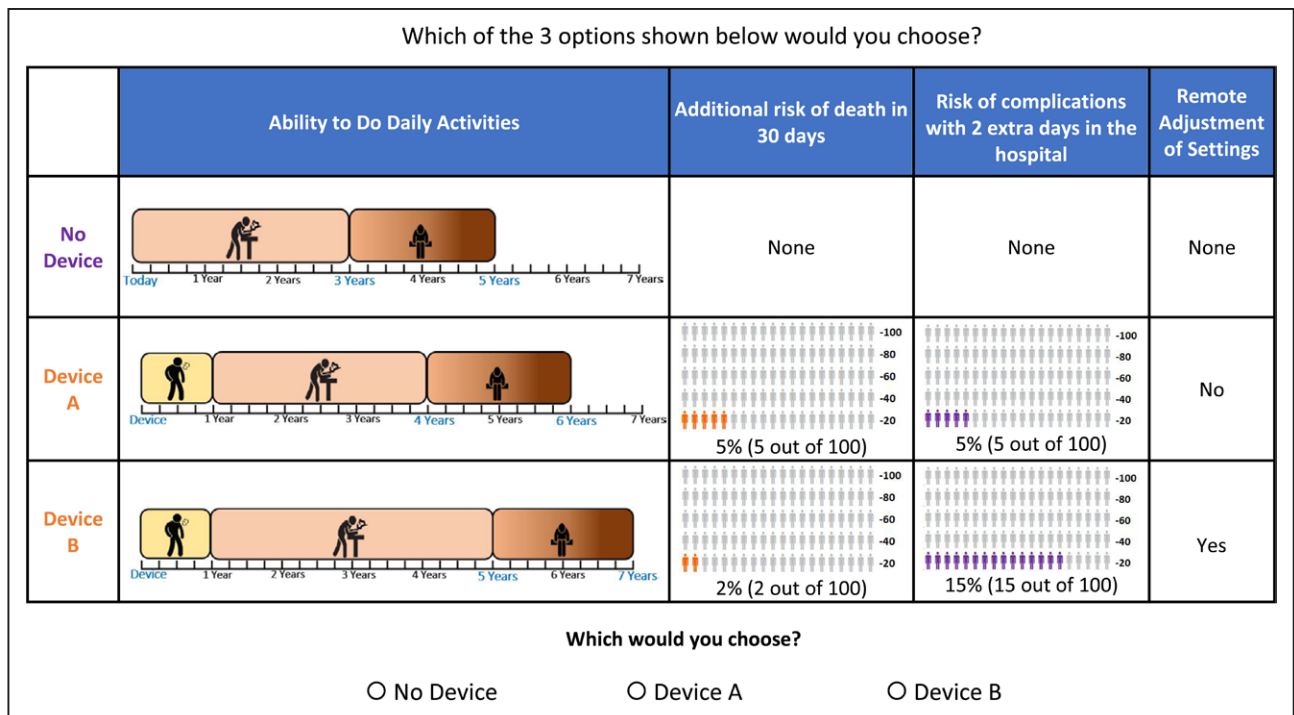
Figure 1. Example discrete-choice question.
One example of 320 possible choice questions

function. The appropriate form was determined based on whether model fit was significantly improved ($P<0.05$) with these transformations for each risk attribute versus a linear function on the basis of log-likelihood tests. An interaction term was included to account for a potential multiplicative effect of an additional year in both NYHA class II and class III. An additional parameter was included to test for a potential pro- or antidevice effect that was independent of benefit, risk, and remote-monitoring attributes.[12]

Separate regression models were fit to the web panel and DUHS cohort. Methods described by Hensher et al[13] that adjust for potential differences in preference consistency were used to compare the relative importance of attributes between the web panel and DUHS samples.

In the final model, population-level preference weights were estimated using fully correlated random-parameters logit regression using Stata/SE 16 (Stata, College Station, TX) since this approach (as opposed to uncorrelated parameters) avoids potential bias and we expected some correlation between attribute levels.[14] Heterogeneity in preferences for each attribute level was assumed to follow normal distributions across respondents.[15] A complementary and efficient approach to examining preference heterogeneity was undertaken using logit-based latent-class analysis with Latent-GOLD software, version 5.1 (Statistical Innovations, Inc. Arlington, MA) to identify underlying variations in preference patterns and to evaluate whether predefined

subgroup characteristics were independently associated with class membership. The predefined subgroups were age, gender, NYHA class, race, presence of arrhythmia, previous device, type of heart failure, and time since diagnosis.

Maximum-Acceptable Risk

Preference-weight estimates generated from the random-parameters and latent-class models were used to compute ex-post estimates of maximum-acceptable risk (MAR) of 30-day mortality or in-hospital complications that participants would independently accept to achieve specified improvements in health outcomes or desirable features. MARs were calculated by equating the decrease in preference utility associated with mortality or complication risks to the increase in preference utility associated with a specified gain in physical functioning and associated survival.[16] CIs for MAR estimates were estimated using the Krinsky-Robb procedure.[17]

Sensitivity Analysis

To address potential concerns about participants' understanding of survey content and decision-making logic choices, sensitivity analyses were conducted to evaluate the impact of limiting the analysis to smaller subsets of participants who exhibited better performance on comprehension questions and the initial

dominant-choice question. Two subsets of participants were evaluated: the better subset included participants who appropriately chose no device in the dominant-choice question and correctly answered at least 5 of 10 comprehension questions, and the best subset only included participants who appropriately chose no device in the dominant-choice question and correctly answered at least 8 of the 10 comprehension questions.

RESULTS

In March 2020, 500 panelists from the web panel, and between April and July 2020, 126 patients from DUHS completed the web-based survey. Median time to completion was shorter in the web panel (21 minutes) than in the DUHS sample (29 minutes, *P*<0.001). Mean numbers of ten comprehension questions correctly answered were 6.9 in the web panel and 7.4 in the DUHS group (*P*=0.018). Both samples were evenly split by gender (51% male) with similar educational attainment (Table 2). However, the DUHS sample included a larger percentage of African American respondents (32% versus 8%, *P*<0.001). The DUHS sample also reported significantly more comorbidities, including anemia, arrhythmias, arthritis, hypertension, and obesity.

Forty-four (8.8%) and 10 (7.9%) of participants in the web panel and DUHS samples, respectively, always chose the no-device option across eight choice questions. Eleven participants in the web panel and 2 participants in the DUHS sample were excluded from choice-data analyses because they chose the same device alternative in all 8 choice questions (the probability this would occur by chance is 0.02%).

Aggregate-Level Preferences

Results from logit-based analysis of discrete-choice experiment data is interpreted as the log of the odds of treatment choice given differences in treatment characteristics. These log-odds are also known as preference weights because they convey the impact of specific attribute levels on treatment preferences.[18] In general, better clinical outcomes were preferred, and having a device was preferred over status quo or having no device (corresponding to a total utility of 0; summing across attributes). Both a 1-year improvement to NYHA class II symptoms and one more year in NYHA class III were preferred over no improvement in physical functioning. There was a significant negative interaction indicating that the joint utility gain for the combined effect of an additional year in NYHA class II and an additional year in NYHA class III was less than the sum of the individual utility gains from a year in NYHA II and an additional year in NYHA III

Table 2. Demographic and Disease Characteristics for Web Panel and DUHS Samples

Characteristic	Web panel (N=500)	DUHS (N=126)	P value
Male, %	52%	47%	0.10
Age in years, mean (SD)	64 (12)	66 (11)	0.06
Hispanic or Latino, %	5%	3%	0.74
Race, %			
White	89%	66%	<0.001
Black	8%	32%	<0.001
American Indian/Alaskan Native	3%	3%	0.75
Asian	2%	0	0.37
Other	1%	0	0.59
Education, %*			
High school or less	19%	23%	0.16
Some college but no degree	28%	19%	
Associate degree/technical school	16%	19%	
4-y degree (± some graduate studies)	23%	20%	
Graduate or professional degree	14%	19%	
Days over the past 7 d with mean (SD)†			
NYHA I	2.0 (2.8)	2.0 (2.8)	0.95
NYHA II	2.4 (2.6)	2.8 (2.8)	0.15
NYHA III	2.1 (2.7)	1.8 (2.5)	0.32
NYHA IV	0.5 (1.5)	0.4 (1.2)	0.44
Heart failure management			
Take prescription medications	95%	92%	0.13
Changed diet	59%	68%	0.05
Heart valve device	3%	2%	0.78
Heart valve repair/replacement	7%	7%	0.92
Implantable cardiac defibrillator	24%	31%	0.13
Cardiac resynchronization device	2%	2%	1.0
Pacemaker	20%	23%	0.52
Stents	29%	27%	0.69
Comorbidities, %			
Anemia	14%	22%	0.02
Arrhythmias	16%	26%	0.002
Arthritis	36%	53%	<0.001
Cancer	14%	18%	0.36
Chronic obstructive pulmonary disease	24%	19%	0.26
Diabetes	37%	33%	0.44
Prior myocardial infarction	40%	31%	0.06
Hypertension	65%	76%	0.02
Kidney disease	14%	20%	0.08
Obesity	29%	38%	0.04

DUHS indicates Duke University Health System; and NYHA, New York Heart Association.
* Percentages do not include missing responses in the denominator.
†Statistics do not include respondents who did not answer the question about the number of days they experienced each level of physical functioning.

(*P*<0.001). Participants logically preferred lower risks of death and complications and slightly preferred the remote programming device feature. And, despite the

finding that the DUHS sample revealed stronger pro-device preferences compared with the web panel, the samples overall were similar in terms of both consistency, or scale ($P=0.17$), and benefit-risk preferences ($P=0.78$).

The largest difference in preference weights associated with an attribute define how much an attribute can influence treatment choice in the experiment. Hence, it is commonly considered a form of attribute importance.[19] There were no statistically significant differences in importance of the risk of death or risk of complications relative to effectiveness across the 2 samples (Figure 2). Based on these similarities, subsequent results are based on the pooled sample (model estimates in Table S2).

Based on the preference estimates obtained with the pooled sample, participants would accept a MAR of 30-day mortality of 6% (95% CI, 5%–8%) with a device offering one year of improved physical functioning (NYHA class II) and a MAR of 5% (95% CI, 4%–6%) with a device offering one additional year of functioning equivalent to NYHA class III. Acceptance of in-hospital complications was greater with participants accepting MARs of 27% (95% CI: 22%–34%) and 21% (95% CI,

16%–25%) for 1-year gains in NYHA class II and NYHA class III, respectively.

In sensitivity analysis, compared with the pooled sample, risk-tolerance estimates were higher for respondents in the better subset (ie, n=449 who chose the better device in the dominant-choice question and correctly answered 5 or more comprehension questions). Risk tolerance was even greater among the more narrowly defined best subset (259 participants who also correctly answered at least 8 comprehension questions; Figure 3).

Latent-Class Preferences

A 3-class latent-class model provided: (1) good fit to the data, relative to a one- or two-class model, based on Bayesian information criterion and Akaike information criterion statistics; (2) parsimony; and (3) consistency with results in the pooled sample. Sixty-six percent of all participants were best represented by the more pro-device, more risk-tolerant class. This class was willing to accept a 9% (95% CI, 7%–11%) or 7% (95% CI, 6%–8%) maximum risk of mortality for 1-year gains in survival with NYHA II or III functioning, respectively (Figure 4).
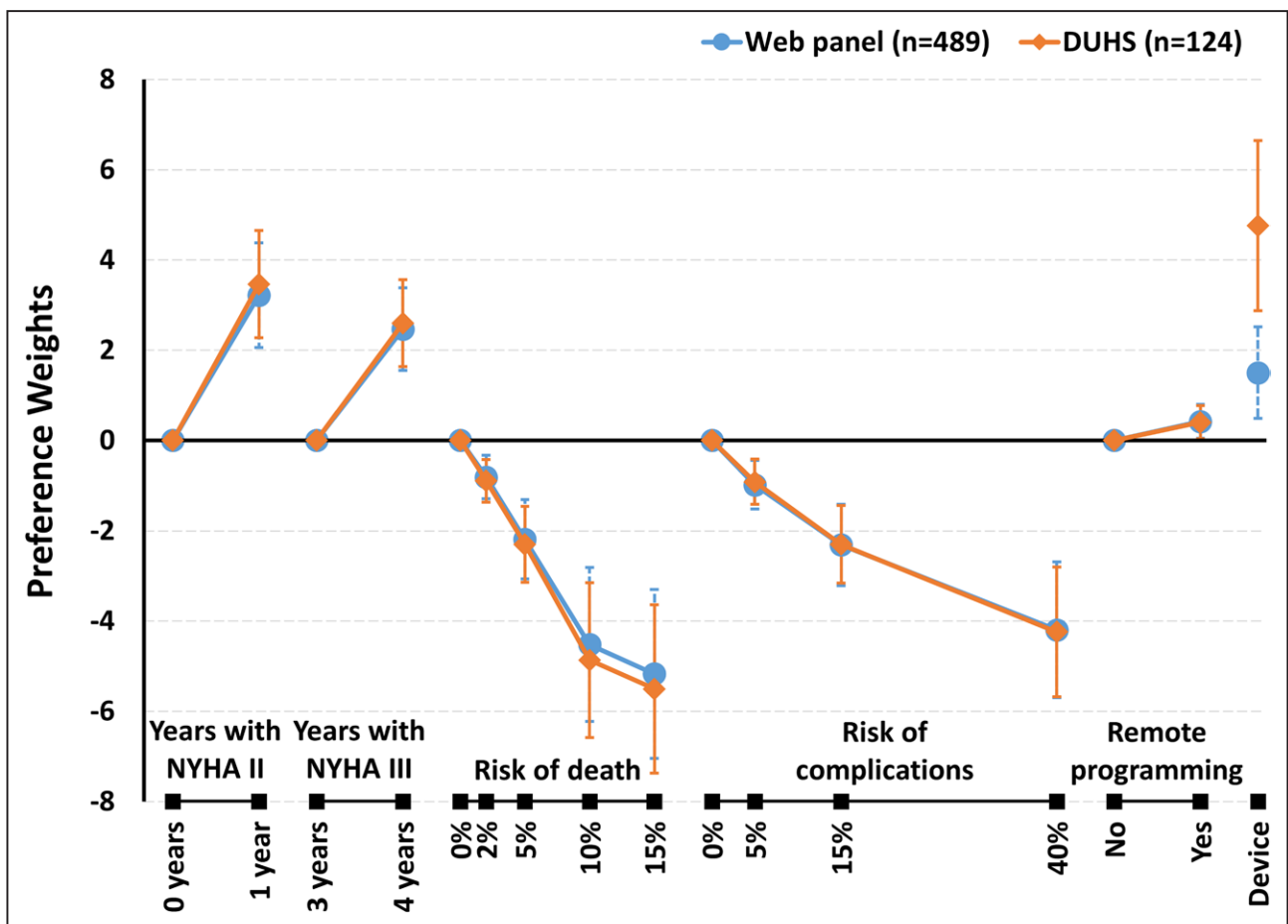
Figure 2. Preference weights for online-panel and Duke University Health System (DUHS) samples.
Larger weights represent more positive preference, and smaller weights represent more negative preference. Vertical bars represent 95% CIs. NYHA indicates New York Heart Association.
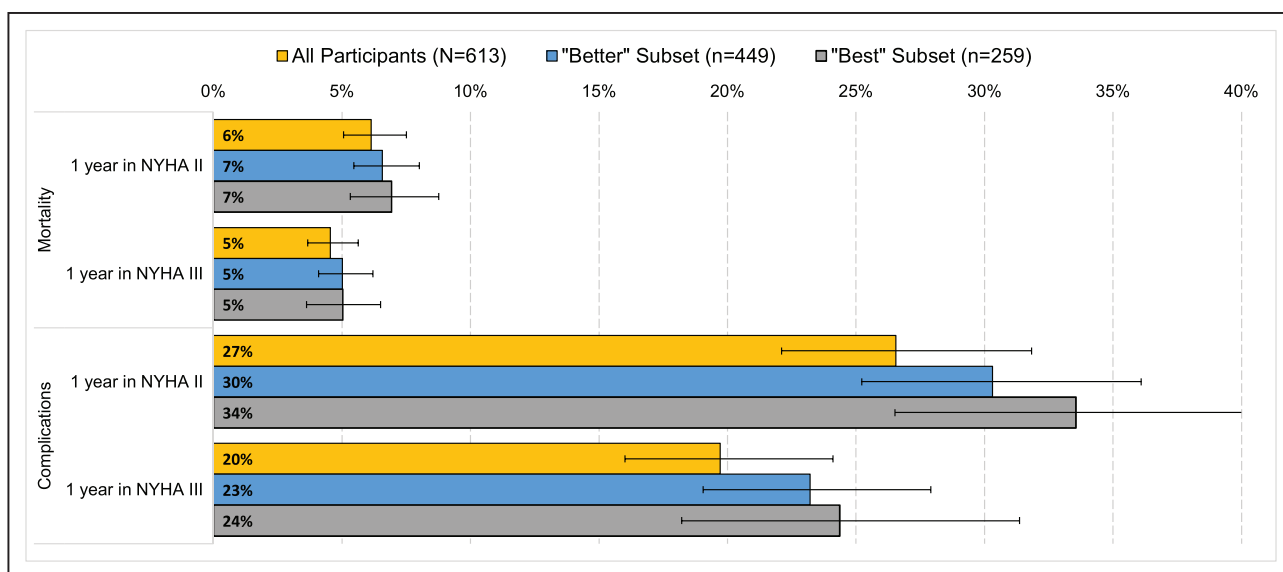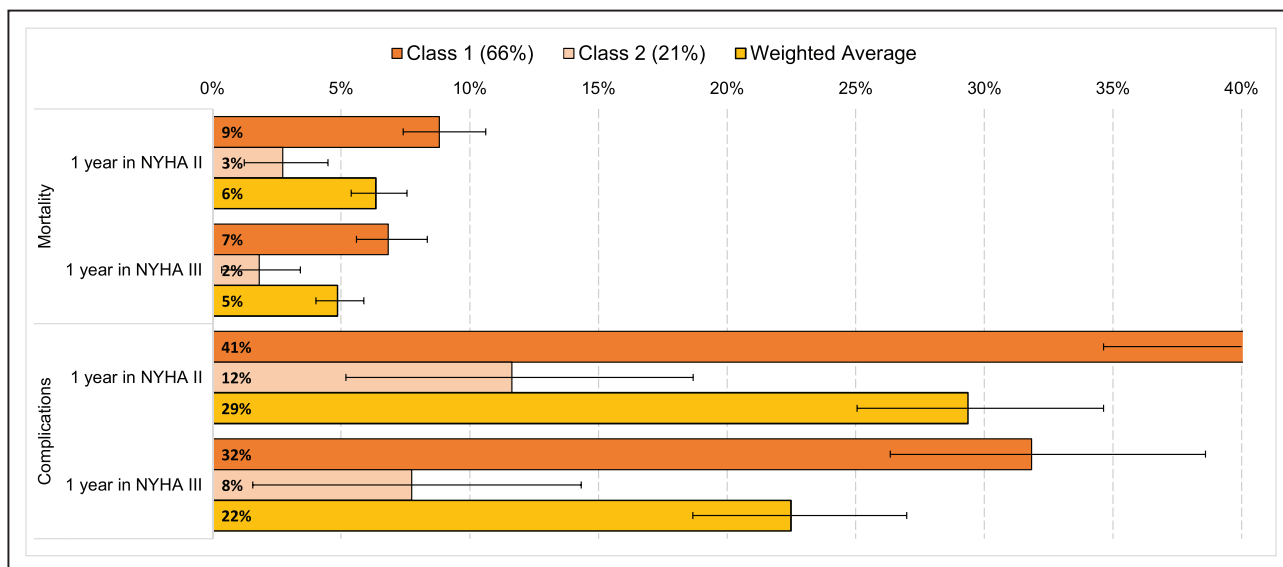
Figure 3. Maximum-acceptable risks (MAR) of 30-day mortality and complications: sensitivity analysis.
Better includes participants who appropriately chose no device in the dominant-choice question and correctly answered at least 5 of 10 comprehension questions. Best further limits the subset to participants who correctly answered at least 8 of the 10 comprehension questions. All MAR estimates are censored at the maximum risk levels included in device scenarios (15% for death and 40% for in-hospital complications). NYHA indicates New York Heart Association.

Twenty-one percent were predicted to be members of a less risk-tolerant class, accepting a device with mortality risks of 3% (95% CI, 1%–4%) or 2% (95% CI, 0.4%–3%) for 1-year gains in survival with NYHA II or III functioning, respectively. The pro-device, more risk-tolerant class was accepting of a 41% risk or a 32% risk of in-hospital complications with devices offering one-year gains in survival with NYHA II or NYHA III functioning, respectively. The corresponding MAR levels were 12% and 8% in the less risk-tolerant class. The third class that best characterized about 14% of participants had a strong antidevice sentiment across the device profiles

presented in the survey; thus, MARs were assumed to be 0 for this group. The weighted-average MARs using the 2 large latent classes approximated the MARs calculated from the parameter weights from the pooled model.

Participants with highest probability of being a member of the large, more risk-tolerant class were more likely to be male, have symptoms consistent with NYHA class I or II rather than III or IV, have been diagnosed with an arrhythmia, and have a previously implanted cardiac device (all $P<0.05$). None of the prespecified characteristics were associated with membership in the more risk-averse class. However, females and individuals without

Figure 4. Maximum-acceptable risks (MAR) of 30-day mortality and complications by latent class.
All MAR estimates are censored at the maximum risk levels included in device scenarios (15% for death and 40% for in-hospital complications). NYHA indicates New York Heart Association.

an arrhythmia diagnosis or a previously implanted medical device were more likely to be members of the class least likely to choose a device option (all $P<0.05$).

DISCUSSION

Heart failure trials increasingly have incorporated patient-centered outcomes such as physical, psychological, social functioning as well as signs and symptoms of heart failure.[20] While this trend has been well received, a comprehensive view of an intervention's net benefit to a patient also accounts for their views on negative aspects such as side effects and risks. Our study provides evidence that most participants with heart failure were willing to accept single-digit risks of mortality for 1-year gains in survival during which their physical functioning would be stable or improve. For these health gains, the majority would accept in-hospital complication risks occurring as frequently as one in 4 to one in 5 patients. Our estimates of risk tolerance would have been greater if we had accounted for the pro-device effects that were observed, particularly in the DUHS sample. Conversely, as revealed in the latent-class analysis, about one in 7 patients generally were unwilling to accept these risks despite improvements in functioning and survival consistent with the device scenarios offered.

Short-term mortality risk frequently is used in patient preference studies to assess risk tolerance for gains in health outcomes and other desirable characteristics, thereby facilitating comparisons across studies. A previous study found that patients would accept a nearly 2 percentage-point increase in mortality risk if their physical functioning could improve from levels consistent with NYHA class III to NYHA class II,[10] an MAR lower than this study's pooled estimate of 6%. The difference may be attributable to how the effectiveness gains were characterized. The former study did not inform participants how long they would maintain improvements in physical functioning, but the current study explicitly depicted a one-year gain with physical functioning improvements with a corresponding one-year gain in survival. Although adding a time dimension more fully describes the health gain to be valued, it increased the complexity of the choice questions. Two comprehension questions designed to test participants' understanding of portrayed disease trajectories revealed that only about 60% of participants provided correct responses. Even though this level of performance may be concerning, when participants provided incorrect responses, the web-based survey was programmed to provide an explanation about the appropriate interpretation. Also, in sensitivity analysis in which we increasingly limited the analysis to participants demonstrating higher levels of understanding, study results were consistent with findings from the pooled cohort.

The study team chose to represent device-associated risks as a group of potential complications that could occur, representing a departure from the convention in health-preference research of choosing a specific adverse event to measure risk tolerance. With this decision, the team implicitly assumed that patients' concerns about these complications were not driven by the specific physiological or device-related problem, but by the incident's impact on medical care required and long-term health impact. This approach facilitates application of the study results across different types of devices and may be more reflective of clinical and regulatory decision making in which it is necessary to simultaneously consider risks of multiple adverse events.

The study also contributes to a scarce body of literature on the impact of different sources for participants in preference studies. Our finding that there were no statistically significant differences in benefit-risk preferences between participants recruited from a web panel with self-reported heart failure versus an academic medical center with a physician-confirmed diagnosis is consistent with another preference study in heart failure patients.[10] Both samples were similar in terms of age and gender with a national sample of patients with heart failure,[21] but were slightly more educated (Table S3). Although both samples represent a broader spectrum of disease severity than just stage C heart failure, limiting inclusion criteria would not have availed us to the finding that participants with NYHA class III and IV symptoms were less risk tolerant than patients with less advanced disease.

Quantifying patients' preferences for potential benefits and risks is helpful to a broad range of stakeholders committed to improving the lives of patients with heart failure. For medical-device development companies, these findings can help support: (1) internal product development decisions; (2) planning clinical trials; and (3) drafting submission packages to inform regulatory decision-making. Our study shows that companies choosing to work together in a precompetitive space can develop a patient-preference study that has the potential to be used in support of multiple Food and Drug Administration submissions. The device-agnostic nature of this patient preference study also may reduce concerns about potential biases that could favor one type of device relative to another. If a company has a device with a unique feature that could influence its benefit-risk profile, this survey instrument could be augmented with one or more additional device-related attributes and used in a new patient preference study. For example, Center for Devices and Radiological Health used a tool that incorporated results from a patient preference study designed to evaluate various benefits and risks associated with different features of medical devices for obesity to help inform an approval decision for a novel weight-loss device.[22]

Results of preference studies also can be used to plan clinical trials, select patient-centered end points, influence power calculations where effect-size assumptions are based on estimates of minimum-acceptable benefit that

patients would require to accept a device with a defined risk, and provide rationale for adjusting acceptable type I and type II statistical errors.[23] Through coordination with the Medical Device Innovation Consortium, the study group is engaged in a collaborative effort to apply Bayesian decision analysis to account for patient preferences in determining the optimal threshold for statistical significance for a 2-arm randomized controlled trial.[23]

From a health care perspective, our findings confirm that patients' acceptance of risks in exchange for therapeutic benefits is not uniform. Even though several demographic and disease factors were significantly associated with different preference patterns and levels of risk tolerance, reliance on patient characteristics is insufficient to guide treatment selections. This observation underscores the need for robust shared decision-making that goes beyond education about alternative treatments and includes eliciting a patient's preferences about associated benefits and risks as well as a clinician's acumen in regard to personalized risks that account for a patient's health history. It should be acknowledged that providers may find it difficult to elicit meaningful preferences from patients in an unstructured manner. Although there are innumerable factors beyond the specific patient preferences evaluated in this study that can influence real-world treatment decisions (eg, coverage policy, recovery time, medical comorbidities), use of a structured, repeatable elicitation process such as the use of experimentally controlled stated-choice questions could be useful in documenting patients' benefit-risk preferences to support elective treatment decisions or as a covariate when examining practice variations. Although this approach holds great potential, evidence on the effects of individualized preference elicitation for use in shared decision making is limited, and it is acknowledged that some patients wish not to be actively engaged in procedure-related decisions.[24]

One may reasonably question how a regulator would use a continuum of risk tolerance across patients to inform a dichotomous approval decision. The Center for Devices and Radiological Health recognizes the importance of preference heterogeneity and has expressed a willingness to consider the preferences of subsets of patients in whom benefits outweigh risks.[5] Additionally, examining the spectrum of risk-benefit preferences across racially and socioeconomically diverse patient samples may offer important insights on factors that motivate or dissuade individuals from enrolling in clinical trials or opting for interventions offered in clinical practice.

Study Limitations

Several limitations are common to all stated-preference studies. One is the hypothetical context in which participants are asked to make choices. To increase the perceived consequentiality of their responses, participants were informed that their effort was needed for the study results to be correct and that the findings would be used. Another potential limitation pertains to participants' understanding of the survey content. However, the extent to which participants had lesser or greater understanding of the tasks and content did not significantly influence the study's findings. A study-specific constraint was the lack of a specific device upon which to select relevant benefit and risk attributes (see Table S4 for predicted choice probabilities for two hypothetical device profiles). On the contrary, the multi-stakeholder project was intended to generate transferable information across different types of devices.

Conclusions

Bringing together medical device companies, regulators, patients, and clinical and stated-preference research experts in a precompetitive collaborative effort yielded quality evidence on benefit-risk tradeoffs acceptable to patients for heart failure devices. Approximately two-thirds of the patients were more risk tolerant, willing to accept up to 9% mortality risk to improve one NYHA functional class for a year. These data can be used to incorporate patients' views during product development, clinical-trial planning and recruitment, regulatory decision-making, and clinical practice.

Supplemental Material
Methods and Results
Tables S1–S4

REFERENCES

1. Griffin JM, Borlaug BA, Komtebedde J, Litwin SE, Shah SJ, Kaye DM, Hoendermis E, Hasenfuß G, Gustafsson F, Wolsk E, et al. Impact of interatrial shunts on invasive hemodynamics and exercise tolerance in patients with heart failure. *J Am Heart Assoc.* 2020;9:e016760. doi: 10.1161/JAHA.120.016760

2. Kaye DM, Hasenfuß G, Neuzil P, Post MC, Doughty R, Trochu JN, Kolodziej A, Westenfeld R, Penicka M, Rosenberg M, et al. One-year outcomes after transcatheter insertion of an interatrial shunt device for the management of heart failure with preserved ejection fraction. *Circ Heart Fail.* 2016;9:e003662. doi: 10.1161/CIRCHEARTFAILURE.116.003662

3. Karki R, Friedman PA, Killu AM. The future of percutaneous epicardial interventions. *Card Electrophysiol Clin.* 2020;12:419–430. doi: 10.1016/j.ccep.2020.04.007

4. Zile MR, Lindenfeld J, Weaver FA, Zannad F, Galle E, Rogers T, Abraham WT. Baroreflex activation therapy in patients with heart failure with reduced ejection fraction. *J Am Coll Cardiol.* 2020;76:1–13. doi: 10.1016/j.jacc.2020.05.015

5. U.S. Department of Health and Human Services Food and Drug Administration, Center for Devices and Radiological Health and Center for Biologics Evaluation and Research. Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and other Stakeholders. Available at: http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446680.pdf. Issued August 24, 2016.

6. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Application.* 2000; Cambridge University Press.

7. Medical Device Innovation Consortium (MDIC). Patient Centered Benefit-Risk (PCBR). Available at: https://mdic.org/project/patient-centered-benefit-risk-pcbr/. Accessed September 28, 2020.

8. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, Bresnahan BW, Kanninen B, Bridges JF. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental design good research practices task force. *Value Health.* 2013;16:3–13. doi: 10.1016/j.jval.2012.08.2223

9. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, Johnson FR, Mauskopf J. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health.* 2011;14:403–413. doi: 10.1016/j.jval.2010.11.013

10. Reed SD, Fairchild AO, Johnson FR, Gonzalez JM, Mentz RJ, Krucoff MW, Vemulapalli S. Patients' willingness to accept mitral valve procedure-associated risks varies across severity of heart failure symptoms. *Circ Cardiovasc Interv.* 2019;12:e008051. doi: 10.1161/CIRCINTERVENTIONS.119.008051

11. Alexander B, Haseeb S, Baranchuk A. Are implanted electronic devices hackable? *Trends Cardiovasc Med.* 2019;29:476–480. doi: 10.1016/j.tcm.2018.11.011

12. de de Bekker-Grob EW, Hol L, Donkers B, van Dam L, Habbema JD, van Leerdam ME, Kuipers EJ, Essink-Bot ML, Steyerberg EW. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value Health.* 2010; 13:315–323. doi: 10.1111/j.1524-4733.2009.00670.x

13. Hensher DA, Rose JM, Greene WH. Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects. *J Transp Geogr.* 2008; 16:126–133.

14. Hess S, Train K. Correlation and scale in mixed logit models. *J Choice Model.* 2017; 23, (Suppl C):1–8.

15. Train K. *Discrete Choice Methods with Simulation.* Cambridge University Press; 2003.

16. Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. *Med Decis Making.* 2011;31:469–480. doi: 10.1177/0272989X10386116

17. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ.* 2007;16:827–840. doi: 10.1002/hec.1197

18. McFadden D. Conditional logit analysis of qualitative choice behavior. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P, ed. *Frontiers in Econometrics.* Academic Press. 1973; 105–142.

19. Gonzalez JM. A guide to measuring and interpreting attribute importance. *Patient.* 2019;12:287–295. doi: 10.1007/s40271-019-00360-3

20. Blom JW, El Azzi M, Wopereis DM, Glynn L, Muth C, van Driel ML. Reporting of patient-centred outcomes in heart failure trials: are patient preferences being ignored? *Heart Fail Rev.* 2015;20:385–392. doi: 10.1007/s10741-015-9476-9

21. Komanduri S, Jadhao Y, Guduru SS, Cheriyath P, Wert Y. Prevalence and risk factors of heart failure in the USA: NHANES 2013 - 2014 epidemiological follow-up study. *J Community Hosp Intern Med Perspect.* 2017;7:15–20. doi: 10.1080/20009666.2016.1264696

22. Ho MP, Gonzalez JM, Lerner HP, Neuland CY, Whang JM, McMurry-Heath M, Hauber AB, Irony T. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc.* 2015;29:2984–2993. doi: 10.1007/s00464-014-4044-2

23. Chaudhuri SE, Ho MP, Irony T, Sheldon M, Lo AW. Patient-centered clinical trials. *Drug Discov Today.* 2018;23:395–401. doi: 10.1016/j.drudis.2017.09.016

24. Matlock DD, Nowels CT, Masoudi FA, Sauer WH, Bekelman DB, Main DS, Kutner JS. Patient and cardiologist perceptions on decision making for implantable cardioverter-defibrillators: a qualitative study. *Pacing Clin Electrophysiol.* 2011;34:1634–1644. doi: 10.1111/j.1540-8159.2011.03237.x