Voting Questions

For each voting question, please use the following scale identifying your level of confidence - with a score of 1 being low or no confidence and 5 representing high confidence.

1	—	2	—	3	—	4	—	5
Low Confidence				*Intermediate*				*High Confidence*

1. * How confident are you that the following are standalone, meaningful primary health outcomes in research studies of heart failure treatment technologies:
    a. Heart failure hospitalization; **RESPONSE:4**
    b. Heart failure hospitalization or heart failure hospitalization equivalent events (i.e., outpatient IV therapy for heart failure); **RESPONSE: 4.5**
    c. Total Hospitalizations? **RESPONSE: 3.5**

Discussion:

- For each health outcome with greater than or equal to intermediate confidence (≥ 2.5), please discuss the appropriate length of follow-up post-heart failure intervention for assessing this outcome;

  RESPONSE:

  The follow-up duration is dependent upon (a) specific therapy and intent; (b) time required for intervention to take place and have it's effect; and, (c) the underlying disease. Hospitalization-based endpoints typically require extended follow-up whereas intermediate endpoints can be assessed in a more timely manner. Many heart failure interventions, such as Cardiac Resynchronization Therapy (CRT), have demonstrated that six months, when using appropriate surrogate/intermediate endpoints, is a sufficient timeframe to establish and confirm durability of clinical effect. Studies using morbidity and mortality, as primary endpoints, assessed the effects for at least a year, or longer. In the CRT therapy area, studies with morbidity and mortality endpoints were confirmatory of studies with surrogate/intermediate measures.[1, 2]

- Please discuss important considerations when assessing the merits of composite outcomes in research studies of heart failure treatment technologies which include the combination of mortality, heart failure hospitalization, or heart failure hospitalization equivalent events.

  RESPONSE:

  Randomized double-blind placebo-controlled trials with combined hospitalization and mortality endpoints have been the gold standard to demonstrate safety and efficacy and provide high confidence as a standalone, meaningful health outcome (score of 5). However, these trials delay access to important new technologies given large sample

[1] N Engl J Med 2002
[2] JAMA 2003; 289:2685-2694

size requirements, extended enrollment timeframes, and significant costs. Also, these trials may miss clinical benefits that are important to patients.

Heart failure hospitalizations provide meaningful information, however this endpoint is becoming diluted as many heart failure events are now being treated outside of a hospital admission. Heart failure equivalent endpoints capture a larger yield of true heart failure events and therefore this endpoint is scored higher (score 4.5) than heart failure hospitalizations alone (score 4). Heart failure equivalent endpoints warrant careful consideration of office-based infusion clinics to minimize risk of Investigator bias since Investigators can routinely schedule infusions in clinic. The endpoint of heart failure equivalents is stronger if the administration of intravenous (IV) therapy (e.g. IV diuretics/vasodilators) is urgent and unplanned.

Composite outcomes of combined endpoints of mortality and heart failure hospitalization (or heart failure hospitalization equivalent) provide high confidence but studies looking at these hard endpoints demand large populations and may miss clinical benefits that are important to patients, also delaying access to the technology. Intermediate endpoints can be valuable to provide earlier access to new technology while still allowing regulatory authorities and reimbursement agencies the opportunity for ongoing evaluation as the evidence of the new technology accumulates.

2. How confident are you that surrogate and intermediate endpoints are predictive of standalone, meaningful primary health outcomes (e.g., reduction in mitral regurgitation, cardiac remodeling, ejection fraction, or biomarkers) in clinical research studies of heart failure treatment technologies for:
    a. Heart failure with preserved ejection fraction; **RESPONSE: 3**
    b. Heart failure secondary to mitral regurgitation where the focus of therapy is mitral valve repair/ replacement; **RESPONSE: 3**
    c. Heart failure with reduced ejection fraction (e.g., cardiac remodeling, ejection fraction)? **RESPONSE: 4.5**

Discussion:

- If greater than or equal to intermediate confidence (≥ 2.5), please identify the specific surrogate or intermediate endpoints and associated disease or therapy which you believe are sufficiently predictive of meaningful health outcomes.

  RESPONSE:

  a) **Packer Clinical Composite Score (CCS)**
  b) **Left Ventricular End Systolic Volume**
  c) **Others that are important but will not be discussed include Quality of Life, Brain Natriuretic Peptide (BNP), patient activity, etc.**

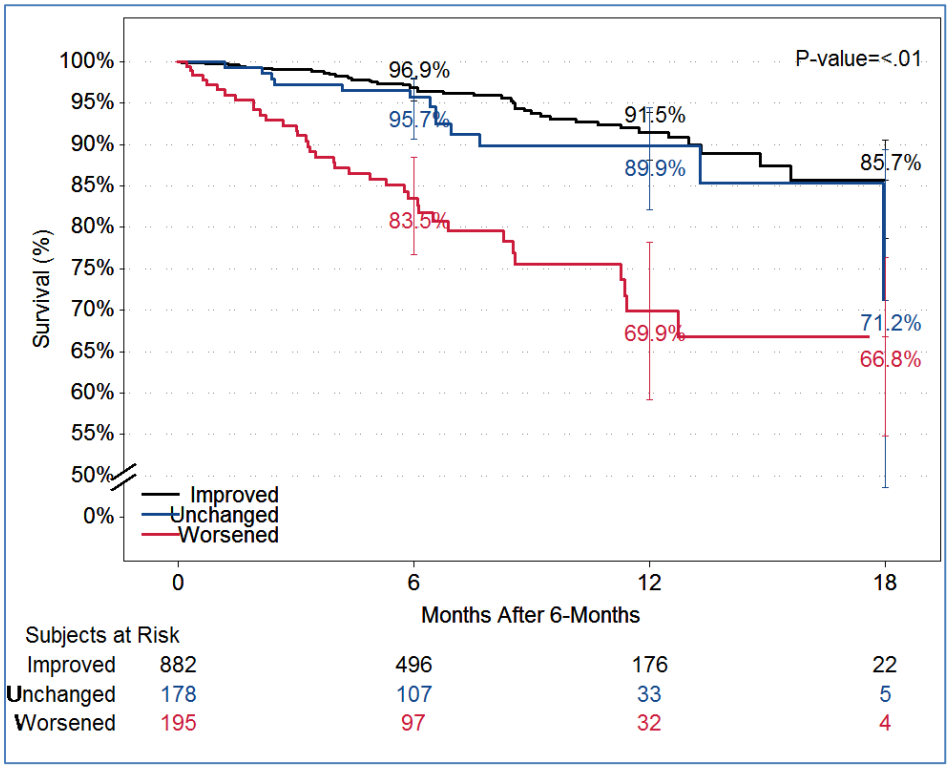
- Please discuss how these intermediate and surrogate endpoints meaningfully contribute towards the evidence base for heart failure treatment technologies.

  RESPONSE:

  Packer Clinical Composite Score (CCS): Introduced in 2001 as a means to measure intermediate effect on outcomes of an HF intervention, the CCS uses a hierarchical

approach to identify improving, unchanging, and worsening status for all patients studied. The CCS endpoint has been used in nearly 30 clinical trials for chronic HF, inclusive of US FDA IDE trials. CCS measured at 4-24 months reliably detected favorable effects in hard endpoints of mortality and HF hospitalization of β-blockers, ACE-inhibitors, and Cardiac Resynchronization Therapy and predated positive outcome results that were to come months to years thereafter. CCS was more sensitive than measures of cardiac function, effort tolerance, or patient-centered outcomes at identifying disappointing results with endothelin antagonists, inhibitors of tumor necrosis factor, xanthine oxidase inhibitors, and positive inotropic agents. In a review of 9 pivotal CRT trials, 4 trials using CCS as a key endpoint were conducted with approximately half the study duration and less than a third the sample size compared to 5 trials that measured morbidity/mortality as the primary endpoint. Using the intermediate endpoints for CRT allowed for earlier access to patients of a life-extending and life improving therapy that also reduced their risk of HF hospitalization. Additionally, a multivariate-adjusted analysis on pooled data from five prospective CRT studies (N= 1,603) that included CCS as an adjudicated endpoint[3], indicated that patients identified as improved or unchanged 6 months after CRT implantation, have considerably better survival compared to worsened (see Figure 1 below). In the same analysis, improved patients had a lower risk of HF hospitalization compared to unchanged and worsened.

Figure 1: Clinical Composite Score Correlates with Survival Outcomes



[3] Chung ES et al. Clinical and economic value of maximizing response to cardiac resynchronization therapy (CRT): Evidence from 5 randomized controlled trials. Circulation 2014, Issue 130, Supplement 2.

Left Ventricular End Systolic Volume (LVESV): Adverse myocardial remodeling has an important relationship to HF and is associated with symptoms, morbidity, and mortality, and thus is used widely in clinical practice. LVESV has been used in multiple trials, notably as a prospectively powered secondary endpoint in REVERSE and as a component in the BLOCK-HF and MADIT-CRT primary endpoints. Reverse remodeling, an improvement in myocardial function has been shown to be predictive of long-term outcomes and is unlikely to be influenced by patient or clinician bias.

- Please discuss important factors to consider when assessing the utility of surrogate and intermediate endpoints.

RESPONSE:

Endpoints should be:

- **Meaningful to patients, clinicians, providers, and payers**
- **Predictive of longer-term outcomes (i.e. high positive predictive value) and biologically plausible**
- **Widely available with minimal measurement burden**
- **Minimally influenced by bias**
- **Cost efficient for trial design and execution**

Importantly, the benefit-risk evaluation of a therapy does not stop at the pivotal phase. Post-approval study is an important tool to assess real-world and long-term effectiveness and to corroborate the intermediate findings.

3. How confident are you that quality of life measures [e.g., Kansas City Cardiomyopathy Questionnaire (KCCQ), Minnesota Living With Heart Failure Questionnaire (MLWHFQ):
   a. Are adequate measures which reflect the patient experience; **RESPONSE: 4.5**
   b. Should be included as the standalone, meaningful primary health outcomes in research studies; **RESPONSE: 2.5**
   c. Should be included as a composite standalone, meaningful primary health outcomes in research studies? **RESPONSE: 3.5**

RESPONSE:

Overall, Health-Related Quality of Life (HRQoL) measures, specific to the measurement of HF, can be meaningful endpoints. Systematic reviews on those measures have been performed recently[4] and in the recent past[5]. Kelkar et al. identified nine HRQoL instruments largely overlapping with the earlier Garin et al. systematic review. Both reviews agreed that the Kansas City Cardiomyopathy Questionnaire (KCCQ) and the Minnesota Living With Heart Failure Questionnaire (MLWHF) are the tools with most usage and evidence supporting their use. These instruments have been shown to be reliable (they produce a similar result on the same subject after multiple administrations), valid (items being tested are appropriate for the condition used) and responsive (the instruments are able to detect changes in HRQoL over time and across subject groups). Garin at al. identified KCCQ as more responsive when compared to

[4] Kelkar AA, Spertus J, Pang P et al. Utility of Patient-Reported Outcome Instruments in Heart Failure. JACC Heart Fail 2016;4:165-75.
[5] Garin O, Herdman M, Vilagut G et al. Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. Heart Fail Rev 2014;19:359-67.

MLWHF, but not drastically so. MLWHF[6] and KCCQ[7] have been reliably correlated with survival free from hospitalization; KCCQ has recently been additionally correlated with the risk of 30-day readmissions[8]. Both questionnaires can be burdensome to administer and retain, especially among the sickest patients[9]. A shorter version of the KCCQ has been proposed[10]. Despite the above, we are not aware of any effort to associate responses to either MLWHF and/or KCCQ with healthcare utility (valuation). This foundationally limits the ability to both measures to be used for comparisons across different disease states.

4. How confident are you that functional assessments [e.g., 6 min walk test (6MWT), VO2max, ventilator threshold]:
    a. Are adequate measures which reflect the patient experience; **RESPONSE: 3.5**
    b. Should be included as the standalone, meaningful primary health outcomes in research studies; **RESPONSE: 4**
    c. Should be included as a composite standalone, meaningful primary health outcomes in research studies? **RESPONSE: 4**

Discussion:

- Please discuss whether additional patient-reported measurement [e.g., Short Form-36 (SF-36), EuroQol five dimensions questionnaire (EQ5D)] should be considered to capture burdens associated with the heart failure therapy under study.

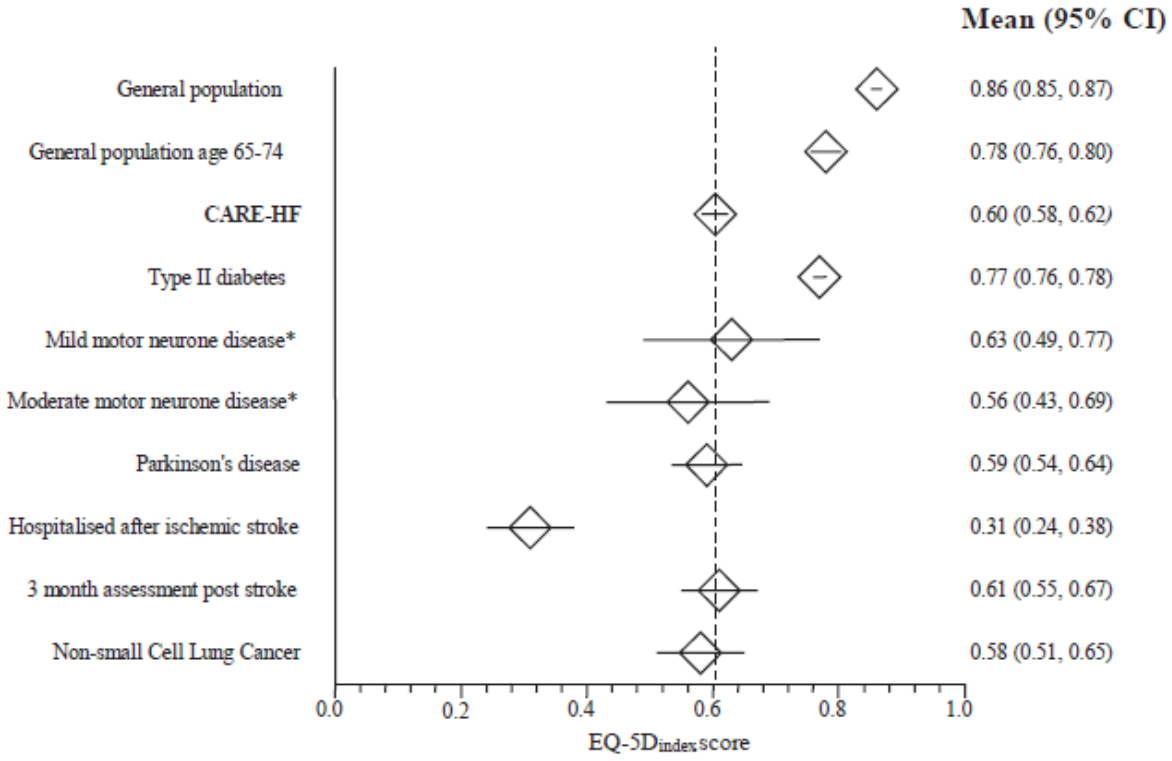
RESPONSE:

Generic health elicitation instruments (e.g. SF-36, EQ-5D, Health Utility Index (HUI)-III) have seen wide spread use. Contrary to HF-specific instruments discussed above, advantages of generic instruments include the more frequent existence of valuation studies for the latter and, thus, comparisons outside the HF disease space. Nevertheless, these instruments may be too broad to accurately assess health status. As HF becomes increasingly well treated, incremental improvements over Standard-of-Care (SoC) diminish in magnitude and deliver lower impact when put in perspective of holistic health. Calvert et al.[11] have outlined how different health states vary in terms of utility (Range 0-1) when calculated using the EQ-5D (Figure 2). With the difference of significantly debilitating HF (NYHA Class III) in CARE-HF to the General Population being 0.18, and the limited amount of distinct answer combinations available in these instruments; it is unlikely these instruments will continue to be sensitive enough to capture ever smaller incremental improvements. Recently, a more granular version of EQ-5D (EQ-5D-5L) has been introduced. Even if the sensitivity of EQ-5D-5L is improved

[6] Kato N, Kinugawa K, Seki S et al. Quality of life as an independent predictor for cardiac events and death in patients with heart failure. Circ J 2011;75:1661-9.
[7] Heidenreich PA, Spertus JA, Jones PG et al. Health status identifies heart failure outpatients at risk for hospitalization or death. J Am Coll Cardiol 2006;47:752-6.
[8] Dai S, Manoucheri M, Gui J et al. Kansas City Cardiomyopathy Questionnaire Utility in Prediction of 30-Day Readmission Rate in Patients with Chronic Heart Failure. Cardiol Res Pract 2016;2016:4571201.
[9] Gupta BP, Grady KL, Fendler T, Jones PG, Spertus JA. Variation of Quality of Life Data Collection Across INTERMACS Sites. J Card Fail 2016;22:323-37.
[10] Spertus JA, Jones PG. Development and Validation of a Short Version of the Kansas City Cardiomyopathy Questionnaire. Circ Cardiovasc Qual Outcomes 2015;8:469-76.
[11] Calvert MJ, Freemantle N, Cleland JG. The impact of chronic heart failure on health-related quality of life data acquired in the baseline phase of the CARE-HF study. Eur J Heart Fail 2005;7:243-51.

over the ordinary version (now called EQ-5D-3L), it is difficult to see how the new questionnaire will resolve all floor and ceiling effect issues. Nevertheless, EQ-5D and/or other instruments valued for healthcare utility can be administered simultaneously with disease-specific questionnaires to generate utility scores for the latter. The technique has been used successfully in CARE-HF[12].

Figure 2: Healthcare Utility Scores elicited using EQ-5D for Health States of Varying Severity



In addition to patient-reported measurements, patient activity is also emerging as an intermediate endpoint showing promise since it correlates closely with long-term heart failure outcomes. Many implantable devices include an accelerometer which collects physical activity continuously, providing real world assessment of patient ability. Recently, an expert panel (industry, academia and government) defined a meaningful endpoint from accelerometer data in clinical trials to be change in daily ambulatory activity measured in minutes of walking per day using an accelerometer in patients with class II/III heart failure[13]. Physical activity levels measured via accelerometer are known to be correlated with clinical outcomes in heart failure patient (mortality, HF hospitalization, age at implant and NYHA at implant).[14, 15]

One limitation of quality of life questionnaires and functional measures is that certain patient conditions may preclude appropriate use of those measures on the patients concerned. Simple examples are patients suffering with dementia in the case of quality

[12] Calvert MJ, Freemantle N, Yao G et al. Cost-effectiveness of cardiac resynchronization therapy: results from the CARE-HF trial. Eur Heart J 2005;26:2681-8.
[13] Clinical Trials Transformation Initiative Multi-stakeholder expert meeting. Developing Novel Endpoints Generated by Mobile Technology for use in Clinical Trials. Silver Spring MD, September 29-30, 2016.
[14] Conraads VM, Spruit MA, Braunschweig F, Cowie MR, Tavazzi L, Borggrefe M, Hill MR, Jacobs S, Gerritse B, van Veldhuisen DJ. Physical activity measured with implanted devices predicts patient outcome in chronic heart failure. Circ Heart Fail. 2014; 7(2): 279-87.
[15] Linde C, Tang A, Cowie M, Bergemann T, Abraham WT. Physical activity measured with implanted devices predicts heart failure outcomes. Submitted to ESC HF conference, 2017.

of life questionnaires and patients with arthritis, or other activity-limiting comorbidities, and functional measures such as the 6MWT.

- Please discuss the appropriate length of follow-up post-heart failure intervention for assessing patient-reported measurements.

  RESPONSE:

  The follow-up duration is dependent upon (a) specific therapy and intent; (b) time required for intervention to take place and have it's effect; and, (c) the underlying disease. Hospitalization-based endpoints typically require extended follow-up whereas intermediate endpoints can be assessed in a more timely manner. Many heart failure interventions such as Cardiac Resynchronization Therapy (CRT) have demonstrated six months, when using appropriate surrogate/intermediate endpoints, as a sufficient timeframe to establish and confirm durability of clinical effect. Studies using morbidity and mortality, as primary endpoints, assessed the effects for at least a year, or longer. In the CRT therapy area, studies with morbidity and mortality endpoints were confirmatory of studies with surrogate/intermediate measures.

- For some studies of heart failure treatment technologies it may not be practical for patients to be blinded. Please discuss the impact of unblinded study participants on patient-reported measurements and functional assessments.

  RESPONSE:

  Blinding strengthens trial designs by reducing bias in assessment of outcomes. Blinding is particularly important in trials whose outcome measures include subjective assessment of signs and symptoms. Blinding has successfully been implemented in device trials using intermediate endpoints with shorter evaluation timeframes. Longer evaluations may be considered unethical in device trials where a subject would be asked to undergo a device implant without receiving active therapy for an extended period of time. Blinding in device studies is complicated because the device itself may have an electrocardiographic or radiographic "signature" which can be detected by the investigator or other health care practitioners. Inadvertent crossovers could occur due to reasons such as Investigators viewing chest x-rays or echos, device upgrade discussions, patients receiving device identification cards in the mail, etc. Independent core labs and clinical events committees are also important to minimize the bias.

- Please discuss how to best consider the impact of adverse events associated with heart failure technologies while balancing the potential for improvements to meaningful health outcomes.

  RESPONSE: Several methods have been developed over the years to quantitatively assess risks and benefits of healthcare technologies. Increasingly, the term "risk" is associated more with uncertainty around a choice and less with the downside of particular choices. Therefore, these techniques, traditionally referred to as "Risk-Benefit Analysis" are now increasingly termed "Harm-Benefit Analysis". Number-Needed-to-Treat (NNT) and Number-Needed-to-Harm (NNH) represent basic epidemiological approaches to Harm-Benefit Analysis that may fail to accurately capture the totality of increasingly complex decision analytic problems. Figure 3 provides an overview of

several decision analytic techniques that have been developed over the years alongside strengths and weaknesses based on several criteria. Minelli et al. applied decision modelling to conduct a harm-benefit analysis using QALYs as the outcome measure; the foundational advantage of such an approach being that benefits from an intervention in a particular body area can be weighed against adverse events impacting a different bodily area/function[16]. Similar work is increasingly being conducted, with the latest evolution in holistic assessments over the entire expected patient lifespan including Multi-Criteria Decision Analysis (MCDA)[17] and Incremental Net Benefit[18]. We would like to underline that, importantly, these techniques do allow for trade-offs to be made in the absence of cost being part of the assessment.

Figure 3: Overview of Decision Analytic Techniques in Quantitative Harm-Benefit Analysis and Performance against Standardized Criteria

	METHODS													
CRITERIA	Chuang-Stein r_1 & r_2	RV-NNT	NNT	NNE	NNT_{US} & NNT_{UF}	RB Contour	Surplus Efficacy	RV-MCE	$NNT_{U\&T}$	Benefit-less-risk	NNT_T & MERT	Benefit-Risk Ratio	Q-TWiST	INB
Patient-sensitive - Accounts for differing baseline risks among patients	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Universal - Can be applied across interventions and diseases; Can be used to assess the population at risk or the patient at risk	+	+	+	+	+	+	-	+	+	+	+	+	+	+
Flexible - Can be conducted rapidly and with few resources; Can easily incorporate new knowledge acquired over time	-	+	+	+	+	+	+	+	+	+	+	+	+	+
Explicit Preferences - Weighs individual harms and benefits according to an explicit set of preferences from a relevant group	+	+	-	-	-	+	+	+	+	+	+	-	+	+
Easily-interpreted - Produces a graphical harm-benefit profile to facilitate comparison against no therapy or an appropriate comparator	-	-	-	-	-	+	+	+	-	-	+	+	+	+
Threshold - Has an intrinsic harm-benefit acceptability threshold	+	-	-	-	-	-	+	+	-	+	+	+	+	+
Integrates with Economic Evaluation Measures	-	-	+	+	+	-	-	-	+	+	-	+	+	+
Incorporates Uncertainty - Accounts for the quality and source of the benefit / harm information entered into the model; Provides a measure of precision (uncertainty) around the harm-benefit metric	-	-	+	+	+	-	-	-	-	-	-	+	-	+
Inclusive - Can incorporate multiple harms AND multiple benefits	-	-	-	-	-	-	-	-	+	-	+	-	-	+
Comprehensive - Can quantify both objective harms (e.g. mortality) and subjective benefits (e.g. QoL); Can quantify the duration, intensity, and reversibility of harms and benefits	-	-	-	-	-	-	-	-	-	-	-	-	+	+

[16] Minelli C, Abrams KR, Sutton AJ, Cooper NJ. Benefits and harms associated with hormone replacement therapy: clinical decision analysis. BMJ 2004;328:371.
[17] Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. Pharmacoepidemiol Drug Saf 2007;16 Suppl 1:S2-S15.
[18] Lynd LD, Najafzadeh M, Colley L et al. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making--a case study of alosetron in irritable bowel syndrome. Value Health 2010;13:411-7.

- Please discuss how to balance the benefits and harms of therapies which may improve near-term patient-reported health outcome assessments or clinical measurements (e.g., 6 MWT or symptoms) but may decrease length of life.

  RESPONSE: Patient preferences must be included in this assessment. In certain circumstances, patients may choose improvements in quality of life over quantity of life.

Additional Discussion Topics:

- Please discuss health outcomes of interest and appropriate follow-up duration in studies of technologies designed for diagnosis of acute decompensation of heart failure.

  RESPONSE:

  We have no additional comments beyond the information previously stated.

- With the health outcomes and information that we have discussed today, how confident are you that there will there be enough accurate information provided to patients for them to make informed decisions?

  RESPONSE:

  The amount of information available to patients has grown significantly in recent years and currently patients are able to obtain accurate information to make obtained decisions through a variety of sources. Patients or their caregivers can search for study data on devices through the ClinicalTrials.gov Web site, which provides a summary of each study protocol and requires that study results are posted within one year after study completion.

- Please discuss how studies can be designed to accurately capture patient preferences and how their preferences can best be considered and operationalized once the study has concluded.

  RESPONSE: We welcome the interest in incorporating patient preferences in the decision problem. We are unaware of a scientific consensus on the subject but we would like to specifically cite Bridges et al. on Patient-Based Health Technology Assessment (HTA)[19] and recent reviews on the use of patient preference studies in the US[20] and Internationally[21]. Finally, a recent review on choice experiments and their use in quantitative patient preference elicitation may also be informative[22].

[19] Bridges JF, Jones C. Patient-based health technology assessment: a vision of the future. Int J Technol Assess Health Care 2007;23:30-5.
[20] Johnson FR, Zhou M. Patient Preferences in Regulatory Benefit-Risk Assessments: A US Perspective. Value Health 2016;19:741-745.
[21] Muhlbacher AC, Juhnke C, Beyer AR, Garner S. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. Value Health 2016;19:734-740.
[22] Muhlbacher A, Johnson FR. Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. Appl Health Econ Health Policy 2016;14:253-66.