**How to visually integrate value judgment with clinical evidence**

Updated 10JAN2024

**Journal manuscript title: Integrating Value Judgment with Clinical Evidence**

**Authors:**

**Proposed submission date:**

**Outline *(brainstorming)***

**Abstract**

Xyz

**Terms:** benefit, risk, decision-enhanced visualizations, value judgment, thresholds, scoring, weighting

Objective

SBRA is a transparent and explicit decision-making process for assessing clinical evidence. Many aspects of the process rely on value judgements, which are either expert opinions or patient preferences. Even a simple clinical decision relies on value judgements such as selecting outcomes, assigning their importance, and adding thresholds for risk tolerance and benefit acceptance. More complex clinical or benefit-risk decisions require scoring and weighting dissimilar outcomes before combining them into a single score. In this paper, we use decision enhanced visualizations to communicate how value judgments are integrated with clinical evidence. The objective is easing the interpretation of visualizations by decision-makers.

**Terminology**

**[1] Clinical evidence** Objective study results based on a representative sample of study participants.

**[2]** **Value judgment** A subjective value based on expert opinion or personal preference, including levels of utility, risk tolerance, and benefit acceptance, which varies between stakeholders.[11]

**[3] Threshold – regarding minimal acceptable benefit and maximum acceptable risk**

**[4] Scoring -**

**[5] Swing weighting** – “It’s not a process of comparing apples with oranges, rather one of comparing preference for apples with preference for oranges.” -Larry Phillips, LSE

**[6] Ranking -**

**Background**

There are many challenges to clinical decision-making. These challenges include assessing multiple outcomes measured on different scales and having different consequences for the patients. This creates the scenario of comparing apples to oranges, something not encountered with analyzing a single endpoint (Ref. us guys). To overcome these challenges, enhancements can be made to graphs of clinical evidence. Our enhancements for integrating value judgement include scaling, weighting, thresholds and reference lines.

Value judgments are elicited from various stakeholders, including program leads, key opinion leaders, prescribers, and patients. They might not agree initially, but visualization can aid consensus. Elicitation techniques range from small, informal focus groups to large, rigorous patient preference studies. Recent regulatory guidance (ref. FDA & CIOMS) recommends incorporating patient experience into the decision process. Incorporation can be slow for two reasons. **First**, value judgments are perceived as subjective and prone to bias. This argument makes it even more imperative to make value judgment elicitation as transparent, explicit, and consistent as possible. The MDIC and IMI PREFER consortium recommend several valid and reliable elicitation techniques (see Appendix A). Bias can be mitigated by involving a representative sample of patients and physicians (ref. ) The second reason for slow incorporation is the unfamiliarity with methods for integrating value judgements with clinical evidence. This paper provides several examples.

Value judgements are used first in trial design specifications, including selection of the target population, treatment goals, clinical endpoints and patient-centered outcomes. Next, value judgements are used to set thresholds of success for the clinical endpoints and patients-centered outcomes. Thresholds are specified as a minimum clinically important difference (MCID), minimum acceptable benefit (MAB), maximum acceptable risk (MAR), or benefit-risk tradeoff curve (TC).

Currently, treatment effects are graphically displayed by comparing the summary statistics for the active treatment and comparator arms. If the confidence intervals for absolute treatment difference excludes zero or the relative difference excludes one, then statistical significance might be assumed. While vertical reference lines (see Figure 1) at zero or one are helpful, another reference point is needed to indicate MCIDs, MABs, MARs, and TCs.

An example of a MCID is achieving a treatment difference of 5mmHg or greater decrease in systolic blood pressure among hypertensive patients. To add interpretation by non-cardiologists, a graph can display reference lines for treatment differences of zero and 5 mmHg. Even a cardiologist will better interpret a graph of a disease-specific measures of quality-of-life outcome if the MAB is displayed as well.

Thresholds for MARs are particularly useful because adverse events (AE) differ in severity, duration, and reversibility. For example, a forest plot of a common, nonserious AE (ex. runny nose) and a rare but serious AE (ex. death) might be misinterpreted. It is too easy for a reviewer to scan down a forest plot looking for incidence proportions that appear farthest to the left of the graph. Displaying the MARs of 20% and 1% for runny nose and death, respectively, will aid interpretation.

Currently, the tradeoffs in benefit and risk outcomes between two or more treatments or doses are displayed in a table. The tradeoff, for example Drug A with high benefit and high risk, versus Drug B with less benefit and less risk, can be better displayed in a tradeoff plot. To make this plot more interpretable, the thresholds for MAB and MAR, as well as the TC (see Figure 2). To graph a TC, physicians and patients are asked what size benefit is expected for each level of risk along the Y-axis. At first glance, adding thresholds is a simple programming task, but eliciting them is not. Stakeholders, including the sponsor, regulator, physicians, and patient, will have different perspectives, hence different value judgements. Physicians and patient preference studies have shown that often physicians have a lower risk tolerance than patients (Ref.\_\_\_).

The next level of integrating value judgements with clinical evidence is ranking and weighting the benefit-risk outcomes. The ranking or ordering outcomes by their level of importance can be elicited using patient focus groups, A forest plot can be enhanced by ordering the outcomes by their importance.

If multiple benefits and risks are to be combined into a single score, they will first need to be assigned numerical weights. A rare but serious AE would be assigned a very high weight relative to a common, nonserious AE. Weights can be elicited from several hundred patients, using a discrete choice experiment (DCEs). Weights can be incorporated in several types of assessments, the most common being a Multi-Criteria Decision Analysis (MCDA). An MCDA sums the weighted clinical evidence, so as to compare total scores across several drug options. An MCDA generates several graphical displays that show the impact of each outcome on its total score, aid the determination of which drug has the most favorable benefit-risk profile. Because weights are more subjective than clinical results, different stakeholder perspectives need to be considered.

A higher level of integrating value judgements with clinical evidence is scoring the outcomes used in an MCDA. Because the treatment-level results are likely measured on different scales (ex. proportions (0-1) vs. means (0-30)), they need to be mapped to a common scale of 0-100. This is called scoring. Scoring requires a value function that visually show the expected linear or curvilinear relationship between the original measurement scale and common scale. (see Figure \_\_\_).

**The Need For Visual Enhancements**

The authors evaluated X graphical displays to according to the criteria in Table 1 below. Given benefit-risk assessment is a transparent, decision-making process, facilitating open team discussions is essential. Last, but not least is the ability to compare estimands. Estimands are difficult but important to understand, and will be visualized in Section x.

|  |  |  |
| --- | --- | --- |
| **Table X. Criteria For Making Enhancements** | | |
|  | **Criteria** | **Description** |
| **1** | **Facilitating team discussions** | Incorporate values of different stakeholders |
| **2** | **Increasing transparency in decision-making process** | Defining success |
| **3** | **Incorporatng explicit decision rules** | Defining various thresholds, including cut-points, minimum acceptable risk, maximum acceptable benefit, and benefit-risk threshold curves. Thresholds for MCID, MAB and MAR provide reference points for pass/fail decisions. |
| **4** | **Comparing Apples to Apples** | Weighting and rescaling require elicitation of value judgments. Elicitation from stakeholders, such as physicians and patients can be conducted using Swing-weighting to obtain relative weights and KOL consensus for creating value functions for rescaling.  Benefits and risks on the same graph  Relative weights – line for B=R  Scaling of different metrics (means, counts, proportions, rates, time to event)  Same outcome on binary, ordinal and continuous scale |
| **5** | **Interpreting uncertainty** | Tipping points  Sensitivity analysis  Confidence intervals and regions |
| **6** | **Achieving consistency across assessment levels** (Treatment vs. Patient) |  |
| **7** | **Quantify impact of evidence & value judgement on decision-making** | Patient & physician preference data  Identify outcomes that most impact decision-making  Excluding correlated outcomes  Excluding unimportant outcomes |
| **8** | **Connecting visual assessments and statistical assessments** |  |
| **9** | **Specifying estimands** | Bridging simple estimates vs. estimands used in the trial |

Seven graphs were selected on their ability to aid decision making by integrating value judgement with clinical evidence.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Graphs Meeting Enhancement Criteria** | | | | | | | |
| **Criteria** | **Forest** | **Tradeoff** | **MCDA** | **Correlation** | **Ordinal** | **Predicted** | **Cumulative** |
| **Discussions** |  |  |  |  |  |  |  |
| **Transparency** |  |  |  |  |  |  |  |
| **Decision Rules** |  |  |  |  |  |  |  |
| **Apples to Apples** |  |  |  |  |  |  |  |
| **Uncertainty** |  |  |  |  |  |  |  |
| **Consistency** |  |  |  |  |  |  |  |
| **Impact** |  |  |  |  |  |  |  |
| **Statistical** |  |  |  |  |  |  |  |
| **Estimands** |  |  |  |  |  |  |  |

**Dot-Forest Plot**

This plot displays the treatment effects and treatment difference between an active drug and comparator (see Figure \_\_). The left panel indicates the size of a placebo effect. The right panel displays a treatment difference for each outcome. It is called a qualitative assessment because the outcomes are not quantitively integrated into a single composite outcome. A quantitative assessment requires quantitative weighting. The outcomes can be ranked by importance. The question is whether the active drug has a more favorable benefit-risk balance.

Reviewers scroll down looking for point estimates (95% CI) farthest to the left or right. This assumes all outcomes carry equal weight (wrong!). The plot can order the outcomes by their relative ranking or weight (but only within benefits and within weights). Reviewers also rely on the vertical reference line (ex. zero difference) as to whether the difference is statistically significant. That can be driven by sample size. What is missing is whether the difference is clinically meaningful. The SAP provides this for primary endpoints, but the team needs to elicit this for all outcomes if they hope to interpret the graph. The plot will foster team discussion. The BRAP can document how they reached consensus.

**Tradeoff Curve**

While the forest plot displays several benefit-risk outcomes for one active and comparator drug pair, the tradeoff plot displays one benefit-risk pair for several drugs or doses. There can be a tradeoff between a high risk / high benefit drug and a low risk / low benefit drug. The first drug might be preferred by severe patients and the second drug preferred by mild patients. The decision of whether to approve both or either drug will depend on which drugs’ point estimates and confidence intervals fall in the white acceptance region. The boundaries of acceptance region are set by the MAB, MAR, and tradeoff curve. The graph can display different drugs, different doses, and different estimands of the same drug.

**Correlogram**

The correlogram is a heatmap of color coded correlation coefficients. It estimates the associations between benefits and risks, using patient-level data, regardless of whether its binary, ordinal or continuous. If two benefits (or risks) are highly correlated, then there is a concern with double counting of outcomes. Which outcome do you prune? We can determine whether a benefit and risk are positively correlated, which might be due to increasing the dose. If two outcomes are highly correlated on the continuous scale but uncorrelated on the dichotomous scale, then perhaps information been lost.

Value judgment is used to categorize levels of correlation ranging from no, low, medium and high. Accompanying the correlogram is a scatter plot of the individual patient clinical values. The scatter plot can be enhanced by overlaying a regression line emphasizing the direction and magnitude of the correlation. The scatter plot can also have concentric circles around the mean focal point that emphasizes the variability in patient paired clinical values.

**Ordinal Composite Stacked Bar Chart**

Integrating value judgment can be used to reduce the dimensionality of benefits and risks measured on a continuous scale to a single composite outcome on the ordinal scale. At the patient level, thresholds are used to dichotomize a continuous or ordinal benefit to categories ‘Benefit’ and ‘No Benefit’, and dichotomizing risk as ‘Risk’ and ‘No Risk’. If there are several adverse events, such as nonfatal stroke, nonfatal myocardial infarction, and cardiovascular death, then it can be combined as adverse event or no adverse event. The next step is to combine the dichotomized outcomes into a single ordinal variable with 5 categories: Benefit/NoRisk, Benefit/Risk, NoBenefit/NoRisk, NoBenefit/Risk, and Withdrawal (Ref. Chuang-Stein). The patients’ composite outcomes are next visualized as a heat map, stacked and unstacked bar chart and two-by-two frequency table (see Figures \_\_\_). These graphs are not easy to interpret, particularly if the benefit and risk outcomes do not carry the same weight or consequence to the patient. A threshold for success, such as using an MAB and MAR for dichotomization, is still needed. Eliciting these value judgments will require lengthy team discussions to reach consensus, however, this prepares a sponsor for discussion with regulators.

{*Note: A threshold cannot not be drawn on a tacked bar charts, but it might work for an unstacked bar chart.}*

{*Note: The CIOMS report states: “One concern with the rank-based methods based on pairwise comparisons is that a decrement in a very important component could be offset by a large advantage in a component outcome of lesser importance.” “Partial credit analyses can be conducted to directly address the concerns with pairwise comparison methodologies.” What can we do with this?}*

**Predicted Scatter Plot**

A scatter plot of individual patient data displays the variability among patients and the correlation between a benefit and risk. It can also identify those patients with a favorable benefit-risk profile. Scatter plots require continuous patient-level data, but unfortunately risks (i.e. adverse events) are typically ‘yes/no’ data. One approach is to use logistic regression models to estimate each patient’s probability of a benefit and probability of a risk. With probabilities ranging from 0 to 100%, a scatter plot can be produced. To aid interpretation, a diagonal threshold can be use to delineate an acceptance region where the probability of benefit exceeds that of risk. Value judgments can be made to divide the area into four quadrants with one quadrant being acceptable benefit and acceptable risk. Additionally, a regression line can be displayed to determine if the probabilities of benefit and risk are positively correlated.

{*Note: Iris: “Underpinning all prescribing decisions is the individualized benefit-risk consideration. By collecting safety data in a more standardized way, utilizing accepted grading scales and focusing on key safety topics of interest we will be able to provide more quantitative and detailed benefit-risk assessments. This will continue to evolve as we begin to run trials with seamless access to EHRs and leverage emerging technology to create accessible real-time assessments of emerging benefits and safety over time.”}*

**Cumulative Excess No. Events**

*An important question is whether a favorable benefit-risk profile is sustainable across time. The cumulative plot in Figure \_\_, displays separate time trends for a benefit and risk. Ideally, the benefit trend will increase, while the risk trend will quickly plateau somewhere beneath the benefit trend. The plot can be enhanced with error bars around the trends, which is important when sample sizes are small and variability is high. The thresholds MAB and MAR can also be used to define an acceptance region. If the benefit and risk are not equally weighted, then weighted scores can be used, and the difference in cumulative excess number of events can be displayed as well. Changing the scale can be a substitute for a multiplicate weight. A linear, segmented or nonlinear regression line through the points might show a potential change in slope over time, and a potential lag of risk behind benefit.*

**Multi-Criterial Decision Analysis**

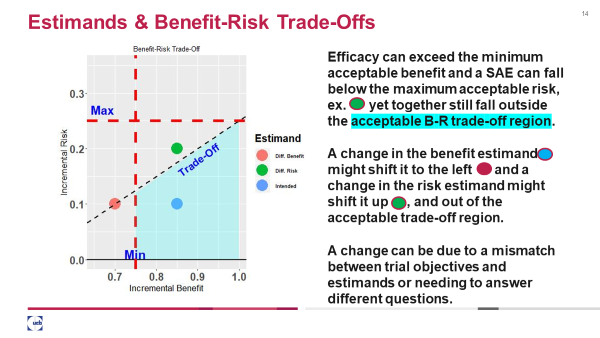
MCDA is a quantitative assessment that assigns relative weights to the outcomes (i.e. decision criteria), and converts the outcomes to a common 0-100 scale, if necessary. It is an excellent approach for internal sponsor decision-making, because it requires a project team to make explicit value judgments. MCDA does not make the decision, but rather aids decision making. If the results do not make sense, then a sensitivity analysis is conducted to see how a shift in weights or change in value functions might tip decision in the opposite direction. MCDA displays include …….

End of Manuscript

Below are my notes, example of graphs, random screen shots, etc.

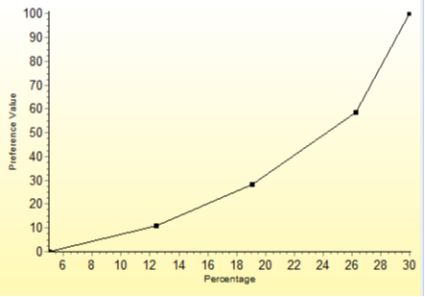
A picture containing text, screenshot, diagram, number

Description automatically generatedA picture containing text, diagram, line, plot

Description automatically generated

A screenshot of a computer

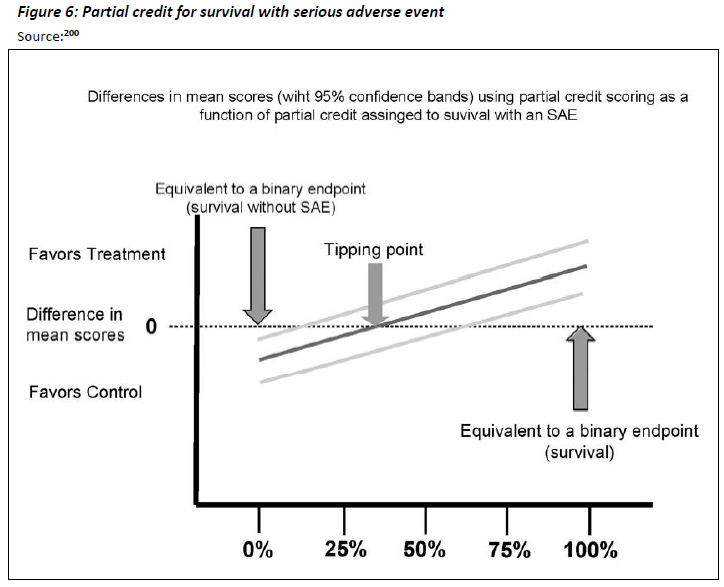
Description automatically generated with medium confidenceA picture containing screenshot, text

Description automatically generated

Value Function

A picture containing text, screenshot, diagram, number

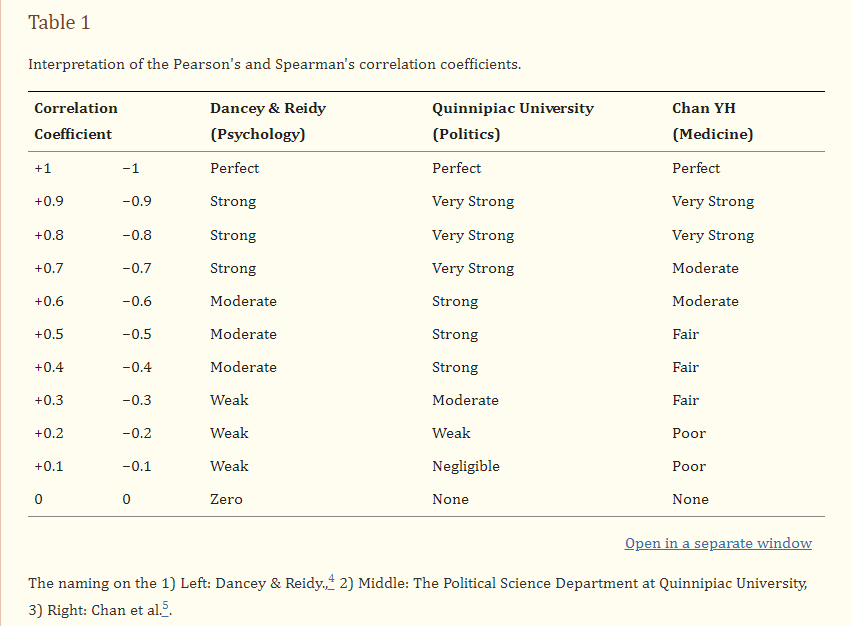
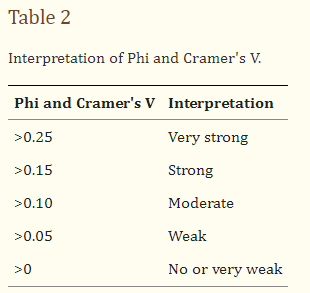
Description automatically generatedA picture containing text, diagram, line, plot

Description automatically generated A picture containing text, screenshot, parallel, design

Description automatically generatedA picture containing text, diagram, line, plot

Description automatically generated

Correlogram / Scatterplot with regression line

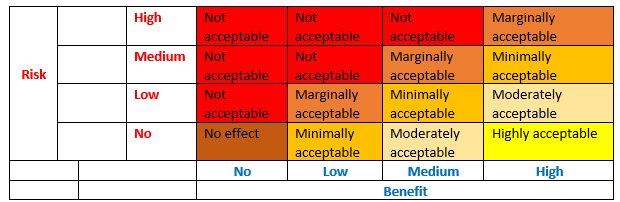
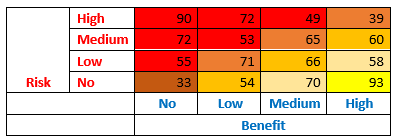
 

**Special Case: Ordinal Composite Stacked Bar Chart** – ***Chen would like to do this one.***

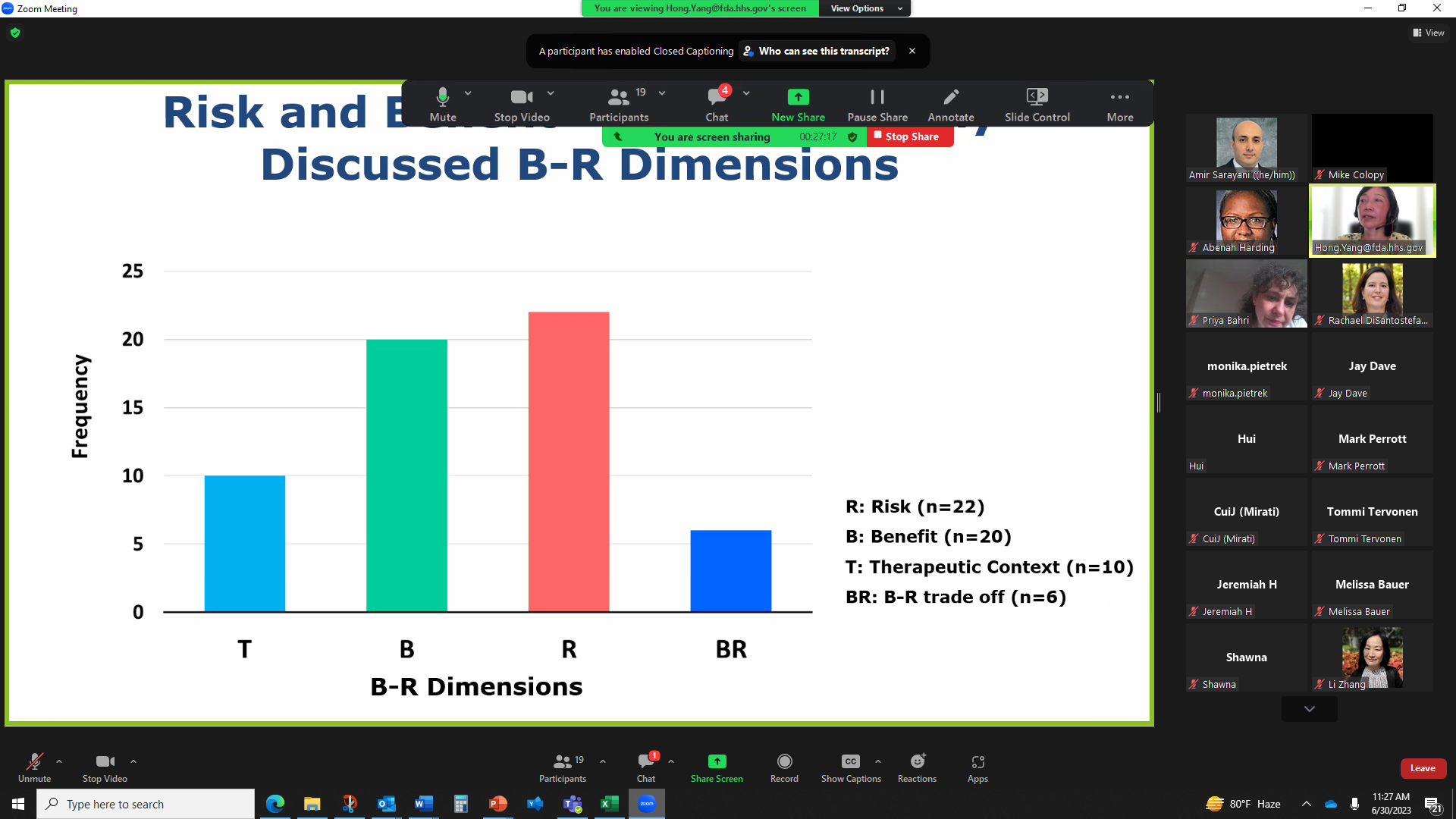
The project team should discuss how far they can collapse the categories for combinations of benefit-risk. This is similar to the standard cross-tabulation where you combine categories, particularly with few patients.

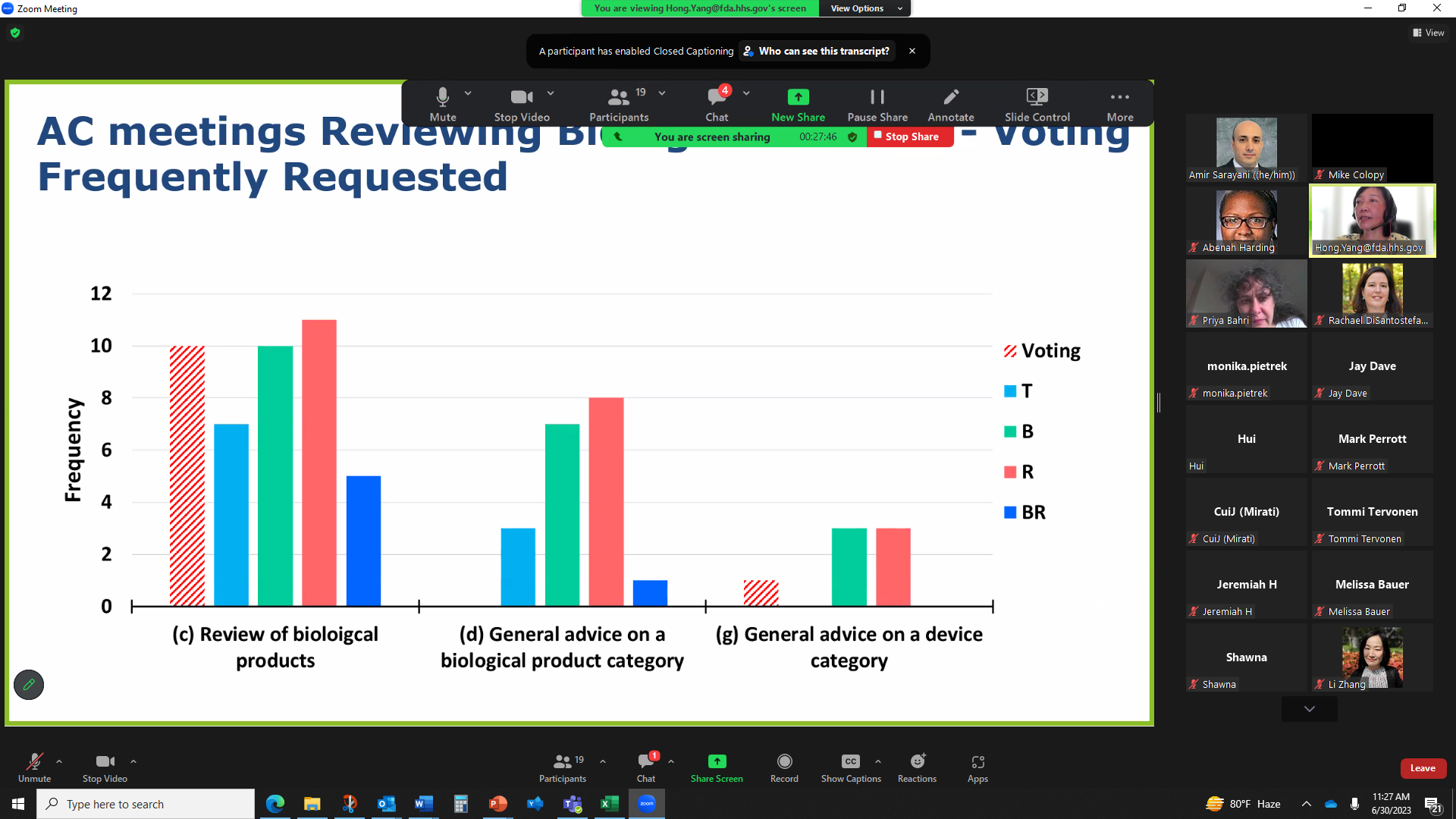
Epidemiologist like to get it down to a 2x2 table. It is not a trivial exercise.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **5 Different Strategies to Collapse 4x4 to 1x16 Table** | | | | | | | | | | | | | | | |
| **Not Acceptable** | | | | | | | | | | | **Acceptable** | **Not Acceptable** | **Acceptable** | | |
| N/Y | | N/N | N/Y | | Y/Y | N/N | Y/Y | N/N | Y/Y | N/N | Y/N | Y/Y | Y/N | | |
| **Risk > Benefit** | | | | | | No Effect | **Benefit=Risk** | | | **Benefit > Risk** | | | | | |
| **AE Only** | | | **Both** | | | **No Effect** | **Both** | | | **Benefit**  **Only** | **Both** | | **Benefit**  **Only** | **Both** | **Benefit Only** |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| No | No | No | Lo | Lo | Med | No | Hi | Lo | Med | Lo | Med | Hi | Med | Hi | Hi |
| Hi | Med | Lo | Hi | Med | Hi | No | Hi | Lo | Med | No | Lo | Med | No | Lo | No |
| 0.09 | 0.072 | 0.055 | 0.072 | 0.053 | 0.049 | 0.033 | 0.039 | 0.071 | 0.065 | 0.054 | 0.066 | 0.06 | 0.07 | 0.058 | 0.093 |

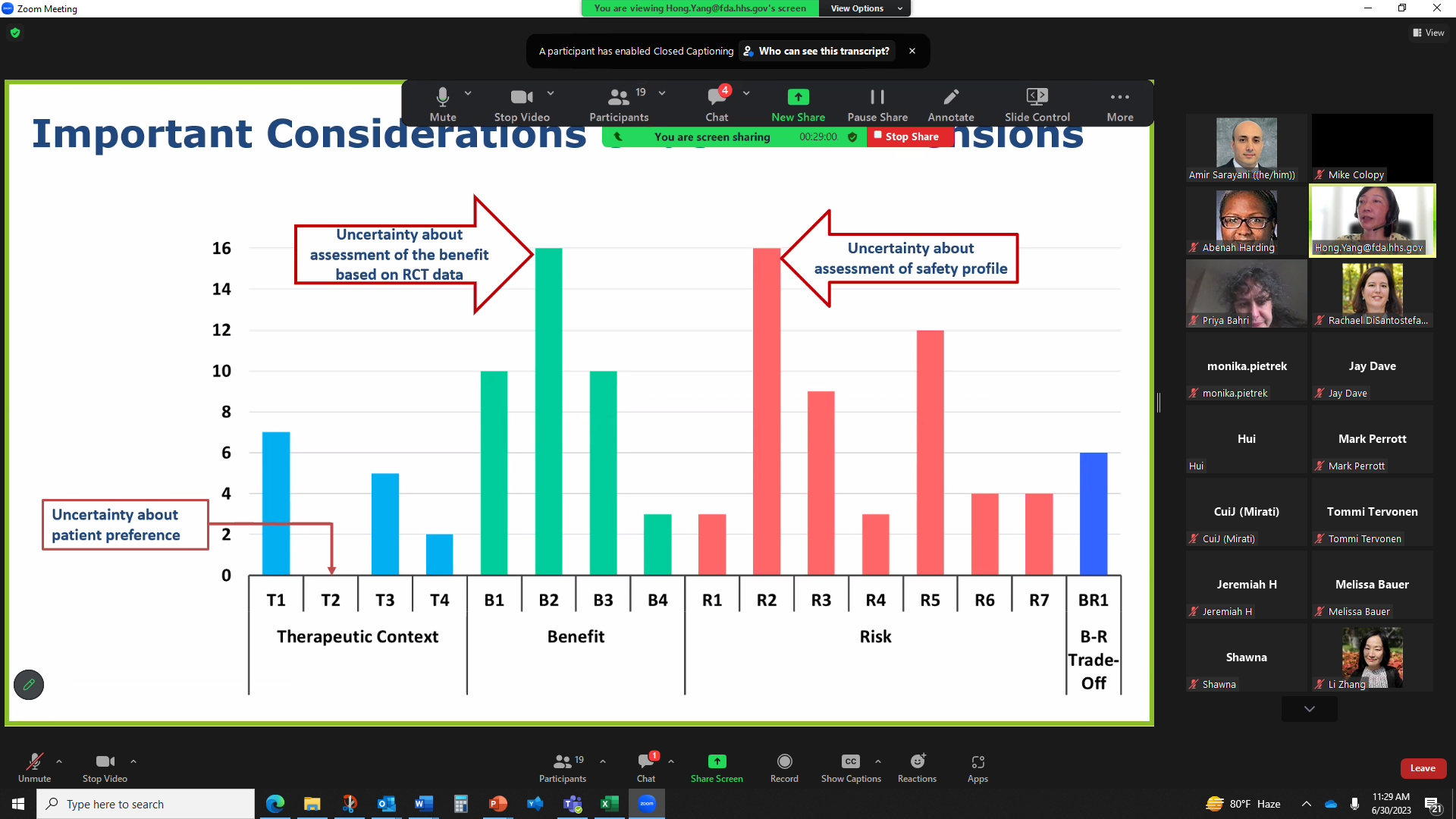
 

A

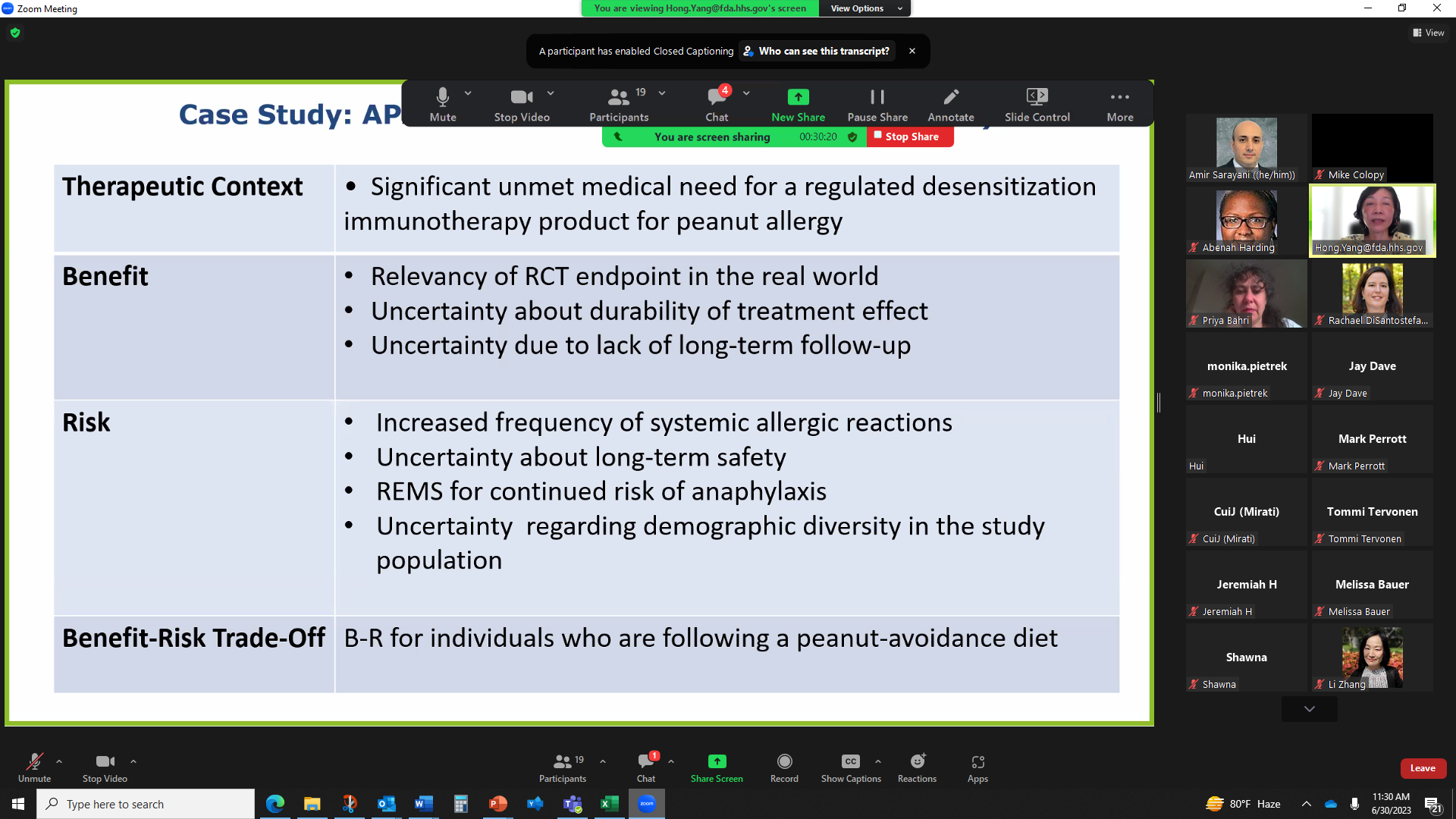


A

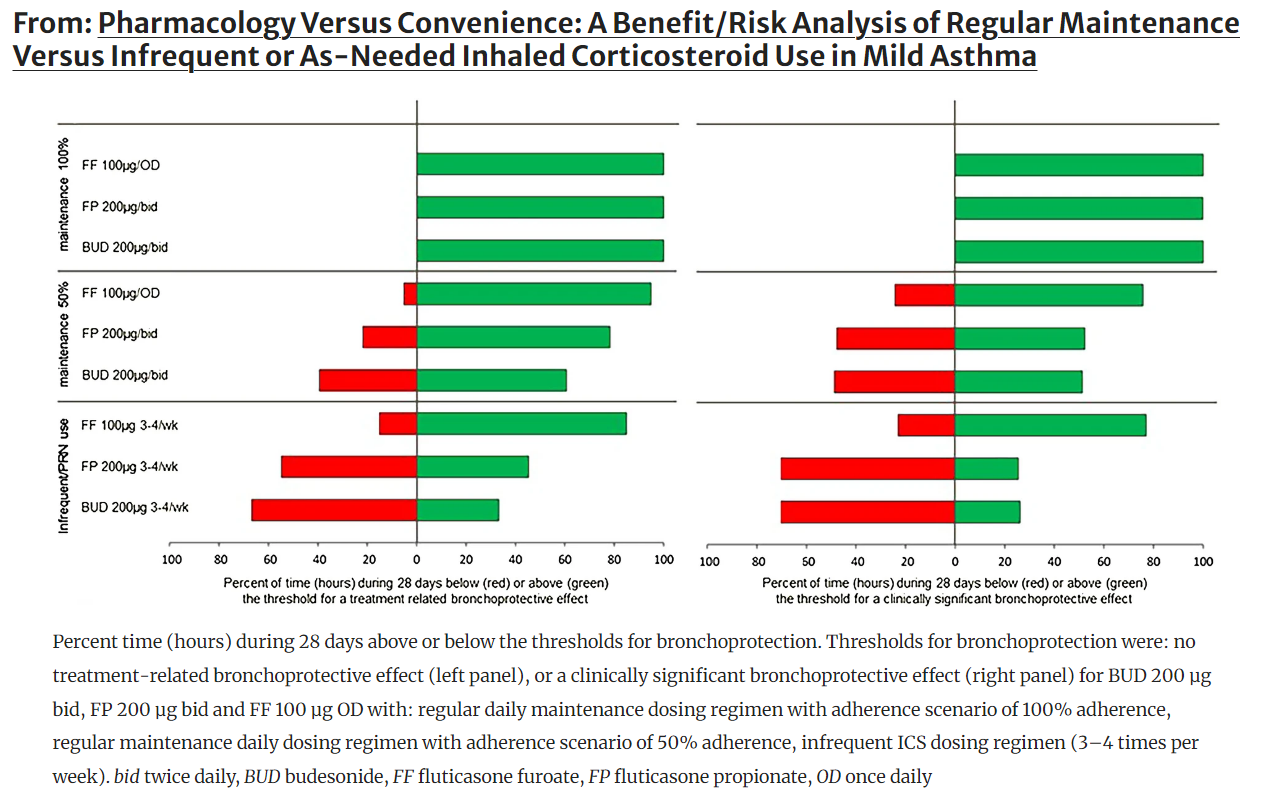
A

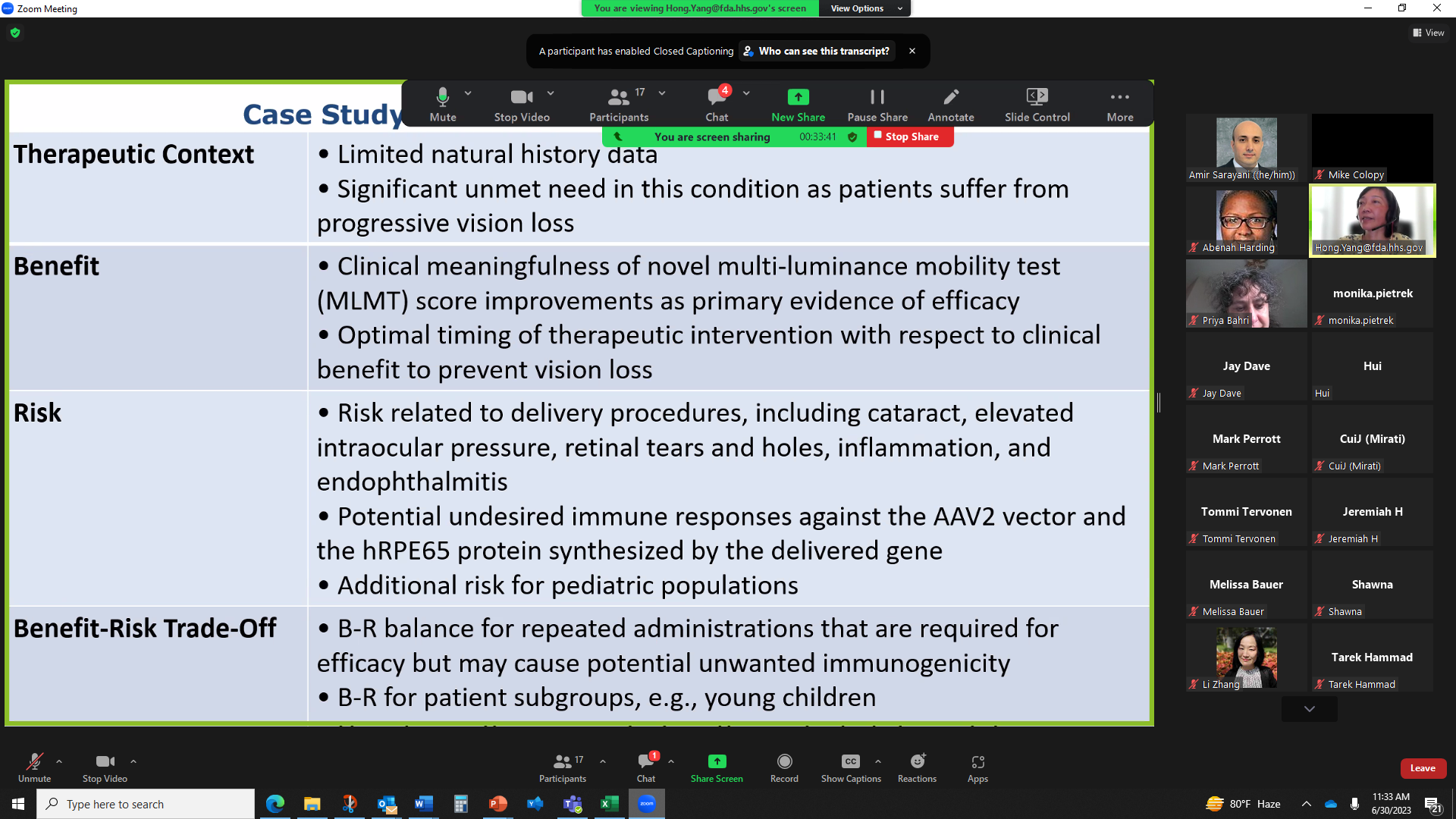


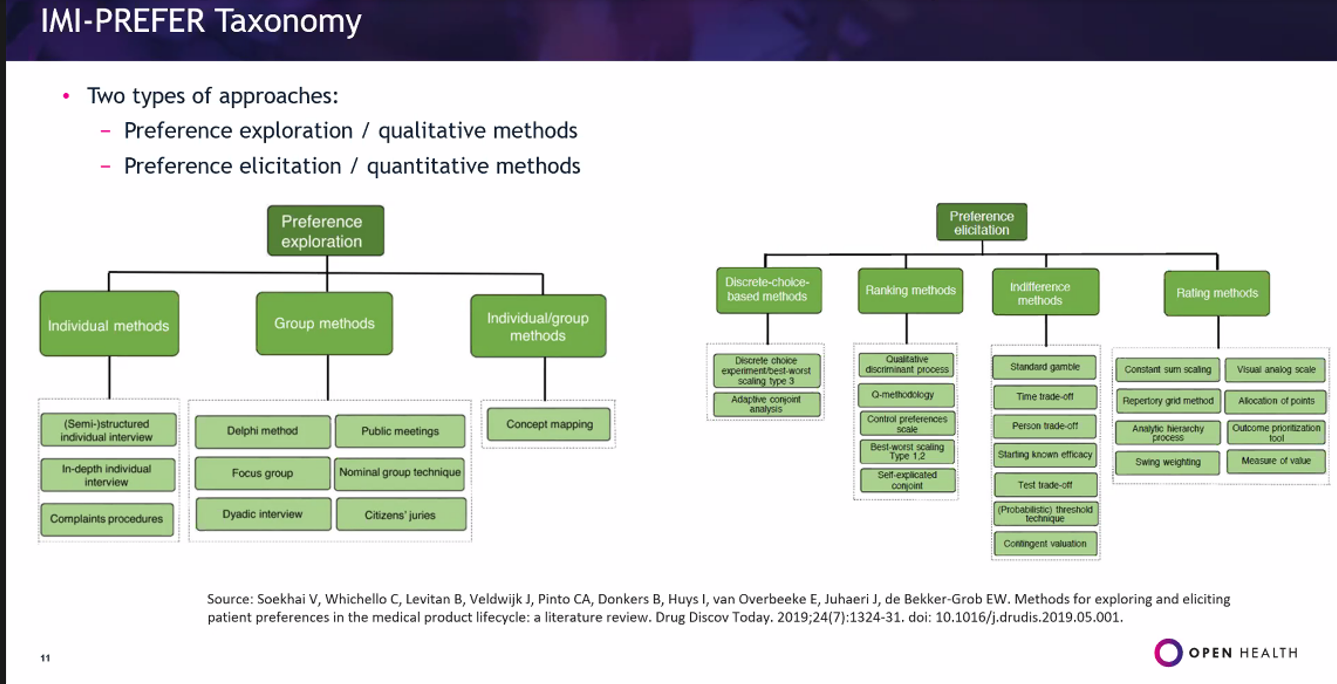
A

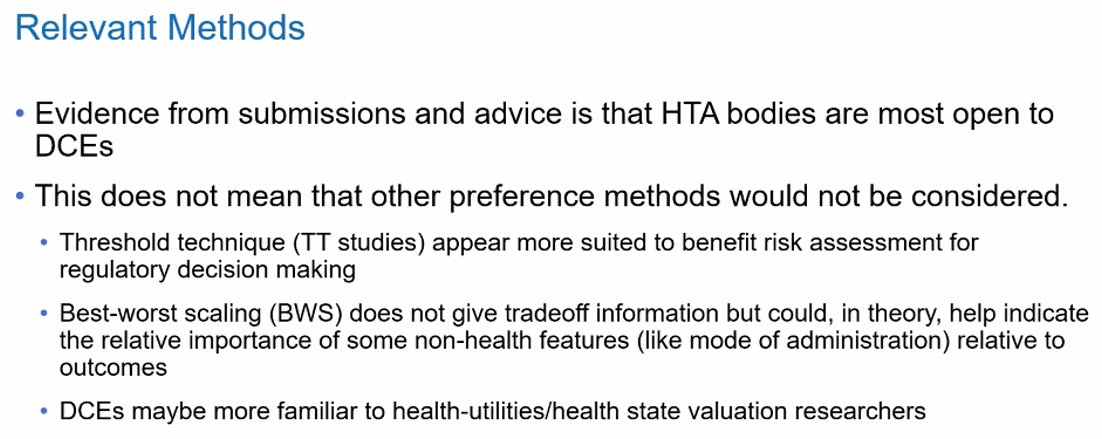


Mike: How do we use “convenience” in a BRA, when it is a factor, attribute, feature or characteristics of the therapy and not a variable? Zilu is a convenient self-administered SC once a day regime. To quantify it, we can assign it a weight or preference score (think MCDA), compared to a patch, oral dose or IV delivery. It can be a PRO score for satisfaction, or measure by days of adherence or compliance. Here is another example where the regimens differ by convenience, but the outcome is efficacy. There is a cut-point that is a value judgement. Look at how it is handled for medical devices.

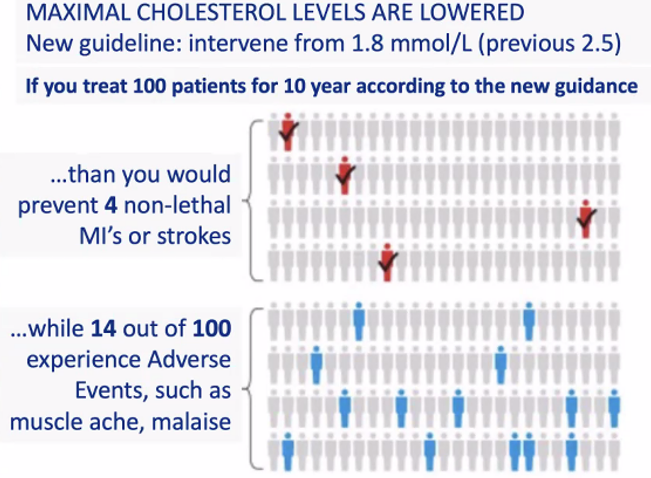




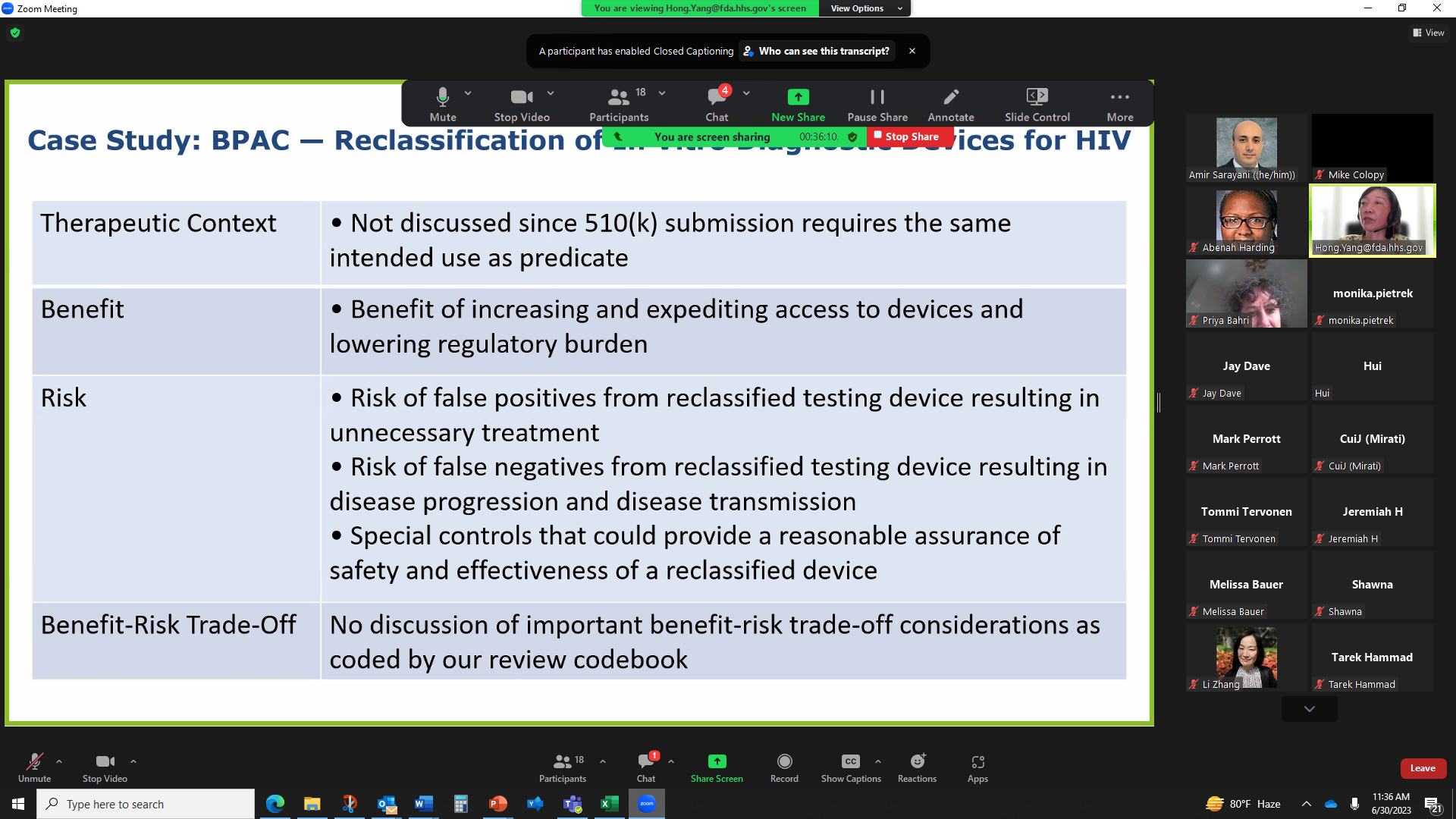




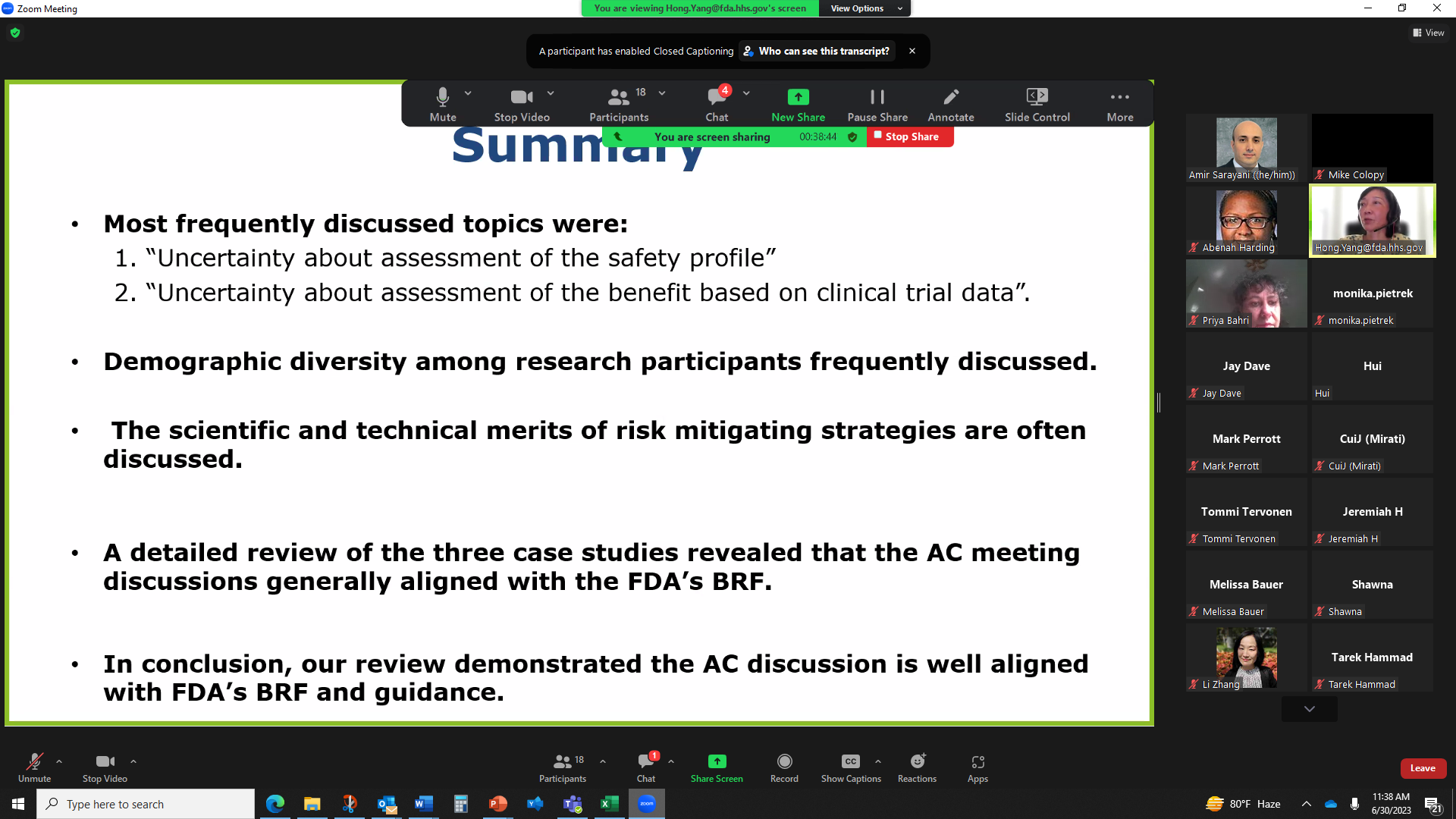
**It is interesting that risk perception is influenced by how random the figures appear below.**



ASA Webinar on Benefit-Risk. by Yang Hong



A



Appendix A. Methodologies that can be used to incorporate patient perspectives: MDIC Taxonomy

|  |  |
| --- | --- |
| **Group** | **Method** |
| Structured-weighting | Simple direct weighting  Ranking exercises  Swing weighting  Point allocation  Analytic hierarchy process  Outranking methods |
| Healh-state utility | Time tradeoff  Standard gamble |
| Stated-preference | Direct-assessment questions  Threshold technique  Conjoint analysis and discrete-choice experiments  Best-worst scaling exercises |
| Revealed-preference | Patient-preference trials  Direct questions in clinical trials  (Real market data – prescriptions, pharmacies, purchase data, GP records, ...) |

Source: Medical Device Innovation Consortium (MDIC) Patient centered benefit-risk project report. Available at: <https://mdc.org/wp-content/uploads/2018/05/MDIC_PCBR_Framework_Web.pdf>.



**DELETED TEXT**

There are scenarios where value judgment is of lesser importance, such as a therapy with a very favorable benefit-risk profile and minimal uncertainty in the data. More often decisions fall into a gray area where what is considered a favorable benefit-risk trade-offs depends on the stakeholder. For example, patients and physicians can differ on the relative importance of outcomes. Another scenario is a treatment for a rare disease with no Standard of Care. Value judgements may be needed in the place of hard data.