From Stories to Evidence

Patient Preferences in Benefit-Risk Evaluations

Juan Marcos Gonzalez, PhD
jm.gonzalez@duke.edu

“Patient preferences are critical in determining when a product’s benefits outweigh its risks… .”

-- Robert M. Califf (JAMA 2017)

“Treat data on patient preferences with the same level of scientific rigor as we would clinical data, and present it to regulators as such.”

--Bennett Levitan, Director Epidemiology Janssen R&D

What are preferences (in health)?

"Qualitative or quantitative statements of the relative desirability or acceptability of attributes that differ among alternative interventions."

Medical Device Innovation Consortium (PCBR Framework Report 2015)

Preference Information

Revealed Preferences

- Inferred from patients' actions

Stated preferences:

- Inferred from patients' statements

Revealed Preferences

- Elicited within real-world decision context
- Confounded with many factors that are not observed
- Current treatment alternatives are limited and may not cover the benefits or risks of interest

Preference Information

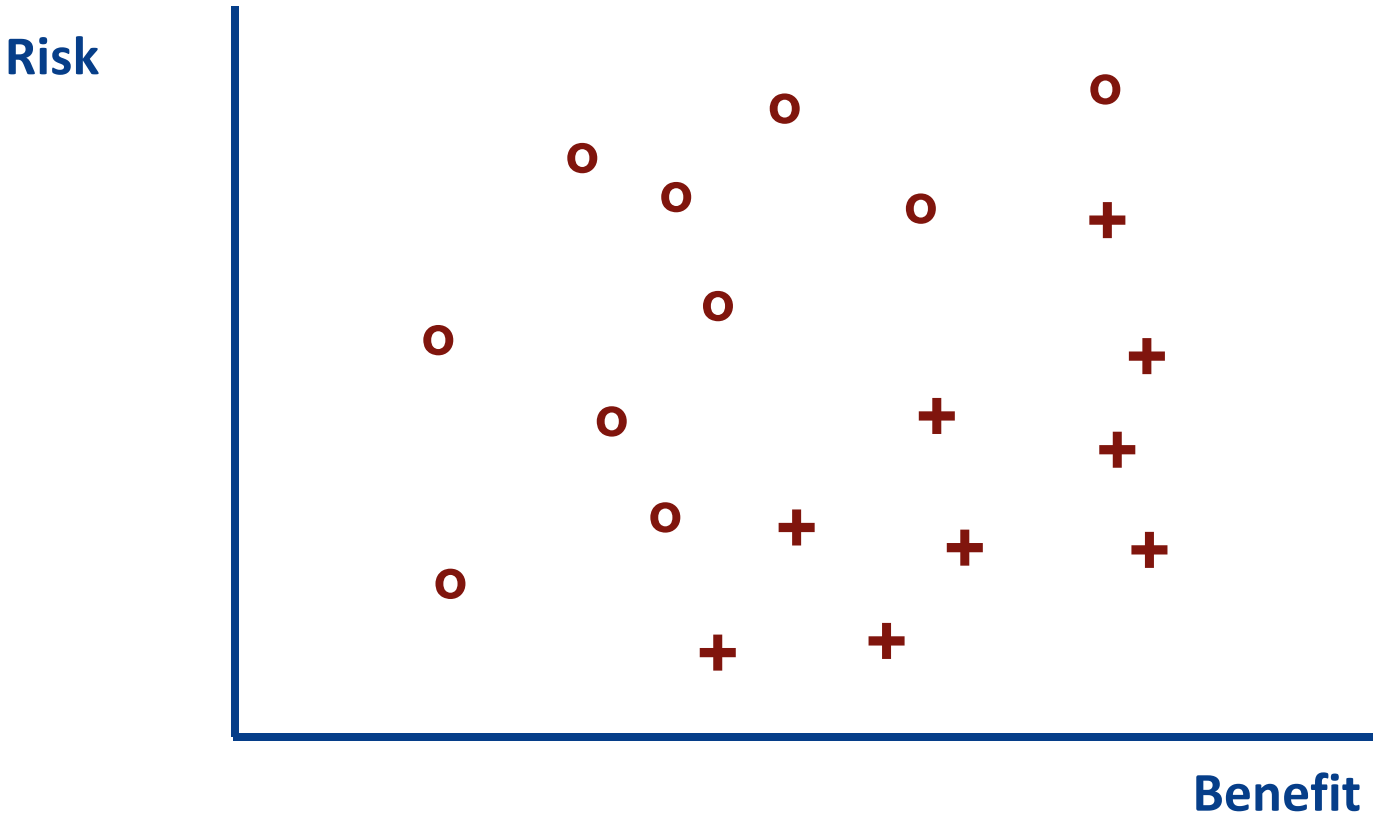
Stated Preferences

- Elicited under experimentally-controlled scenarios
- Alternatives can be new to respondents
- Decisions have no real-world consequences

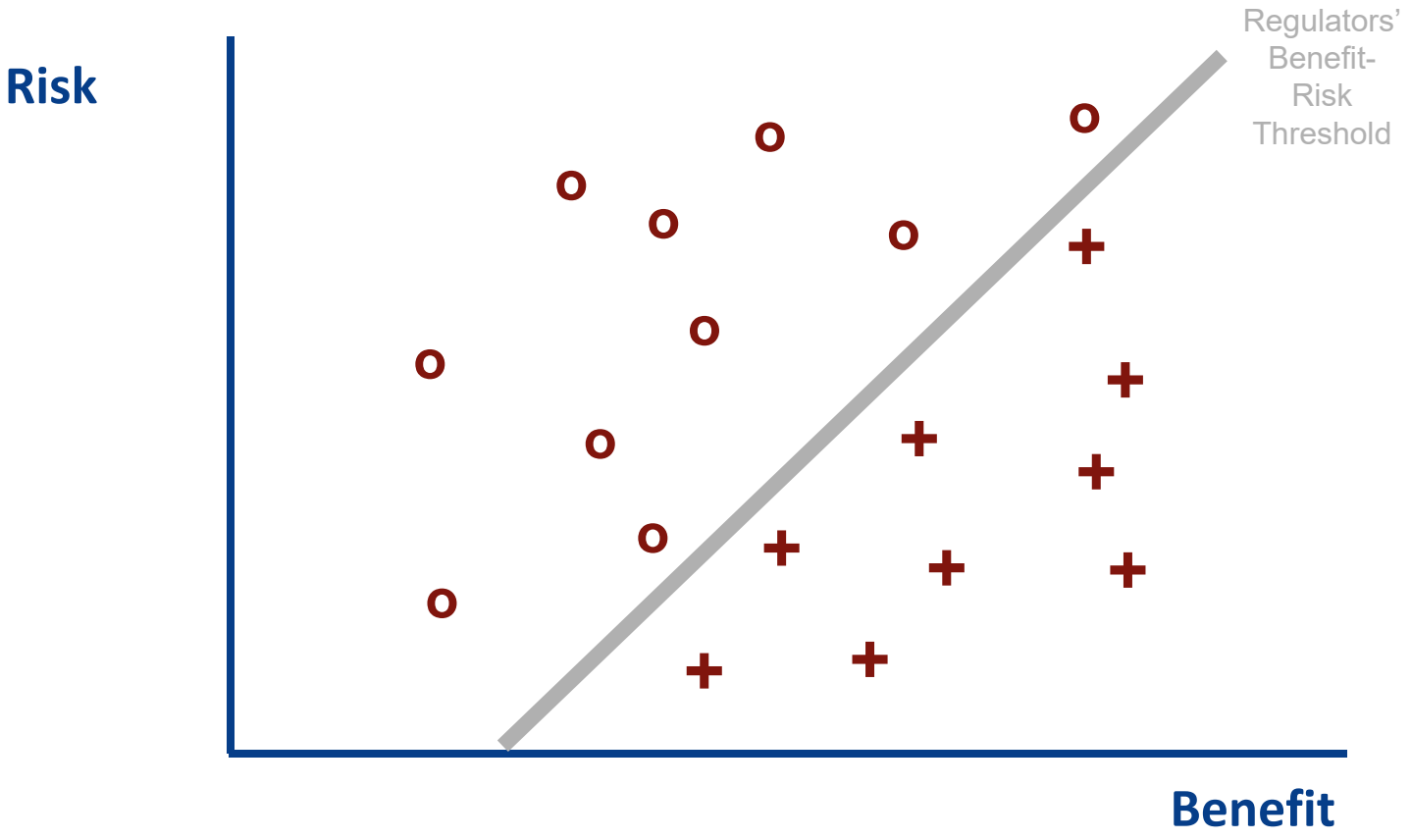
The value of things is defined by what people pay for them

The value of things is defined by what people would give up for them

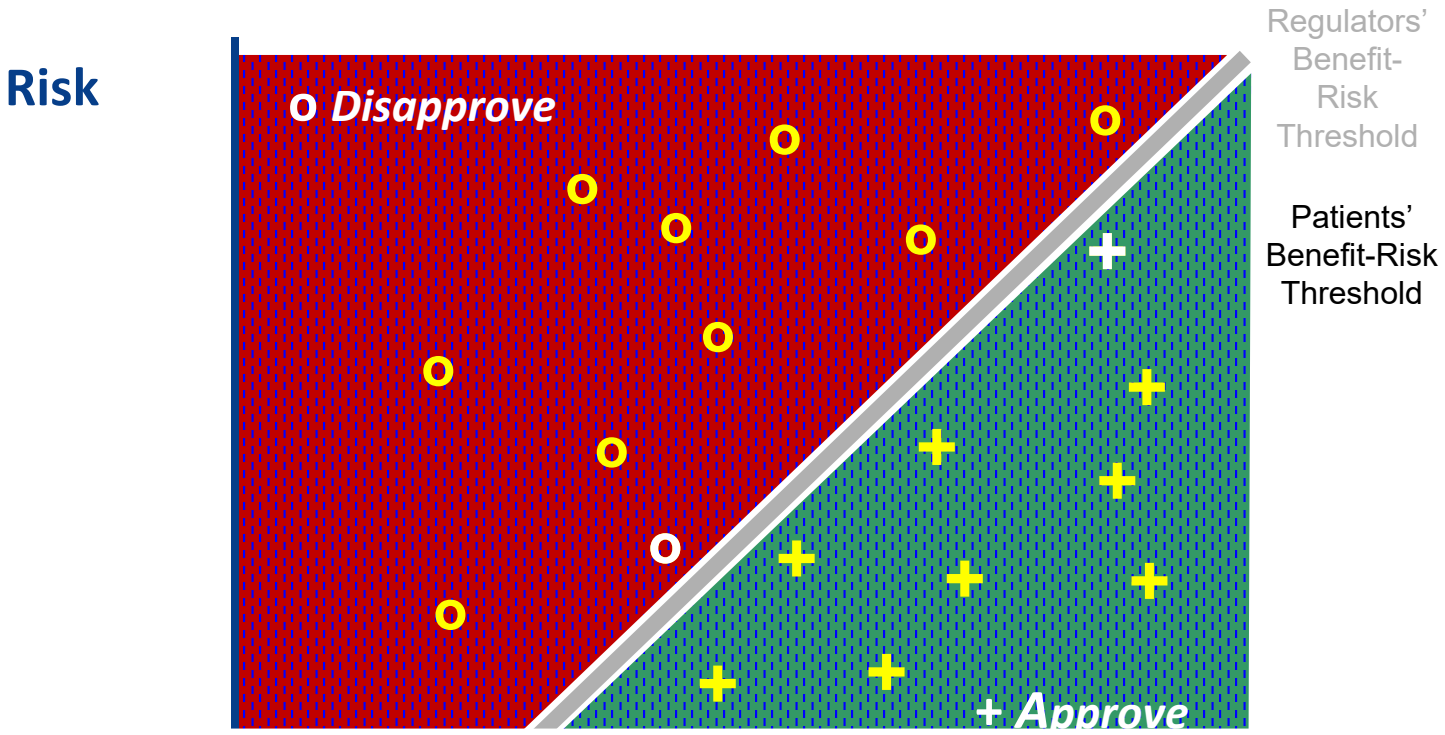
Preferences and Regulatory Decisions



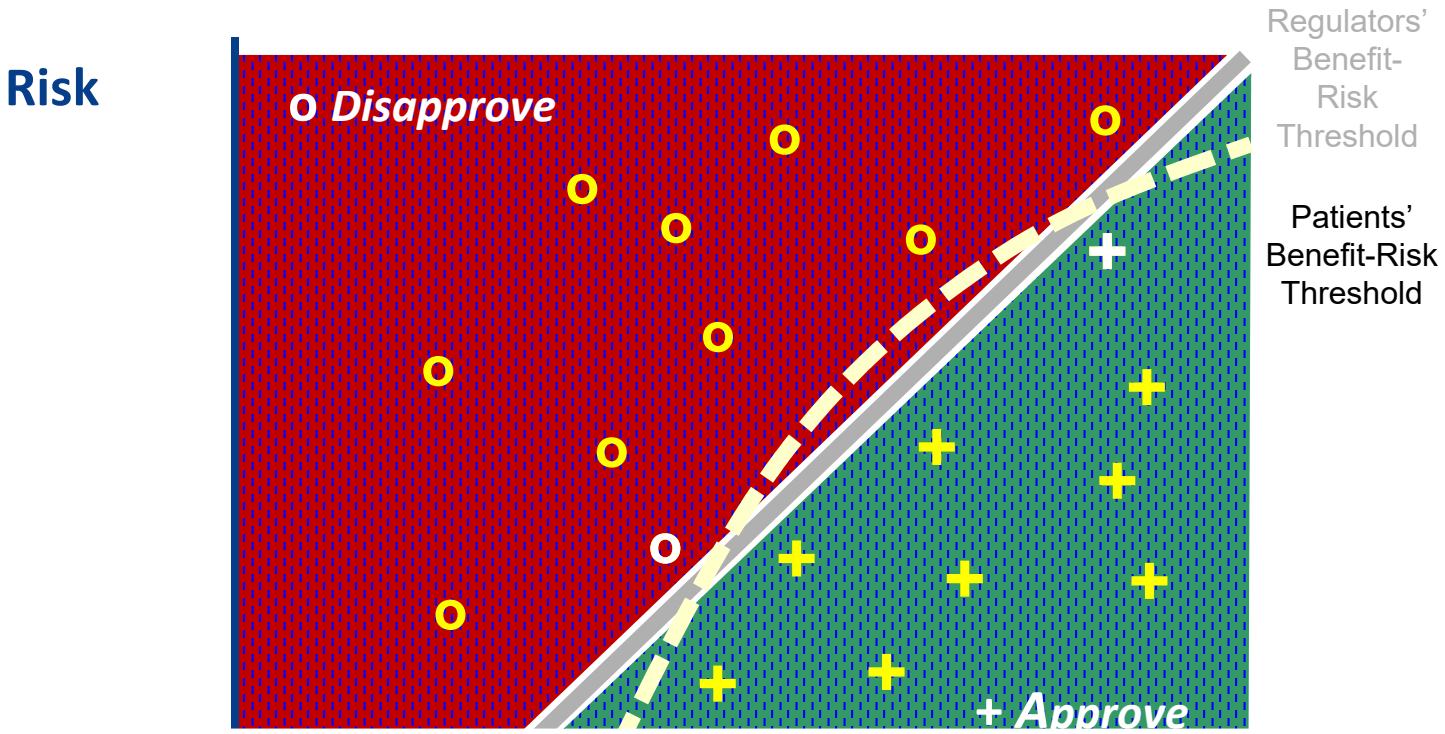
Preferences and Regulatory Decisions



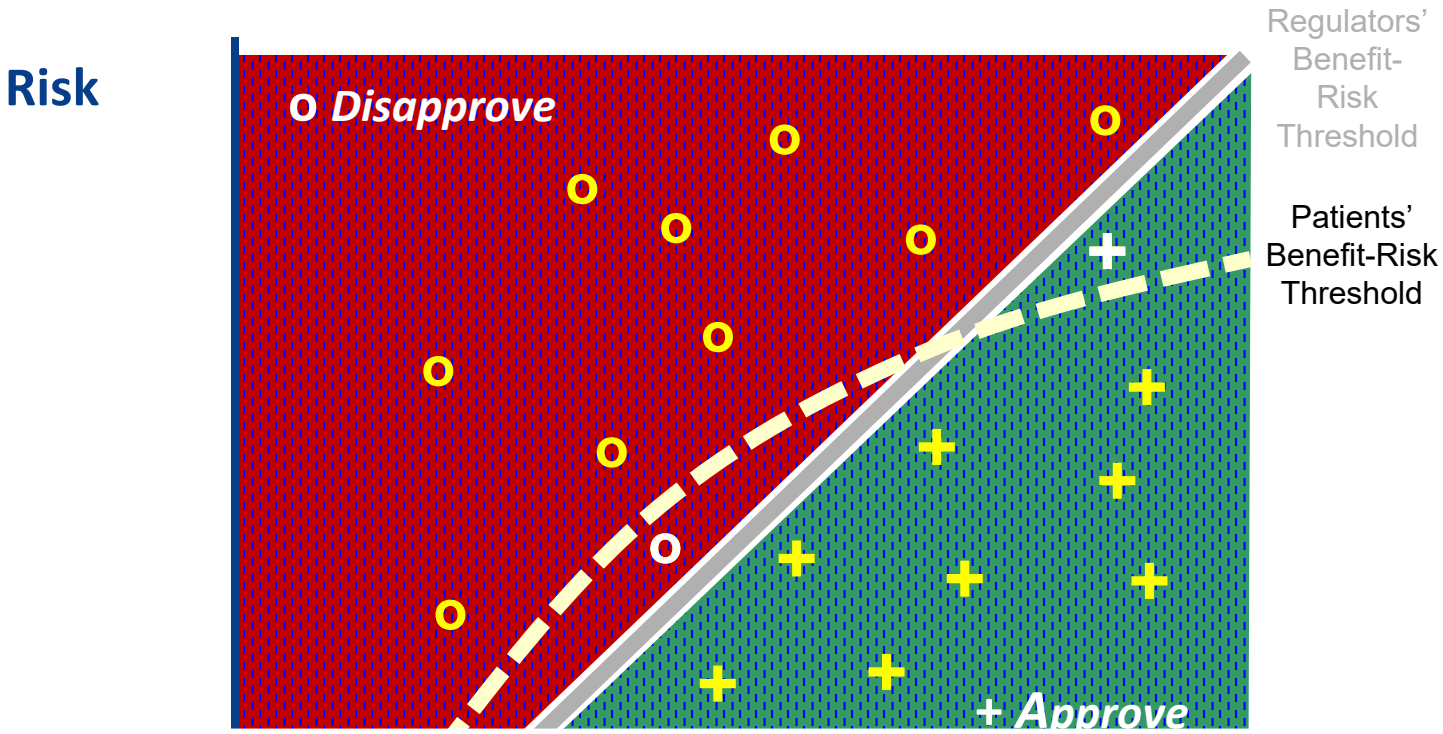
Preferences and Regulatory Decisions



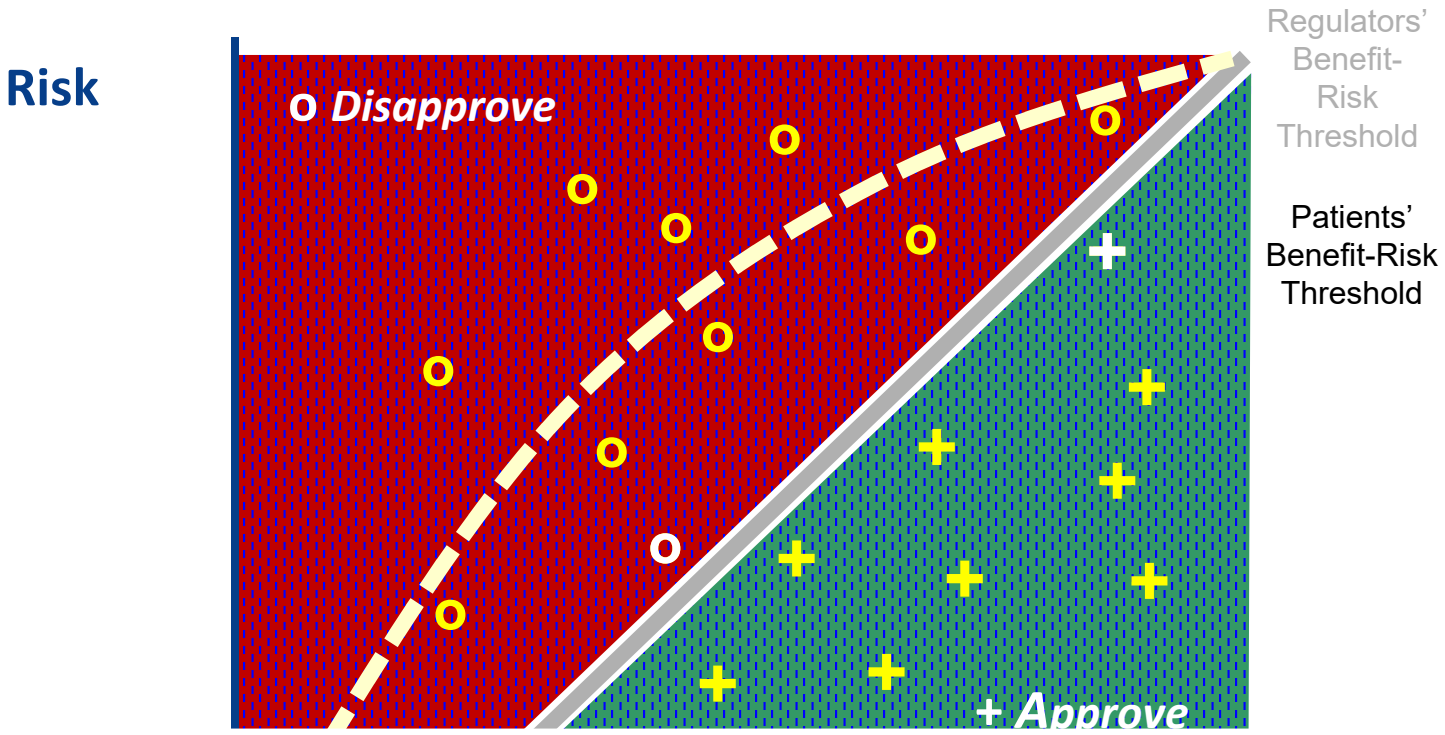
Preferences and Regulatory Decisions

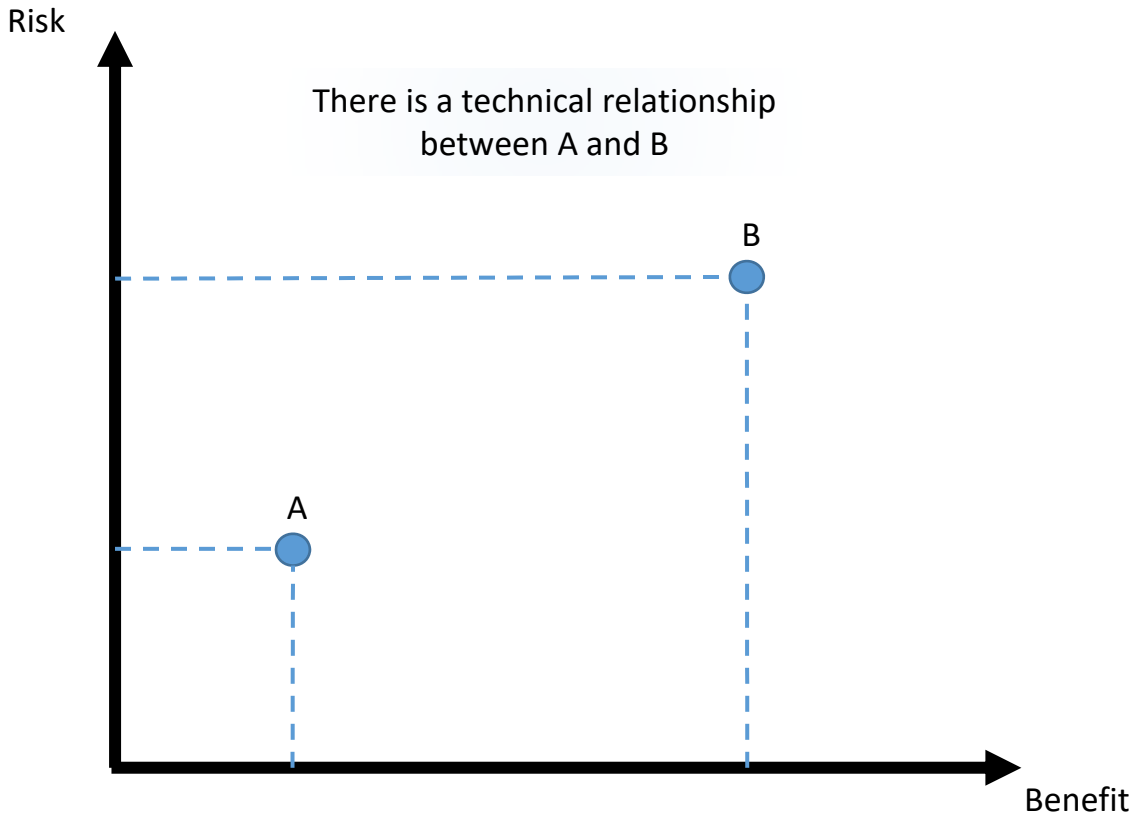


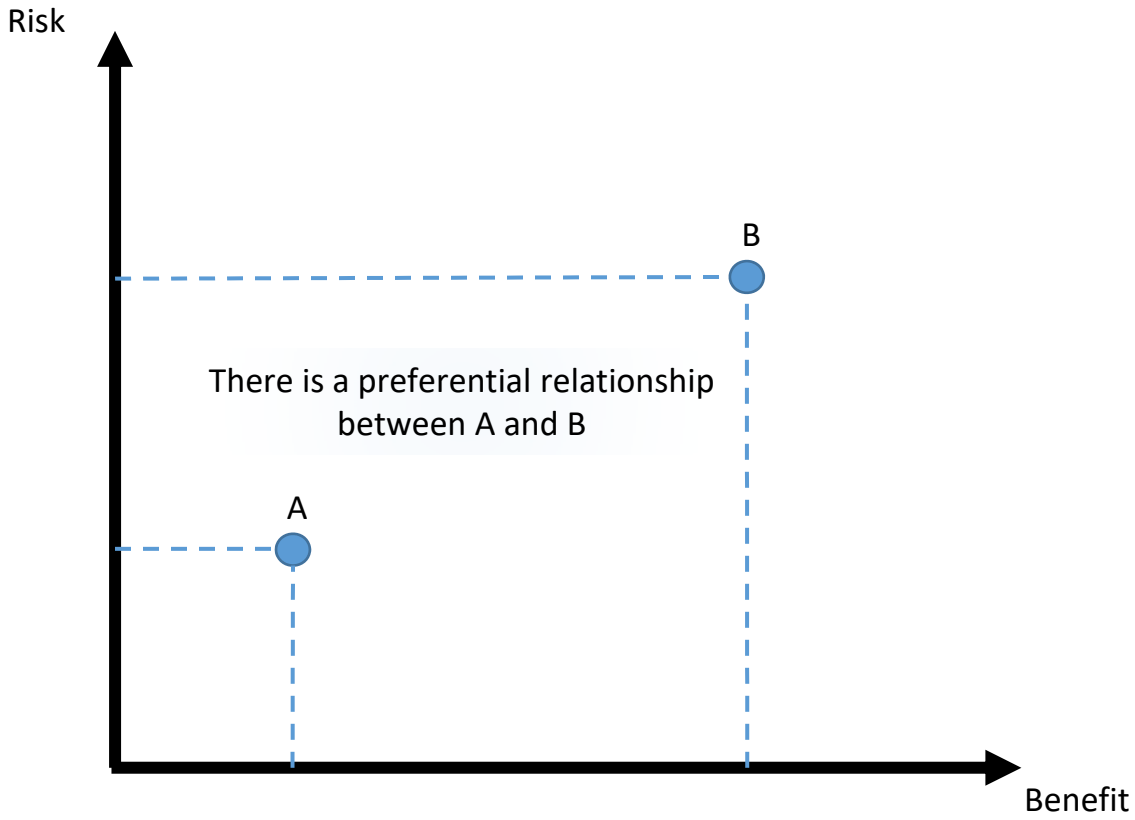
Preferences and Regulatory Decisions

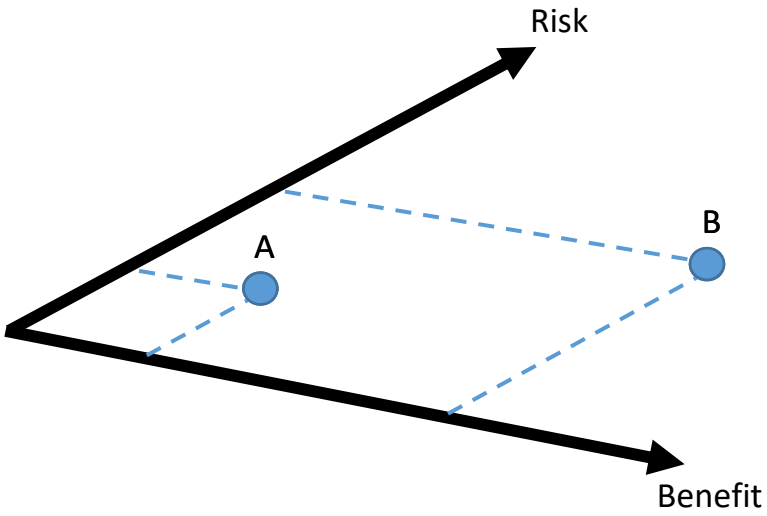


Preferences and Regulatory Decisions

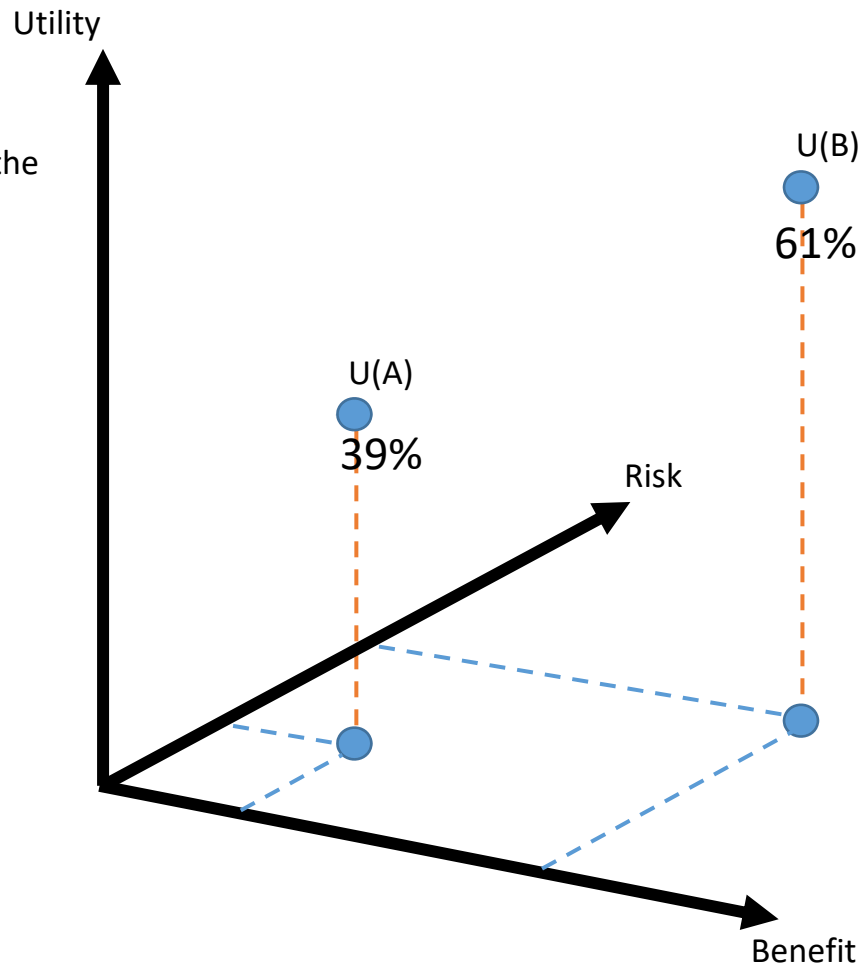



Risk
There is a technical relationship between A and B
B
A
Benefit


Risk
B
There is a preferential relationship between A and B
A
Benefit


Risk
A
B
Benefit

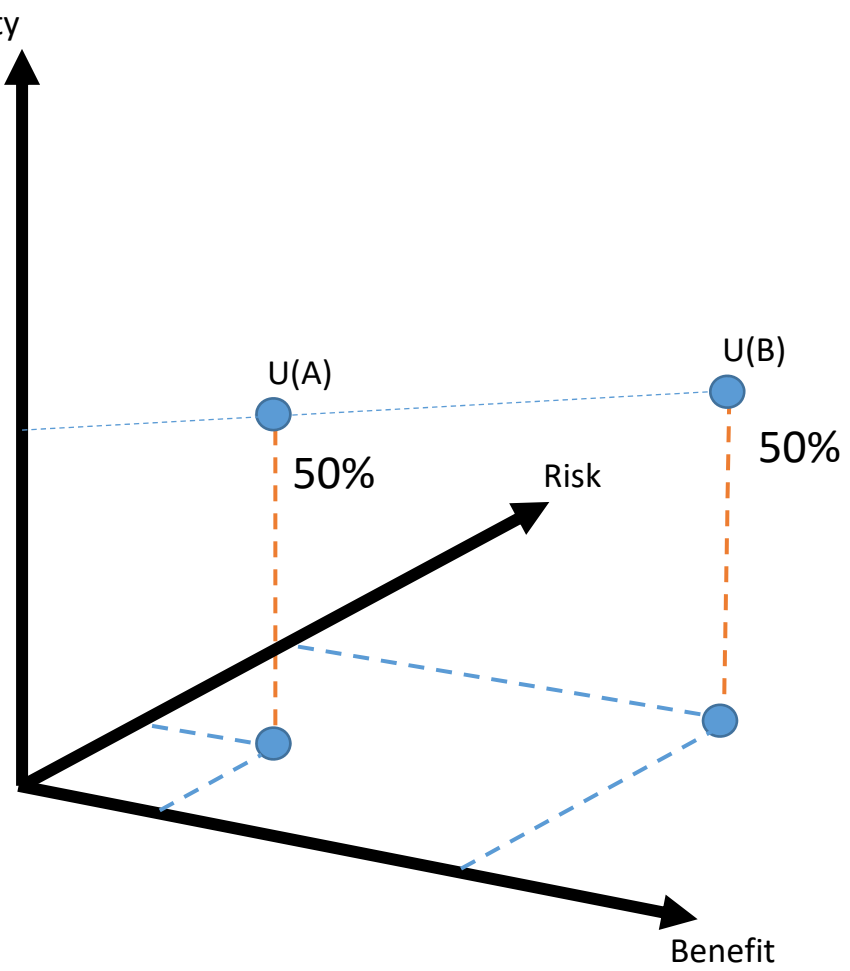
How do we determine the relative height of these bundles?



Statistical analysis turns proportions into utility values

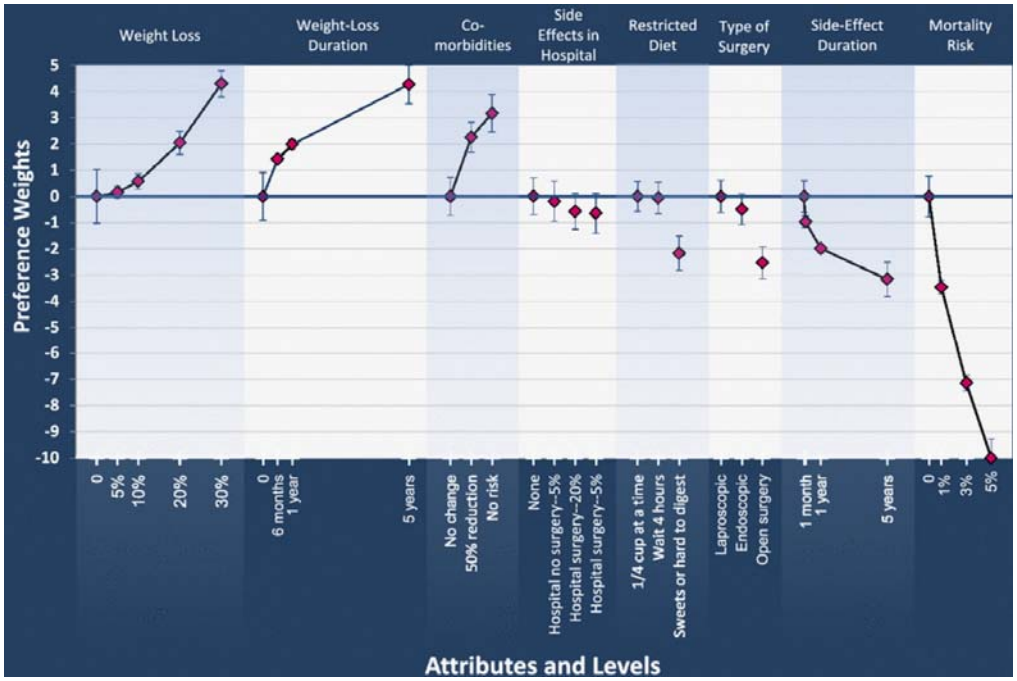
This is why preferences are relative

Utility equivalence is particularly interesting



Utility equivalence indicates a threshold

FDA Obesity Study



Ho et al. Surgical Endoscopy (2015)

FDA Obesity Study

Ho et al. Surgical Endoscopy (2015)

Stated-Preference Methods

- Eliciting stated-preference data

- Analyzing stated-preference data

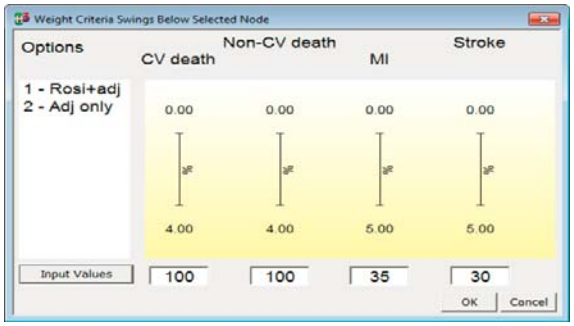
In summary

- Preferences are a key part of judgments about benefits and harms of treatments
- Evidence on preferences must be treated rigorously
- There are two types of preference data
  - Revealed preference data – Messy and not experimentally controlled
  - Stated preference data –Stylized with no direct consequences
- Stated preference methods rely on signals of relative desirability between outcomes or treatments
- Relative preference data allows evaluating utility-equivalence/thresholds between benefits and risk of harms for treatments (stated risk tolerance)

Eliciting Stated-Preference Data

Methods

Rating



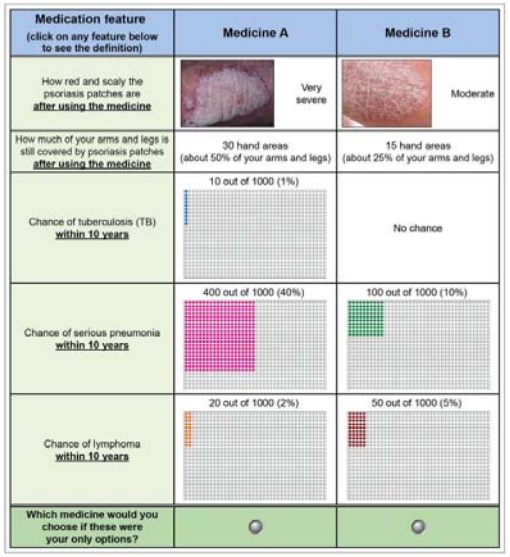
IMI Protect, 2015

Ranking

Most Important		Least Important
○	Washing and Drying Body Completely	○
○	Using the Toilet Without Accidents	○
○	Cooking a Light Meal	○
○	Taking Medicines	○
○	Staying at Home Alone	○

Zhang et al., 2015. JCM

Tradeoff



Gonzalez et al., 2016. BJD

Methods

Rating	Ranking	Tradeoff
Likert scales	Full-ranking exercise	Contingent Behavior
Point-allocation technique	Partial-ranking exercise (e.g., best-worst scaling)	Discrete-choice experiments
Swing weighting		Standard Gamble
Analytic Hierarchy Process		Time-tradeoff

Methods

Rating

- Elicit the intensity of preferences in a cardinal scale
- Provide direct preference weight values per respondent
- Must assume that tradeoff context does not affect rating
- Does not require an experimental design and often require simple statistical analysis tools

Ranking

- Elicit the relative utility of options
- Provide multiple signals for underlying preference weight value
- Must assume that tradeoff context does not affect ranking
- May require experimental design and can require complex statistical analysis

Tradeoff

- Elicit respondents' willingness to accept tradeoffs
- Provide multiple signals for underlying preference weight value
- Requires experimental design and can require complex statistical analysis

Rating Methods

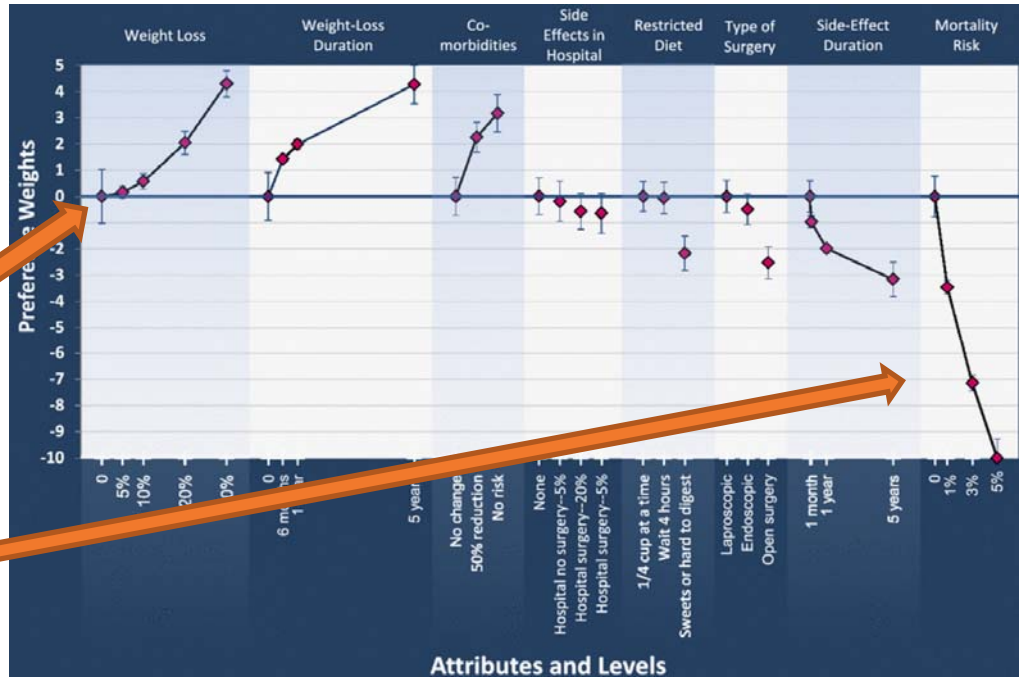
Please rate the following attribute changes between 0 (least important score) and 10 (most important score)

Weight loss from None to 30%

4.1

Mortality risk from 5% to None

10



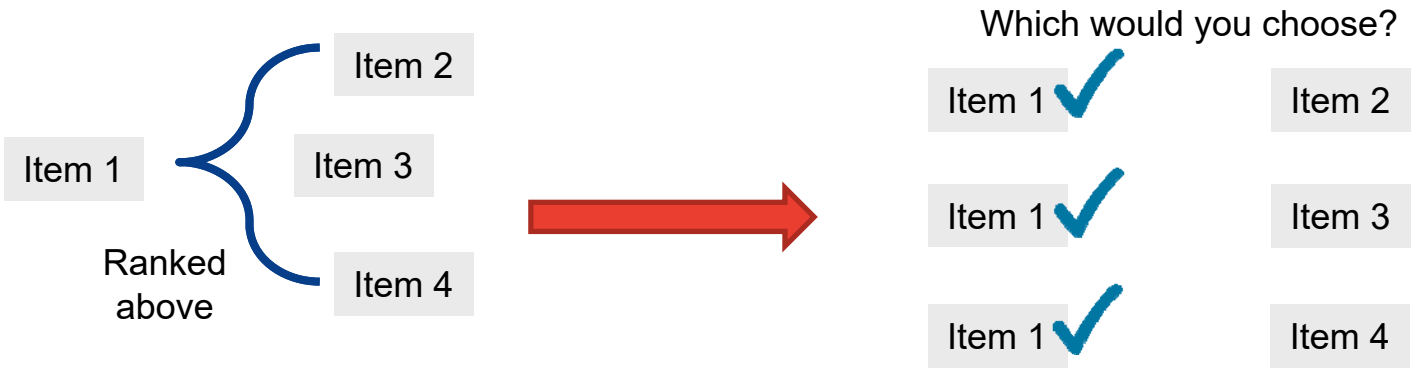
Ho et al. Surgical Endoscopy (2015)

Methods

Rating	Ranking	Tradeoff
• Elicit the intensity of preferences in a cardinal scale • Provide direct preference weight values per respondent • Must assume that tradeoff context does not affect rating • Does not require an experimental design and often require simple statistical analysis tools	• Elicit the relative utility of options • Provide multiple signals for underlying preference weight value • Must assume that tradeoff context does not affect ranking • May require experimental design and can require complex statistical analysis	• Elicit respondents' willingness to accept tradeoffs • Provide multiple signals for underlying preference weight value • Require experimental design and can require complex statistical analysis

Ranking Methods

- Provide a way to infer how many respondents would have chosen one alternative over the other

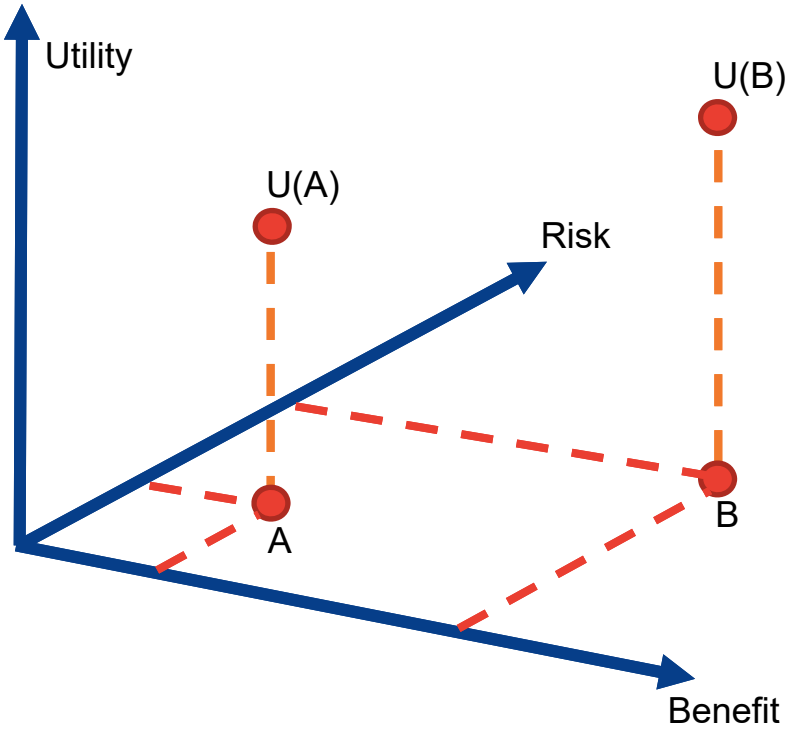


- Ranking is not necessarily a natural way to think about preferences in everyday behavior

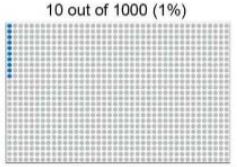
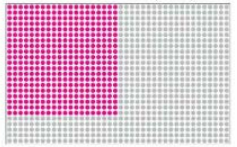
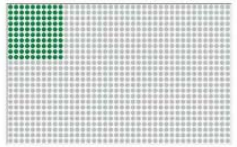
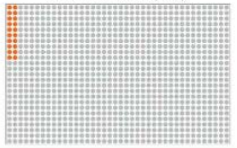
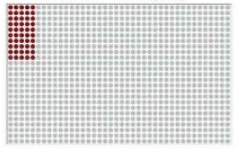
Methods

Rating	Ranking	Tradeoff
• Elicit the intensity of preferences in a cardinal scale • Provide direct preference weight values per respondent • Must assume that tradeoff context does not affect rating • Does not require an experimental design and often require simple statistical analysis tools	• Elicit the relative utility of options • Provide multiple signals for underlying preference weight value • Must assume that tradeoff context does not affect ranking • May require experimental design and can require complex statistical analysis	• Elicit respondents' willingness to accept tradeoffs • Provide multiple signals for underlying preference weight value • Require experimental design and can require complex statistical analysis

Tradeoff Methods

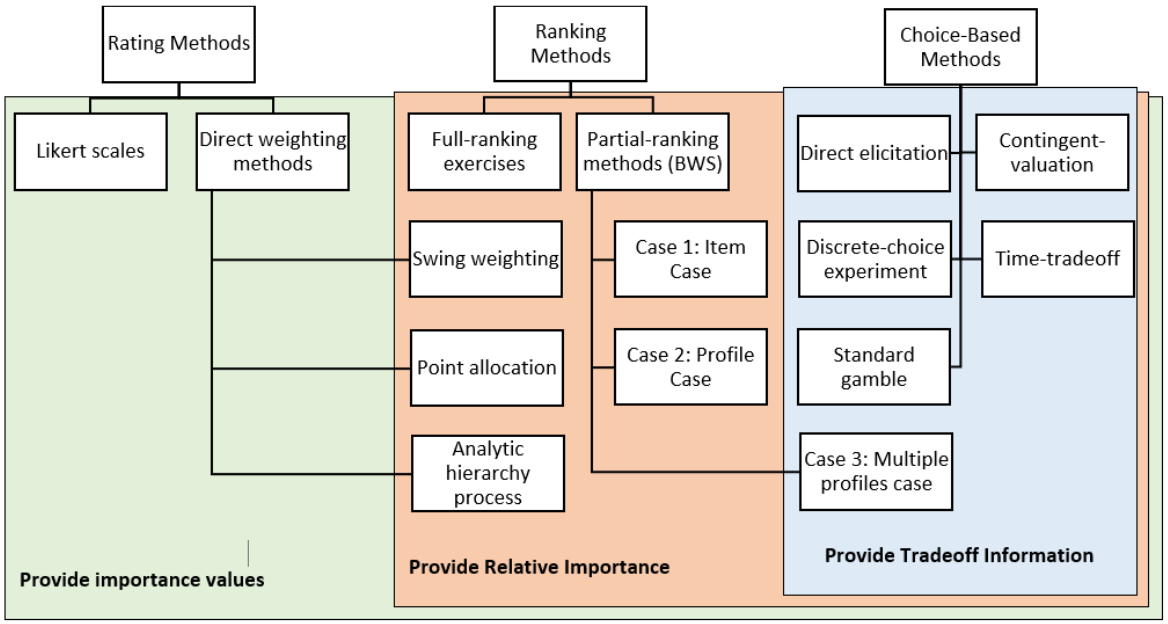


Gonzalez et al., 2016. BJD

Medication feature (click on any feature below to see the definition)	Medicine A	Medicine B
How red and scaly the psoriasis patches are **after using the medicine**	 Very severe	 Moderate
How much of your arms and legs is still covered by psoriasis patches **after using the medicine**	30 hand areas (about 50% of your arms and legs)	15 hand areas (about 25% of your arms and legs)
Chance of tuberculosis (TB) **within 10 years**	10 out of 1000 (1%) 	No chance
Chance of serious pneumonia **within 10 years**	400 out of 1000 (40%) 	100 out of 1000 (10%) 
Chance of lymphoma **within 10 years**	20 out of 1000 (2%) 	50 out of 1000 (5%) 
Which medicine would you choose if these were your only options?		

Gonzalez et al., BMJ 2016

What information is provided by preference-elicitation methods?

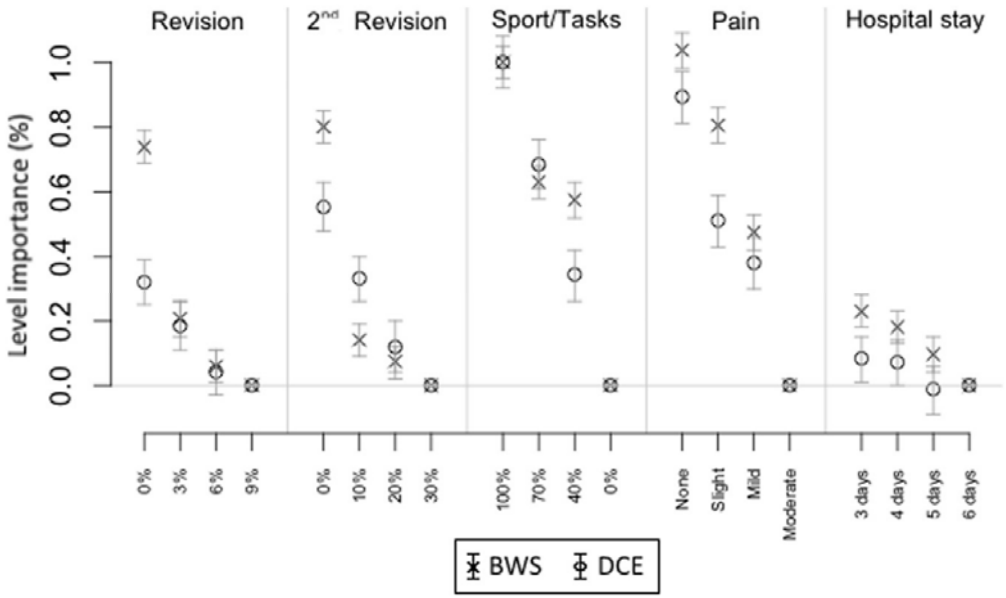


Importance values are defined to be relative to anchor categories. Possible to indirectly compare importance across attributes.

Importance is set for each attribute relative to other attributes.

Tradeoff information represents the rate at which one outcome is accepted for another.

Are Methods Interchangeable?



Van Dijk et al., 2016. Value in Health

Risk tolerance for hip-replacement surgery

Are Methods Interchangeable?

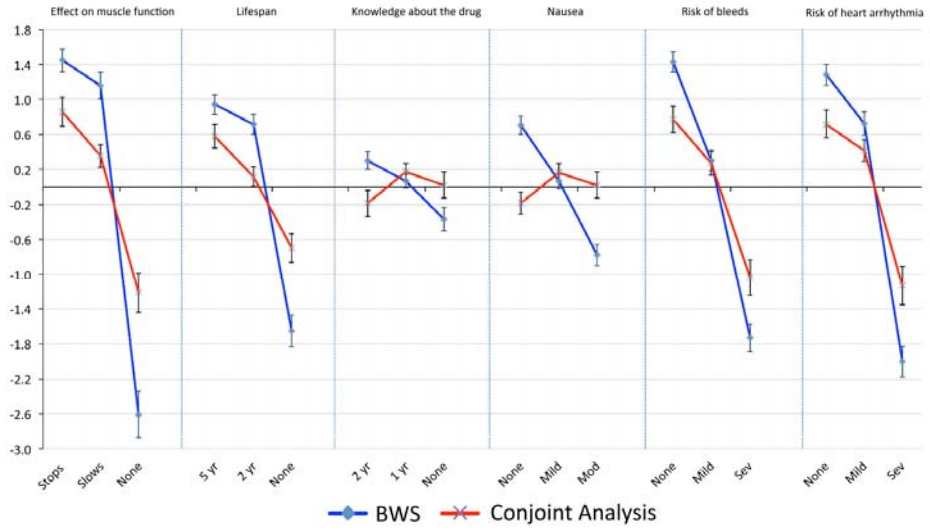
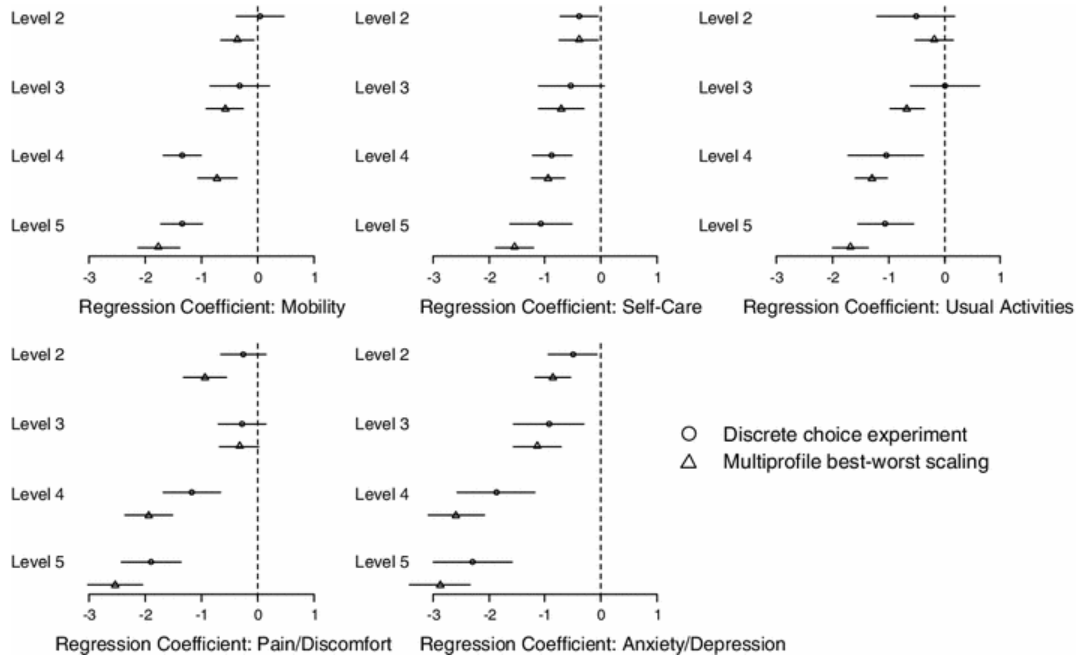


Fig. 2 Comparison of parameter estimates based on best-worst scaling and conjoint analysis. *BWS* best-worst scaling

Hollin et al, 2015. The Patient
Caregiver Preferences for Emerging Duchenne Muscular Dystrophy Treatments

Are Methods Interchangeable?



Xie et al., 2016. European J of Health Econ

Evaluating preferences for EQ-5D-5L

Are Methods Interchangeable?

Table 3. WTP estimate using dichotomous choice WTP technique

Attributes	Coefficient	*P*
Constant	1.26	0.001
Bid	−0.0003	0.001
Number of individuals	325	
Log-likelihood	−197.72	
Chi-Squared	55.06	
	(0.001)	
Individual predictions	69%	
Mean WTP ($0 \Rightarrow \infty$)	£4893	
95% confidence intervals	£4188–£6173	

Table 4. WTP estimate using the choice experiment

Attributes of service	Parameter	*p*	Marginal WTP for unit change in attribute ($\alpha j/\alpha 4$)	Current system[a]	WTP for current system[a]
Non-price attributes					
Attitudes of staff (α_1)	0.767	0.001	£852	0.8	£682
Continuity of care (α_2)	0.208	0.020	£231	0.7	£162
Time on waiting list (α_3)	−0.076	0.001	−£84	6 months	−£504
'Cost per attempt' (α_4)	−0.0009	0.001	—	1500	N/a
Chance of leaving the service with a child (α_5)	0.142	0.001	£157	27.9%	£4380
Follow-up (α_6)	0.288	0.006	£320	0.5	£160
Number of individuals	325				
Number of observations	3893				
Log-likelihood	−1150				
Chi-squared	27.36				
	(0.001)				
Individual predictions	79%				
Welfare measures					
Total WTP					£4880[b]
95% confidence intervals					£4532–£5284

[a]IVF service valued in the DC WTP question.

Ryan, 2004. Health Economics

Willingness to Pay for IVF

Are Methods Interchangeable?

Most Important Attribute-Direct Question on Post-Survey

	Conjoint analysis group (n = 50)	Rating / Ranking (n = 54)
Ability to reduce colorectal incidence and mortality	56%	61%
Discomfort	12%	7%
Nature of test	8%	6%
Frequency	12%	6%
Risk of major complications	2%	4%
Out of pocket costs	10%	17%

Unlabeled Test Preference

	Conjoint analysis (n = 50)	Rating/Ranking (n = 54)
FOBT-Based attributes and levels	26%	20%
Sigmoidoscopy-based attributes and levels	0%	0%
Colonoscopy-based attributes and levels	44%	39%
Radiologic test-based attributes and levels	26%	39%
No test	4%	2%

FOBT = fecal occult blood test

Pignone et al., 2012. Journal of General Internal Medicine

In summary

- 3 ways to elicit stated preferences
  - Rating
  - Ranking
  - Tradeoff
- Tradeoffs are closer to the construct of preferences in economics
  - May not be feasible to obtain tradeoff information
  - May need a different construct related to preferences
- Assumptions need to be made to turn rating and ranking data into tradeoff information
  - Whether the assumptions hold is an empirical question
  - Evidence suggests that sometimes these assumptions are reasonable

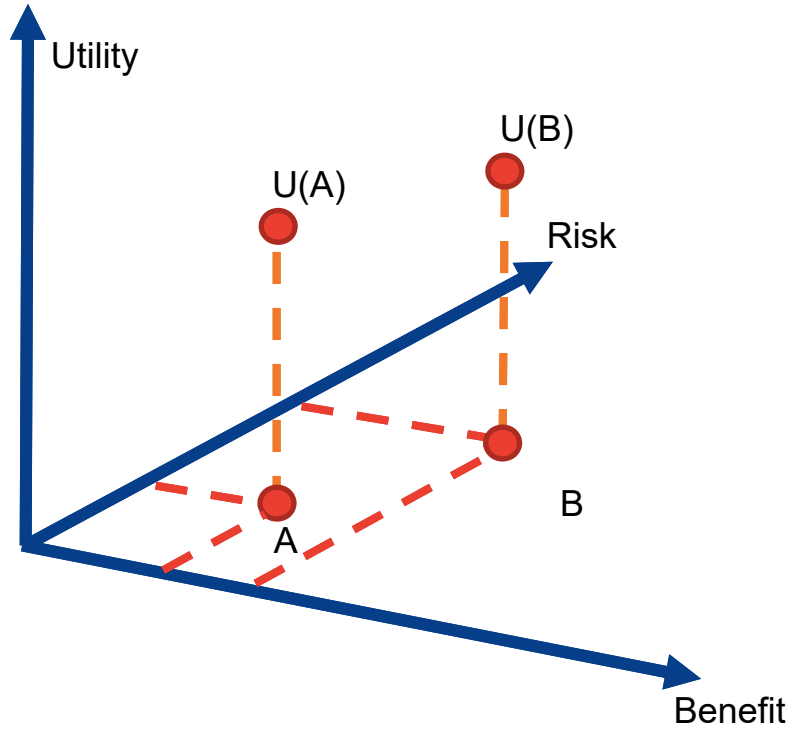
Analysis of Stated-Preference Data

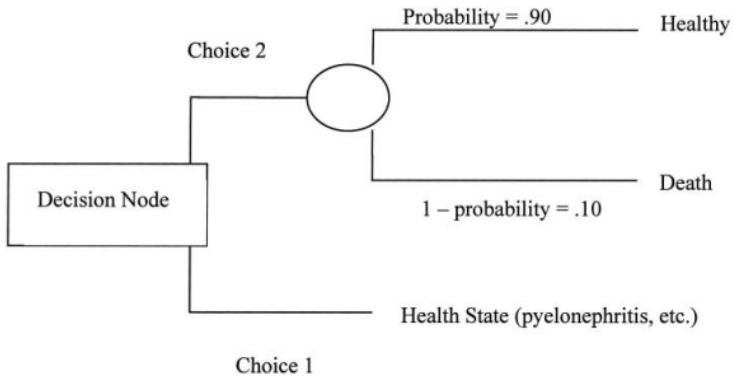
Experimental Design

- First step in data analysis

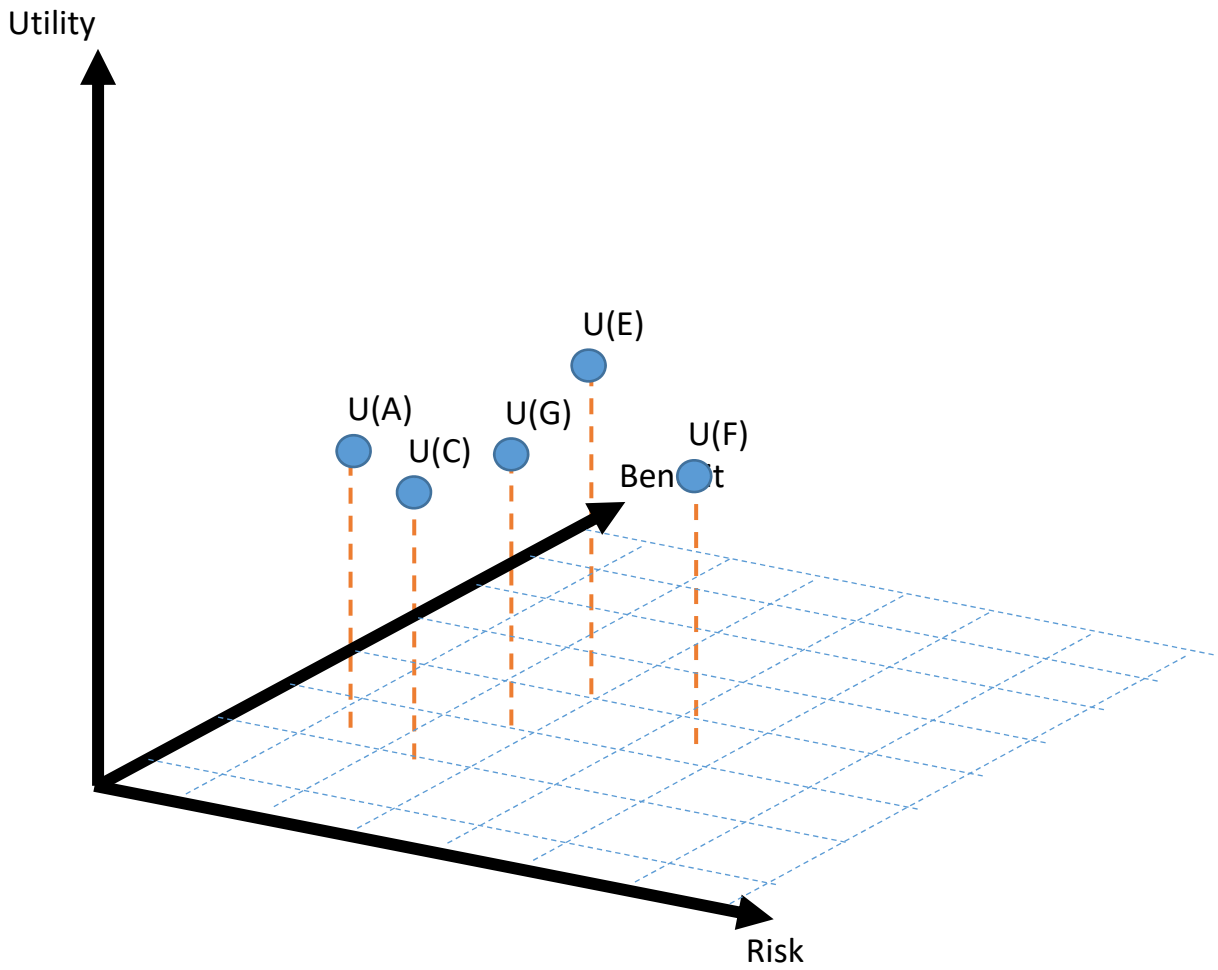
- The experimental design determines
  - The appropriate analysis tools
  - The feasible outputs
  - Interpretation of results

Utility-equivalence-Based

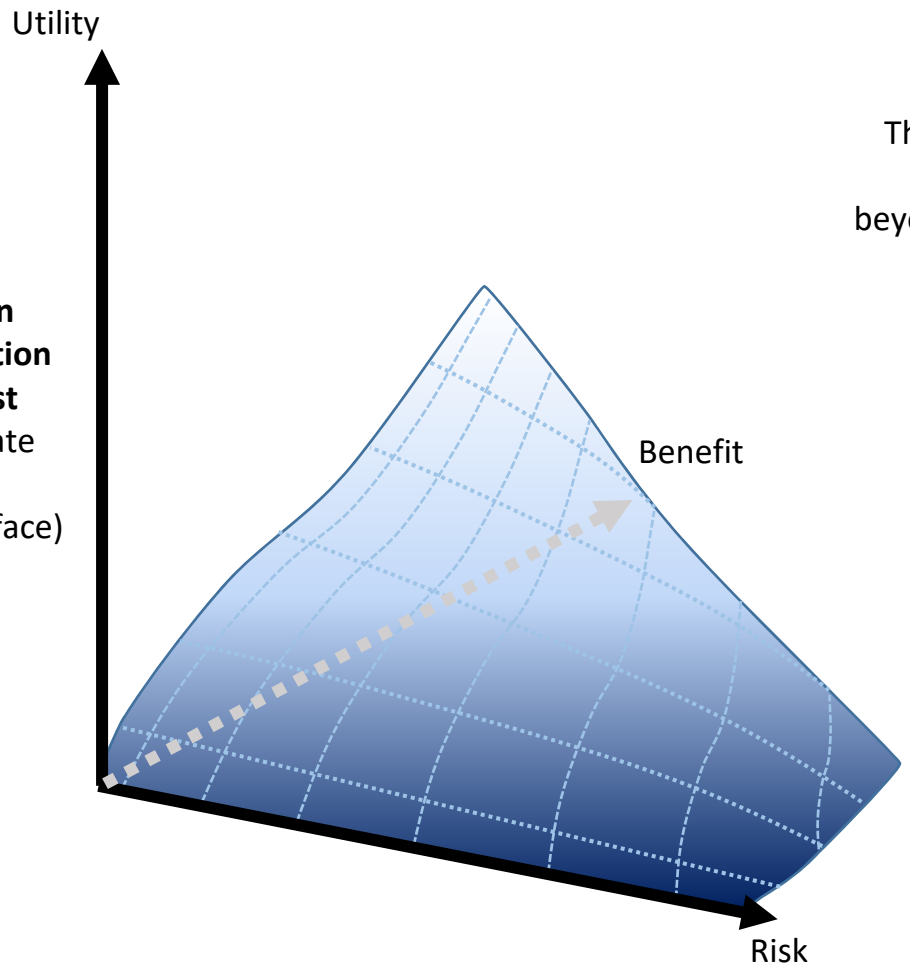




Garza and Wyrwich, Academic Emergency Medicine, 2008

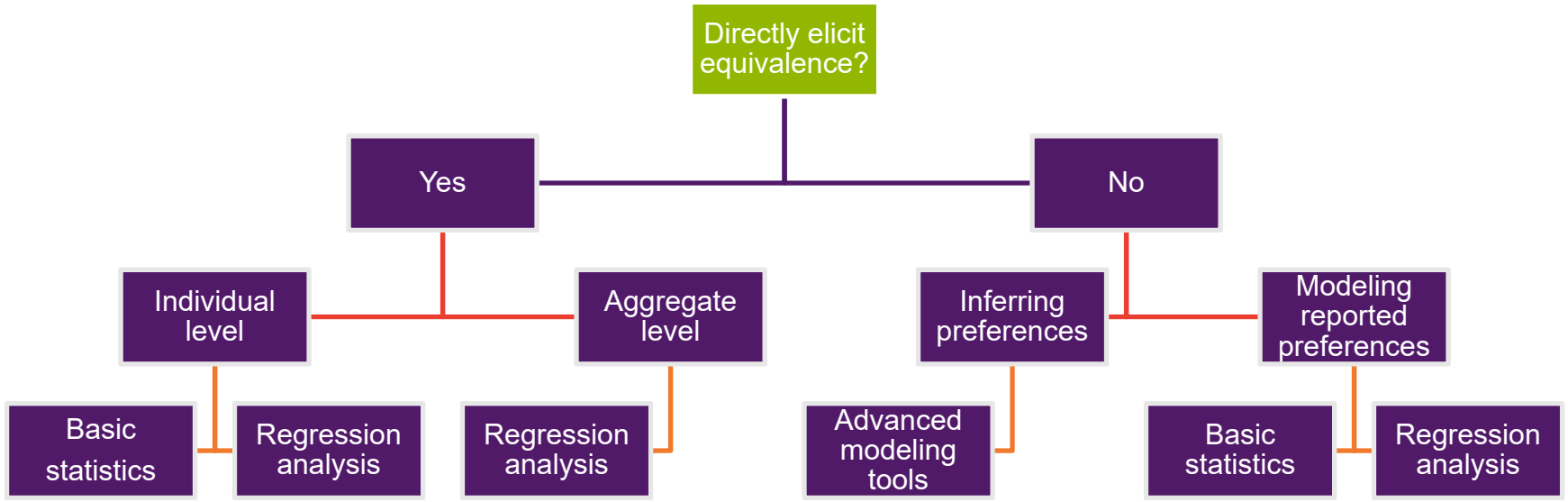

Utility
U(E)
U(A)
U(G)
U(F)
U(C)
Ben t
Risk

The experimental portion of stated-preference elicitation is about selecting the best points to be able to estimate this preference surface (or a specific point in the surface) In an unbiased way

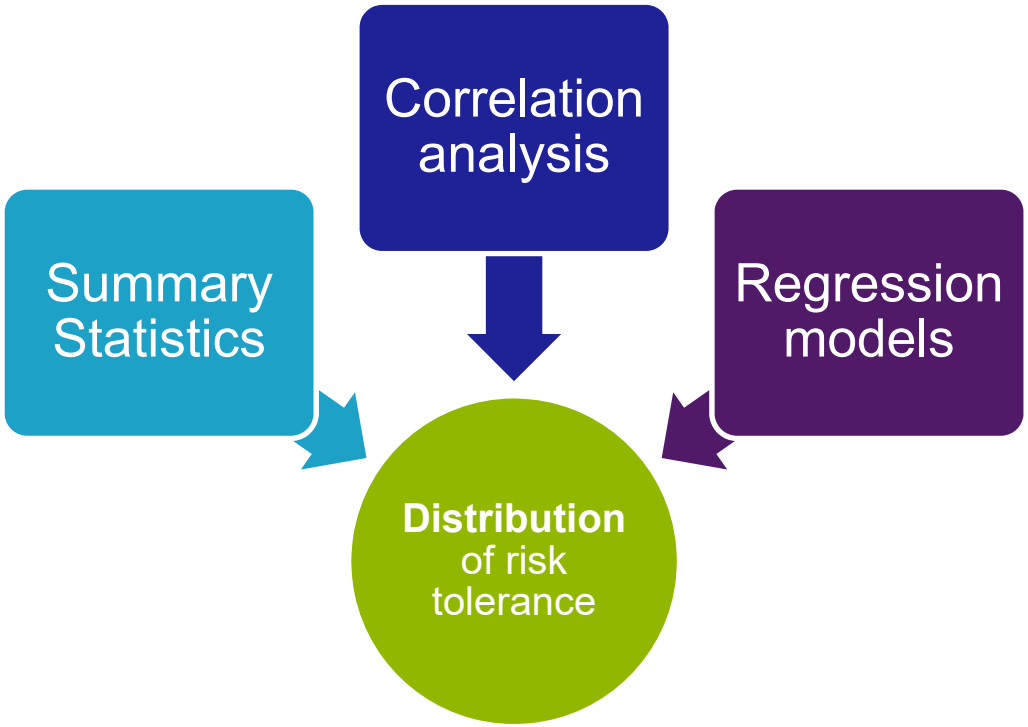


This preference surface allows us to predict utility values beyond the points we asked about

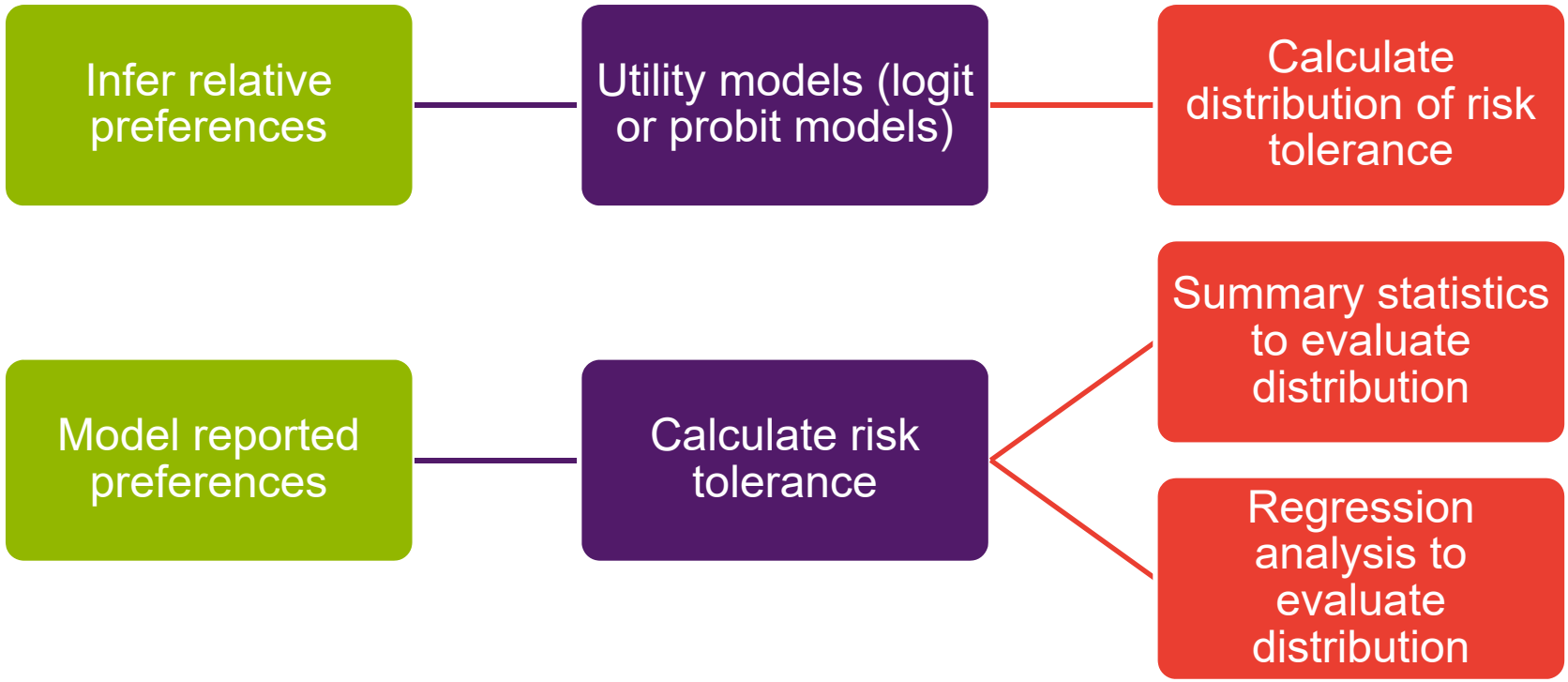
Analysis options



Equivalence-Based Analysis



Non-equivalence methods



In summary

- Experimental design is a crucial part of the analysis of stated-preference data
  - Defines appropriate analysis tools
  - The feasible outputs
  - Interpretation of results
- The information collected—whether preference measures or risk tolerance measures—also defines analysis
  - Data on measures of interest generally require simpler methods (may be more onerous to respondents)
  - Individual-level data generally can be analyzed with simpler statistical tools (require more information per respondent)
- Utility-equivalence data generally require simpler analysis methods (usually have limited information on variations given decision context)