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
Preference Information

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

Risk

Benefit
Preferences and Regulatory Decisions

Risk

Benefit

Regulators’ Benefit-Risk Threshold
Preferences and Regulatory Decisions

Risk

- Disapprove

+ Approve

Regulators’ Benefit-Risk Threshold

Patients’ Benefit-Risk Threshold
Preferences and Regulatory Decisions

Risk

Disapprove

Approve

Regulators' Benefit-Risk Threshold

Patients' Benefit-Risk Threshold
Preferences and Regulatory Decisions

![Diagram showing Regulators' and Patients' Benefit-Risk Thresholds](image)

- **Regulators' Benefit-Risk Threshold**
- **Patients' Benefit-Risk Threshold**

- **Disapprove**
- **Approve**

[Image showing points on either side of the thresholds]
Preferences and Regulatory Decisions

Regulators’ Benefit-Risk Threshold

Patients’ Benefit-Risk Threshold

Risk

Disapprove

Approve
There is a technical relationship between A and B.
There is a preferential relationship between A and B
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.

Utility = 50%, Risk = 50%, Benefit = 50%

Utility equivalence is indicated by the equilibrium point U(A) and U(B).
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

<table>
<thead>
<tr>
<th>Rating</th>
<th>Ranking</th>
<th>Tradeoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMI Protect, 2015</td>
<td>Zhang et al., 2015. JCM</td>
<td>Gonzalez et al., 2016. BJD</td>
</tr>
</tbody>
</table>
## Methods

<table>
<thead>
<tr>
<th>Rating</th>
<th>Ranking</th>
<th>Tradeoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likert scales</td>
<td>Full-ranking exercise</td>
<td>Contingent Behavior</td>
</tr>
<tr>
<td>Point-allocation technique</td>
<td>Partial-ranking exercise (e.g., best-worst scaling)</td>
<td>Discrete-choice experiments</td>
</tr>
<tr>
<td>Swing weighting</td>
<td></td>
<td>Standard Gamble</td>
</tr>
<tr>
<td>Analytic Hierarchy Process</td>
<td></td>
<td>Time-tradeoff</td>
</tr>
</tbody>
</table>
## Methods

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

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

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

- Utility
- Risk
- Benefit

### Medication feature

<table>
<thead>
<tr>
<th>Medicine feature</th>
<th>Medicine A</th>
<th>Medicine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>How red and scaly the psoriasis patches are after using the medicine</td>
<td><img src="Image1" alt="Image of very severe psoriasis" /></td>
<td><img src="Image2" alt="Image of moderate psoriasis" /></td>
</tr>
<tr>
<td>How much of your arms and legs is still covered by psoriasis patches after using the medicine</td>
<td>30 hand areas (about 50% of your arms and legs)</td>
<td>15 hand areas (about 25% of your arms and legs)</td>
</tr>
<tr>
<td>Chance of tuberculosis (TB) within 10 years</td>
<td><img src="Image3" alt="Graph of 10 out of 1000 (1%)" /></td>
<td><img src="Image4" alt="Graph of no chance" /></td>
</tr>
<tr>
<td>Chance of serious pneumonia within 10 years</td>
<td><img src="Image5" alt="Graph of 400 out of 1000 (40%)" /></td>
<td><img src="Image6" alt="Graph of 100 out of 1000 (10%)" /></td>
</tr>
<tr>
<td>Chance of lymphoma within 10 years</td>
<td><img src="Image7" alt="Graph of 20 out of 1000 (2%)" /></td>
<td><img src="Image8" alt="Graph of 50 out of 1000 (5%)" /></td>
</tr>
</tbody>
</table>

Which medicine would you choose if these were your only options?

Gonzalez et al., 2016, BMJ 2016

Gonzalez et al., 2016, BJD
What information is provided by preference-elicitation methods?

- **Rating Methods**
  - Likert scales
  - Direct weighting methods

- **Ranking Methods**
  - Full-ranking exercises
  - Partial-ranking methods (BWS)

- **Choice-Based Methods**
  - Direct elicitation
  - Contingent-valuation
  - Discrete-choice experiment
  - Time-tradeoff

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

**Provide relative importance**
- Importance is set for each attribute relative to other attributes.

**Provide tradeoff information**
- 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

40
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

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.26</td>
<td>0.001</td>
</tr>
<tr>
<td>Bid</td>
<td>−0.0003</td>
<td>0.001</td>
</tr>
<tr>
<td>Number of individuals</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−197.72</td>
<td></td>
</tr>
<tr>
<td>Chi-Squared</td>
<td>55.06</td>
<td></td>
</tr>
<tr>
<td>Individual predictions</td>
<td>69%</td>
<td></td>
</tr>
<tr>
<td>Mean WTP (0 → ∞)</td>
<td>£4893</td>
<td></td>
</tr>
<tr>
<td>95% confidence intervals</td>
<td>£4188–£6173</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. WTP estimate using the choice experiment

<table>
<thead>
<tr>
<th>Attributes of service</th>
<th>Parameter</th>
<th>p</th>
<th>Marginal WTP for unit change in attribute (Δzj/Δy)</th>
<th>Current system (a)</th>
<th>WTP for current system (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-price attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes of staff (z1)</td>
<td>0.767</td>
<td>0.001</td>
<td>£852</td>
<td>0.8</td>
<td>£682</td>
</tr>
<tr>
<td>Continuity of care (z2)</td>
<td>0.208</td>
<td>0.020</td>
<td>£231</td>
<td>0.7</td>
<td>£162</td>
</tr>
<tr>
<td>Time on waiting list (z3)</td>
<td>−0.076</td>
<td>0.001</td>
<td>−£84</td>
<td>6 months</td>
<td>−£504</td>
</tr>
<tr>
<td>'Cost per attempt' (z4)</td>
<td>−0.0009</td>
<td>0.001</td>
<td>—</td>
<td>1500</td>
<td>N/A</td>
</tr>
<tr>
<td>Chance of leaving the service with a child (z5)</td>
<td>0.142</td>
<td>0.001</td>
<td>£157</td>
<td>27.9%</td>
<td>£4380</td>
</tr>
<tr>
<td>Follow-up (z6)</td>
<td>0.288</td>
<td>0.006</td>
<td>£320</td>
<td>0.5</td>
<td>£160</td>
</tr>
<tr>
<td>Number of individuals</td>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>3893</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−1130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-squared</td>
<td>27.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual predictions</td>
<td>79%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Welfare measures**

Total WTP

<table>
<thead>
<tr>
<th></th>
<th>£4880*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>95% confidence intervals</td>
<td>£4532–£5284</td>
<td></td>
</tr>
</tbody>
</table>

*IVF service valued in the DC WTP question.


Willingness to Pay for IVF
Are Methods Interchangeable?

<table>
<thead>
<tr>
<th>Most Important Attribute-Direct Question on Post-Survey</th>
<th>Conjoint analysis group (n = 50)</th>
<th>Rating / Ranking (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to reduce colorectal incidence and mortality</td>
<td>56%</td>
<td>61%</td>
</tr>
<tr>
<td>Discomfort</td>
<td>12%</td>
<td>7%</td>
</tr>
<tr>
<td>Nature of test</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>Frequency</td>
<td>12%</td>
<td>6%</td>
</tr>
<tr>
<td>Risk of major complications</td>
<td>2%</td>
<td>4%</td>
</tr>
<tr>
<td>Out of pocket costs</td>
<td>10%</td>
<td>17%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unlabeled Test Preference</th>
<th>Conjoint analysis (n = 50)</th>
<th>Rating/Ranking (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOBT-Based attributes and levels</td>
<td>26%</td>
<td>20%</td>
</tr>
<tr>
<td>Sigmoidoscopy-based attributes and levels</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Colonoscopy-based attributes and levels</td>
<td>44%</td>
<td>39%</td>
</tr>
<tr>
<td>Radiologic test-based attributes and levels</td>
<td>26%</td>
<td>39%</td>
</tr>
<tr>
<td>No test</td>
<td>4%</td>
<td>2%</td>
</tr>
</tbody>
</table>

FOBT = fecal occult blood test

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

43
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
This preference surface allows us to predict utility values beyond the points we asked about.

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.
Analysis options

Directly elicit equivalence?

Yes
- Individual level
  - Basic statistics
  - Regression analysis
- Aggregate level
  - Regression analysis

No
- Inferring preferences
  - Advanced modeling tools
- Modeling reported preferences
  - Basic statistics
  - Regression analysis
Equivalence-Based Analysis

- Correlation analysis
- Summary Statistics
- Regression models
  - Distribution of risk tolerance
Non-equivalence methods

- Infer relative preferences
- Model reported preferences
- Utility models (logit or probit models)
- Calculate risk tolerance
  - Calculate distribution of risk tolerance
  - Summary statistics to evaluate distribution
  - Regression analysis to evaluate distribution

Summary statistics to evaluate distribution
Regression analysis to evaluate distribution
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)