Subjective Beliefs about the Health Risks of Smoking

Glenn Harrison, Andre Hofmeyr, Harold Kincaid, Brian Monroe and Don Ross

Workshop in Behavioral and Experimental Health Economics
University of Oslo, Oslo Norway
December 12, 2018
Outline

- Beliefs about Smoking
- Reports vs Beliefs
- Quadratic Scoring Rule (QSR)
- Estimating Risk Preferences
- Recovering Beliefs
- Distributional Differences in Beliefs
- Conclusions
Beliefs About Smoking

There is a presumed causal model between beliefs about the risks of smoking and the decision to start, stop, or continue smoking.

Cho et al. (2018): “Since perceived risk promotes behavioral intention and change, our findings suggest that, given its link with perceived risk of smoking-related conditions, knowledge of toxic constituents could further promote cessation behaviors.”
Beliefs about What?

Consider the question:

“For adults 35 years of age and older, what percentage of deaths from coronary heart disease are associated with smoking in the United States between 2005 and 2009?” Answer: 24.07%

There are other risks that are hypothesized to influence the decision to smoke such as the risk of addiction. (Orphanides and Zervos 1995)
How to elicit beliefs?

1. Ask the question qualitatively: “How likely is it that someone who dies of a heart attack died because they smoked?”
   - “Very Likely”, “Likely”, “Not Likely”, ...
   - Responses can’t tell us if subject is correctly informed.
   - Confidence: “Somewhat sure that it is likely?”
   - Kaufman et al. (2016), Kaufman et al. (2018), Steptoe et al. (2002), Glock, Müller and Ritter (2013), El-Toukhy and Choi (2015), Cho et al. (2018), ...

2. Ask the subject for an answer.
   - Example responses: “It’s 42.1890653729% !”, "50%"
   - Confidence: “I’m sure it’s approximately 42,” “No idea, I just said 50%”.
   - Bias from rounding: Manski and Molinari (2010)
   - Viscusi (1990), Viscusi and Hakes (2008)
     - Approach still used in ongoing litigation in Canada
How to Ask?

3 Partition the continuous space into $K$ intervals, give subjects a sack of $T$ tokens, and ask the subject to allocate tokens to the intervals in proportion to their confidence that the true answer lies in the interval.

▶ Example response: "I’ll put 5 tokens in 40-50%, 3 tokens in ..."
▶ Response space finite, but can be very large.
▶ Confidence is inferred from beliefs recovered from the token allocation

4 Lots of other methods
Reports vs Beliefs

Questions that experimenters might raise about any of these elicitation mechanisms are:

“What sense of ‘beliefs’ are using? How good is our evidence that the subject’s stated belief maps to their true beliefs?”

We infer preferences and beliefs from observed choices. The choice environment and the elicitation mechanism both influence the validity of our inferences.

Incentivize the task used to elicit beliefs. Salient outcomes to responses help protect against bias from hypothetical outcomes: Harrison (2014).
Incentivize the Task

How?

- Ask for a qualitative response?
  - Is there any feasible way to incentivize “Very Likely” over “ Likely”?
- Ask the subject for an exact answer?
  - Give them money if they get the exact answer right?
  - That assumes precisely defined beliefs
- Partition the continuous space into K intervals and allocate tokens to bins.
  - Give them money based on the allocation of tokens to bins.
Scoring Rules

Partition the continuous space into $K$ intervals, give subjects a sack of $T$ tokens, and ask the subject to allocate tokens as bets on which interval the true answer lies.

A scoring rule is needed to map the decision to allocate $t$ tokens in bin $k$ to an outcome (Savage 1971).

- Linear Scoring Rule (LSR): $\theta = \alpha - \delta (1 - r_k)$
- Quadratic Scoring Rule (QSR): $\theta = \alpha + \delta 2 r_k - \delta \sum_{i=1}^{K} r_i^2$

where $r_k$ is the “report” in the correct bin $k$, $\alpha$ is some scalar amount of money to ensure positive payoffs regardless of choice, $\delta$ is a multiplier.
Why use the QSR over LSR?

- If subject obeys Subjective Expected Utility (SEU) and is risk neutral or only modestly risk averse, the subject gets the greatest expected utility by putting all her tokens in the bin she believes is most likely to contain the correct answer with the LSR. The QSR doesn’t have this problem. (Andersen, Fountain, Harrison and Rutström 2014, p. 212)

- With the QSR, if the subject obeys SEU and is risk neutral, the reports will reflect the subjects true, subjective probabilities.

- If the subject obeys SEU and is not risk neutral, or if the subject obeys Rank Dependent Utility (RDU), Harrison and Ulm (2016) provides a method to recover subjective probabilities given the risk preferences of the subject.
The Steps for Inference (Harrison and Ulm 2016)

1. Have subjects respond to a risk preference task and a belief task that uses a QSR.
2. Estimate risk preferences from the risk task.
3. Recover the subjective probabilities over the K bins from the beliefs task using the estimated risk preferences.
4. Estimate if the difference in recovered beliefs about smoking risk are different between smokers, non-smokers, and ex-smokers.
Estimation of Risk Preferences by Maximum Likelihood

\[ RDU = \sum_{c} w_c(p) \times u(x_c) \]

where \( w(\cdot) \) is the decision weight of outcome \( x_c \) and \( u(\cdot) \) is the CRRA utility function:

\[ u(x) = \frac{x^{1-r}}{1-r} \]

\[ w_c(p) = \begin{cases} \omega \left( \sum_{k=c}^{C} p_k \right) - \omega \left( \sum_{k=c+1}^{C} p_k \right) & \text{for } c < C \\ \omega(p_c) & \text{for } c = C \end{cases} \]

and \( \omega(\cdot) \) gives the probability weighting function (PWF)
RDU nests EUT as a special case where the PWF is:

$$\omega(p_c) = p_c$$

and the flexible two parameter PWF proposed by Prelec (1998) as the second DGP:

$$\omega(p_c) = \exp(-\eta(-\ln(p_c)))^\phi$$

where $\phi > 0$ and $\eta > 0$.

The Prelect PWF nests EUT when $\phi = \eta = 1$. 
A deterministic model of choice between two options, A and B:

\[ A \succeq B \iff RDU(A) \geq RDU(B) \]

A stochastic model of choice between two options, A and B:

\[ A \succeq B \iff Pr(A) \geq Pr(B) \]

We link utilities to probabilities using the Contextual Utility model of Wilcox (2011) and the logistic CDF.

The parameters needed for estimation are \( \{r, \lambda, \phi, \eta\} \)
Intuition Behind Risk Preference Estimation

CRRA is just a function that can be concave (risk aversion), convex (risk seeking), or linear (risk neutral).

RDU
- Conceptually, think about pessimistic people who overweight the probability of something bad happening.
- I make a choice about purchasing car insurance knowing the real probability of an accident, but acting as if that probability is higher.

Contextual Utility Stochastic Model
- As the difference in utility grows, the probability of choosing the highest valued option grows.
- Some agents generally more attuned to the difference in utilities than others.
Recovery of Beliefs Given Reports and Risk Preferences

Lemma 2: Assume that the individual behaves consistently with RDU, applied to subjective probabilities. If the individual has a utility function $u(\cdot)$ that is continuous, twice differentiable, increasing and concave and maximizes rank dependent utility over weighted subjective probabilities, the actual and reported probabilities must obey the following system of equations:

$$w(p_k) \times \frac{\partial u}{\partial \theta}|_{\theta=\theta(k)} - \sum_{j=1}^{K} \left\{ w(p_j) \times r_j \times \frac{\partial u}{\partial \theta}|_{\theta=\theta(k)} \right\} = 0,$$

$$\forall k = 1, \ldots, K$$

where $\theta$ is the payout to bin $k$ defined by the QSR. Proof in Harrison and Ulm (2016).
A utility function over risky outcomes is defined by $u(\cdot)$ (and estimated over the choices in the risk task).

Decision weights attached to the subjective probabilities associated with these outcomes are defined by the PWF, $w(\cdot)$ (and estimated over the choices in the risk task).

If the token allocation maximizes RDU, and we know the shape of the utility function and the PWF, and we know the outcomes associated with decision weights, we can solve for the decision weights using linear algebra, and then solve for subjective probabilities by inverting the PWF.
Table: Mean (Standard Deviation) Summary Statistics by Smoking Status

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Age</th>
<th>White</th>
<th>Black</th>
<th>Coloured</th>
<th># of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Smoker</td>
<td>0.40 (0.49)</td>
<td>31.35 (12.68)</td>
<td>0.28 (0.45)</td>
<td>0.36 (0.48)</td>
<td>0.22 (0.42)</td>
<td>129</td>
</tr>
<tr>
<td>Smoker</td>
<td>0.53 (0.50)</td>
<td>26.49 (9.72)</td>
<td>0.19 (0.40)</td>
<td>0.25 (0.44)</td>
<td>0.42 (0.50)</td>
<td>88</td>
</tr>
<tr>
<td>Ex Smoker</td>
<td>0.38 (0.49)</td>
<td>33.81 (11.97)</td>
<td>0.44 (0.50)</td>
<td>0.16 (0.37)</td>
<td>0.38 (0.49)</td>
<td>32</td>
</tr>
<tr>
<td>Full Sample</td>
<td>0.44 (0.50)</td>
<td>29.95 (11.89)</td>
<td>0.27 (0.44)</td>
<td>0.29 (0.46)</td>
<td>0.31 (0.46)</td>
<td>249</td>
</tr>
</tbody>
</table>
Subjects faced with 4 tasks, two of which are used to recover beliefs:

- **Risk Preference task**
  - 90 pairs of lotteries*.
  - Subjects were asked to chose one lottery from each pair that they would like to play out for real money.
  - Largest prize was 700 South African Rands (∼48€ at the time)

- **Beliefs Task**
  - Subjects presented with 10 questions, one at a time.
  - Given 100 tokens and asked to allocated the tokens across 10 bins given the QSR payment rule.
  - For each bin, subjects were shown how much they would be paid if the real answer was in that bin, given their token allocation.
Risk Preference Task

Left

- Chance of winning R 20 is 55%
- Chance of winning R 160 is 25%
- Chance of winning R 190 is 20%

Right

- Chance of winning R 20 is 75%
- Chance of winning R 250 is 25%

Select Left

Select Right
## Risk Preference Estimates

### Table: Pooled Risk Preference Estimates, RDU Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Covariate</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>Constant</td>
<td>0.455</td>
<td>0.109***</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>-0.042</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>0.000</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>Coloured</td>
<td>-0.033</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-0.110</td>
<td>0.051**</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Smoker</td>
<td>0.008</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>Ex-Smoker</td>
<td>0.007</td>
<td>0.070</td>
</tr>
</tbody>
</table>

$N = 22410$, Log-Likelihood $= -14236.84$

* $p$-value $< 0.1$, ** $p$-value $< 0.05$, *** $p$-value $< 0.01$
## Risk Preference Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Covariate</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>Constant</td>
<td>0.488</td>
<td>0.064***</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>-0.003</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>0.048</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>Coloured</td>
<td>0.047</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.074</td>
<td>0.035**</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>0.003</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>Smoker</td>
<td>-0.029</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Ex-Smoker</td>
<td>0.026</td>
<td>0.047</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Constant</td>
<td>1.054</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>-0.038</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>-0.053</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>Coloured</td>
<td>0.035</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-0.015</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Smoker</td>
<td>-0.040</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>Ex-Smoker</td>
<td>0.137</td>
<td>0.113</td>
</tr>
</tbody>
</table>

N = 22410, Log-Likelihood = -14236.84

* $p$-value < 0.1, ** $p$-value < 0.05, *** $p$-value < 0.01
Risk Preference Estimates

- Little variation in utility function curvature
  - Male

- Some variation in probability weighting function
  - Male and Age

- No statistically significant differences in (atemporal) risk preferences between smokers, ex-smokers, non-smokers: confirming findings from Harrison, Hofmeyr, Ross and Swarthout (2018)

- Evidence of non-linear probability weighting and utility function curvature, special case where reports = beliefs violated
Probability Weighting Function and Decision Weights of Average Subject

\[ \phi = 0.629 \]
\[ \eta = 1.020 \]

![Graph showing probability weighting function and decision weights](image)
Beliefs Task Questions of Interest

10 questions were posed to the subjects, but we focus on 4 today.

- For adults 35 years of age and older, what percentage of deaths from lung cancer are associated with smoking in the United States between 2005 and 2009? - Answer 82.42%

- For adults 35 years of age and older, what percentage of deaths from other cancers (cancers of the lip, pharynx and oral cavity, esophagus, stomach, pancreas, larynx, cervix uteri, kidney and renal pelvis, bladder, liver, colon and rectum, and acute myeloid leukemia) are associated with smoking in the United States between 2005 and 2009? Other cancers do not include lung cancer. - Answer 20.17%
Beliefs Task Questions of Interest

- For adults 35 years of age and older, what percentage of deaths from coronary heart disease are associated with smoking in the United States between 2005 and 2009? - Answer 24.07%

- For adults 35 years of age and older, what percentage of deaths from chronic obstructive pulmonary disease are associated with smoking in the United States between 2005 and 2009? - Answer 78.76%
Beliefs Task

For adults 35 years of age and older, what percentage of deaths from lung cancer are associated with smoking in the United States between 2005 and 2009?
Reports by Question and Smoking Status

Question 1 (Lung Cancer) vs Question 2 (Other Cancers)

Question 3 (Heart Disease) vs Question 4 (Pulmonary Disease)

Smoking Status
- Smoker
- Ex-Smoker
- Never-Smoker

Bin Labels:
- 0-10%
- 40-50%
- 90-100%
Recovered Beliefs by Question and Smoking Status

- **Question 1 (Lung Cancer)**
- **Question 2 (Other Cancers)**
- **Question 3 (Heart Disease)**
- **Question 4 (Pulmonary Disease)**

Smoking Status:
- Smoker
- Ex-Smoker
- Never-Smoker

Bin Labels:
- 0-10%
- 40-50%
- 90-100%
Beta Distribution of Beliefs

- Bootstrap distribution of risk preferences to allow covariances in risk preference estimates to propagate through to belief recovery
  - Individual level estimates of RDU model
- Use point estimates as means and covariance matrix as the covariance matrix of a multivariate Normal distribution of risk preferences
- Each draw from this joint distribution represents a full set of risk preferences
  - Draw 500 sets from this distribution
- For each set drawn, and for each subject, recover the beliefs from the reports of each question
- Fit Beta distribution to recovered beliefs using maximum likelihood interval regression, estimating mean and variance
  - Allows for statements about bias and confidence
  - Some limitations (Engelberg, Manski and Williams 2009)
- Additional methods currently in development. Stay tuned
### Beta Distribution of Beliefs - Lung and Other Cancer

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Question 1</th>
<th>Question 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E[X]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.574</td>
<td>0.058</td>
</tr>
<tr>
<td>Black</td>
<td>0.034</td>
<td>0.044</td>
</tr>
<tr>
<td>White</td>
<td>0.123</td>
<td>0.042***</td>
</tr>
<tr>
<td>Coloured</td>
<td>-0.009</td>
<td>0.045</td>
</tr>
<tr>
<td>Male</td>
<td>-0.031</td>
<td>0.027</td>
</tr>
<tr>
<td>Age</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Smoker</td>
<td>0.066</td>
<td>0.029**</td>
</tr>
<tr>
<td>Ex Smoker</td>
<td>0.027</td>
<td>0.042</td>
</tr>
</tbody>
</table>

| **var[X]**   |            |            |          |          |
|--------------|            |            |          |          |
| Constant     | 0.044      | 0.010***   | 0.049    | 0.012*** |
| Black        | 0.001      | 0.008      | 0.002    | 0.009    |
| White        | -0.016     | 0.007**    | -0.008   | 0.008    |
| Coloured     | 0.008      | 0.009      | 0.005    | 0.009    |
| Male         | 0.010      | 0.006*     | 0.006    | 0.006    |
| Age          | 0.000      | 0.000      | 0.000    | 0.000    |
| Smoker       | -0.007     | 0.007      | -0.001   | 0.006    |
| Ex Smoker    | 0.002      | 0.009      | -0.001   | 0.010    |

*** p-value < 0.01, ** p-value < 0.05, and * p-value < 0.1

All standard errors are clustered on subject ID
### Beta Distribution of Beliefs - Heart and Pulmonary Disease

#### Question 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E[X]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.418</td>
<td>0.043*</td>
<td>0.430</td>
<td>0.058</td>
</tr>
<tr>
<td>Black</td>
<td>0.007</td>
<td>0.039</td>
<td>0.010</td>
<td>0.045</td>
</tr>
<tr>
<td>White</td>
<td>0.082</td>
<td>0.037**</td>
<td>0.061</td>
<td>0.042</td>
</tr>
<tr>
<td>Coloured</td>
<td>-0.004</td>
<td>0.039</td>
<td>-0.035</td>
<td>0.042</td>
</tr>
<tr>
<td>Male</td>
<td>-0.031</td>
<td>0.024</td>
<td>-0.008</td>
<td>0.026</td>
</tr>
<tr>
<td>Age</td>
<td>0.002</td>
<td>0.001*</td>
<td>0.002</td>
<td>0.001**</td>
</tr>
<tr>
<td>Smoker</td>
<td>0.063</td>
<td>0.027**</td>
<td>0.018</td>
<td>0.027</td>
</tr>
<tr>
<td>Ex Smoker</td>
<td>-0.019</td>
<td>0.034</td>
<td>-0.013</td>
<td>0.037</td>
</tr>
</tbody>
</table>

#### Question 4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>var[X]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.031</td>
<td>0.007***</td>
<td>0.055</td>
<td>0.012***</td>
</tr>
<tr>
<td>Black</td>
<td>0.015</td>
<td>0.007**</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td>White</td>
<td>-0.002</td>
<td>0.005</td>
<td>-0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>Coloured</td>
<td>0.010</td>
<td>0.006*</td>
<td>-0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>Male</td>
<td>0.006</td>
<td>0.004</td>
<td>0.014</td>
<td>0.006**</td>
</tr>
<tr>
<td>Age</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Smoker</td>
<td>0.007</td>
<td>0.005</td>
<td>-0.002</td>
<td>0.007</td>
</tr>
<tr>
<td>Ex Smoker</td>
<td>0.002</td>
<td>0.006</td>
<td>0.001</td>
<td>0.008</td>
</tr>
</tbody>
</table>

*** p-value < 0.01, ** p-value < 0.05, and * p-value < 0.1

All standard errors are clustered on subject ID.
Bias and Confidence

Bias: the extent to which beliefs differ from the true answer

Confidence:

- Overestimation of one’s actually ability or performance,
- Overplacement of one’s self relative to others,
- Overprecision, excess certainty about the accuracy of one’s beliefs.

It is the third of these definitions that we refer to as confidence here

We evaluate confidence as the difference between the estimated variance controlling for observable characteristics and the average variance for the sample.
Beliefs Bias by Covariate

Question 1 (Lung Cancer)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Question 2 (Other Cancer)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Question 3 (Heart Disease)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Question 4 (Pulmonary Disease)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Bias
-0.50 -0.25 0.00 0.25 0.50
-0.50 -0.25 0.00 0.25 0.50
Confidence Deviations from Average Confidence

Question 1 (Lung Cancer)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Question 2 (Other Cancer)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Question 3 (Heart Disease)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Question 4 (Pulmonary Disease)
- Constant
- Black
- White
- Coloured
- Male
- Age
- Smoker
- Ex Smoker

Confidence Deviation from Average Confidence (0.00)
Conclusions

- Our subjects have incorrect beliefs about the health risks of smoking
  - Lung cancer and chronic obstructive pulmonary disease **underestimated**
  - Other cancers and heart disease **overestimated**
  - Lung cancer result contrary to results by Viscusi (1990) and Viscusi and Hakes (2008)*

- Smokers believe a greater portion of lung cancer deaths and heart disease deaths are attributable to smoking compared to non-smokers
  - Contrary to studies showing that smokers believe smoking risks are less severe compared to non-smokers.
  - Suggestive that a causal model between risk and smoking is more complicated than a positive correlation with beliefs

- No evidence of differences in confidence across smoking status
  - Some evidence of variation in confidence by observable characteristics (White)
References


References II


References III


