

# Using an Online Survey to Elicit Preferences in China: Comparison of Different Methods Based on SF-6Dv2

The image shows a screenshot of a research paper abstract page. At the top, the title is "Using an Online Survey to Elicit Preferences in China: Comparison of Different Methods Based on SF-6Dv2". Below the title, the authors are listed: Ahmed M.Y. Osman, PhD candidate [1,2], Jing Wu, PhD [1,2], Xiaoning He, PhD [1,2], and Gang Chen, PhD [3]. The authors' affiliations are provided: 1-School of Pharmaceutical Science and Technology, Tsinghua University, Beijing, PRC; 2-Center for Social Science Survey and Data, Tsinghua University, Beijing, PRC; and 3-Center of Health Economics, Melbourne Business School, Melbourne University, Melbourne, Victoria, Australia. A small portrait of a woman is visible on the right side of the header.

The abstract is divided into several sections:

- Background:** Online surveys are becoming the norm in Social Science research. This is mostly derived from the advantages that this medium provides compared to traditional methods, including lower cost, better efficiency, flexibility, and wider sample. Advancement in Chinese social sciences studies was coupled with several technical hindrances. Surveys are used to collect, estimate and understand the data.
- Methods:**
  - Sample:** A pilot study that was run before the launch of the official survey suggested that 3 methods for every respondent were overwhelming (they spent 50 min on average answering the survey), this reduced to 2 methods (30 min on average). A total of 660 respondents were targeted to ensure having about 660 respondents per arm and account for about 20% unusable responses (design efficiency requirements). The sample was Quota controlled to ensure representation of the population. The Quota is based on the China Statistical Yearbook (2004) and included only Adults 18-70 who are non-married & unretired (Married and
- Results:**
  - Study Sample:** A quota-controlled quota sampled representative population of the Chinese general population in terms of age, gender, geographical allocation, educational level and household registration was surveyed online during the summer of 2010. A complete sample selection flow chart is presented in the following flowchart.

```
graph TD
    A[Total number of Qs  
N0=4080] --> B[Final Quota  
N0=2000]
    B --> C[Final Quota  
N0=2000]
```
- Discussion:** This online survey study generated models to elicit preferences on the SF-6Dv2 showed that the quality of the data from different elicitation techniques is not the same even when applying similar restrictions and controlling measures. It has suggested that an estimation of about 30% of the data to be of poor quality when calculating the sample size is a reasonable assumption. Furthermore, the quota requirements were achieved to a satisfactory level. The quota control and attention checks are crucial to ensure that the data is of high quality.
- Conclusion:** Online surveys are a promising alternative to face-to-face interviewing methods for collecting and generating self-preference values from the general population in China if the proper controls and precautions were considered. The sample should follow a specific quota to represent the general population or be directed purposefully to the group under investigation. Some subgroups have limited access to the internet. Online elicitation methods are a better choice for online surveys than traditional methods in China but still need quality measures to be included. Self-anchoring using death values from the same task is acceptable to convert latent ordinal values to 0-1 death-full health cardinal values.

At the bottom of the page, there are buttons for [ABSTRACT](#), [REFERENCES](#), [SOURCE AUTHORS](#), and [GET FULLTEXT](#).

Ahmed M.Y. Osman, PhD candidate [1,2], Jing Wu, PhD [1,2], Xiaoning He, PhD [1,2], Gang Chen, PhD [3]

- 1-School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, PRC  
2-Center for Social Science Survey and Data, Tianjin University, Tianjin, PRC  
3-Centre of Health Economics, Monash Business School, Monash University, Melbourne, Victoria,  
Australia



PRESENTED AT:



## BACKGROUND

Online surveys are becoming the norm in Social Science<sup>1,2</sup>. This is mainly derived from the advantages that this medium provides compared to traditional methods, including lower cost, better efficiency, flexibility, and wider sample<sup>3</sup>. Adaptation by Chinese social sciences studies was coupled with several technical hindrances<sup>4</sup>.

Surveys are used to collect, estimate and understand the public preference for health and then include them in studies that measure the effectiveness of health technologies<sup>5,6</sup>. Preference can be elicited by comparing hypothetical states to attach utilities (benefit values) to their components. These states are designed from Multi-Attributes Utility Instruments (ex. SF-6D<sup>7</sup>, EQ5D<sup>8</sup>), suitable for economic evaluations<sup>9</sup>.

Preference-based instruments are composed of 2 parts: a) Health Descriptions, i.e., Dimensions decomposed to levels, and b) Utilities attached to the levels unique to each country<sup>10,11</sup>.

Elicitation of preferences can be either by applying direct methods. The task results are exactly the utilities (such as Standard Gamble / Time Trade-Off) or indirect (such as Discrete Choice Experiment / Best Worse Scaling). The direct suffers from several drawbacks such as more cognitively demanding and measuring only one characteristic of health<sup>12-14</sup>.

In the last few years, there has been an ongoing debate about which of the two choice methods is best, DCE or profile-case BWS<sup>15,16</sup>. An assumed advantage of BWS over DCE is its suitability to certain vulnerable subgroups such as the elderly or

certain patients. Other advantages relate to the richer information gained from each task and the analytical models that estimate individual values and relative importance of domains<sup>17,18</sup>.

This report aimed to:

- ★ Present a practical case for health state preference elicitation online survey in China (pros, cons & improvement)
- ★ Comparing the reliability of cardinal and ordinal methods from an online sample
- ★ Generating models for ordinal elicitation methods based on self-anchoring on death

# METHODS

## Sample

A pilot study that was run before the launch of the official survey suggested that 3 methods for every respondent were overwhelming (they spent 50 min on average answering the survey), thus reduced to 2 methods (30 min on average). A total of 990 respondents was targeted to insure having about 660 respondents per arm and account for about 20% unreliable responses (design efficiency requirements).

The sample was Quota controlled to ensure representatives of the population. The Quota is based on the China Statistical Yearbook (2018) and included only Adults (+18) who can read & understand Mandarin and remained in the mainland for at least 7 months/ year during the past 5 years.

## Instrument

The survey is in simplified Chinese and designed using Sawtooth Lighthouse Studio® version 9.7.2 (Sawtooth Software, Inc. Provo, UT) and composed mainly of 3 parts (Introduction and content, 3 preference elicitation methods, i.e., BWS, DCE and TTO, and further sociodemographic and health-related behaviors). Health states were designed based on the SF-6Dv2 (Med Care 2020;58: 557–565).

## Experimental Design

In principle, full factorial design is unfeasible as there are 18750 health states produced from the survey tool. Then, head-to-head comparisons require that tasks be controlled along with different elicitation techniques. Thus, the experimental design was performed using the survey software only for the DCE tasks, and then the other two tasks were built/sampled from the DCE.

The Fractional factorial design insured maximum level balance & near orthogonality (possibly of interactions). A total of 800 health scenario pairs (100 versions) were generated, and the dimension orders were randomized between respondents. The BWS was based on the last 5 tasks in DCE, and the TTO was based on the first 3 BWS tasks, 1 light, 1 heavy and pits states.

DCE Question

1	To select the language of the questionnaire from the drop down box, the original survey language is Chinese
2	Hover your mouse over to get more information about the task
3	The current question (problem)
4	Shows the current question position out of all required questions
5	Here is the description of the first health state
6	Here is the description of the second health state
7	These are the dimensions used to describe each health state (constant through all the task)
8	Choice buttons, please choose only the one under the health state you believe is most desirable
9	A second question to know how bad do you think the worse health state of the above two compared to death
10	Choice buttons, please choose the left if the worse health state is better than death and the right one otherwise
11	These are the buttons to proceed to the next or previous page, you must answer both questions before going to the next page
12	A progress bar that shows how much of the questionnaire did you finish and how much is left (approximately)

1 Select Language | Powered by Tong Translation

 天津大学  
TIANJIN UNIVERSITY

2 帮助

3 请您仔细阅读以下两个健康状态（生活），如果您必须从中选择一个，您认为哪一个健康状态（生活）更好？

4 (第2题, 共10题)

生活一

您很少会由于身体健康或情绪问题，而干扰到您的社交活动

您很少会由于身体健康或情绪问题，而导致实际完成的工作或其他日常活动比您要完成的少

您的身体有虫咬疼痛

您有轻微社会感觉到度量

您的健康很大程度上限制了您去乘坐公交车

您大部分时间会感觉到沮丧或非常紧张

选择

生活二

您一直会由于身体健康或情绪问题，而干扰到您的社交活动

您大部分时间会由于身体健康或情绪问题，而导致实际完成的工作或其他日常活动比您要完成的少

您的身体有非常严重的疼痛

您从来没有感觉到度量

您的健康很大程度上限制了您乘坐高强度活动

您从来没有感觉到沮丧或非常紧张

选择

7

- 社会功能
- 角色功能
- 疼痛
- 活力
- 躯体功能
- 精神健康

8

9 将之前会没有选择的健康状态与“立即死亡”进行比较，您认为哪一个更好？

10

说更差的 | 立即死亡

11

上一步 | 下一步

12 0% | 100%

由同济-伟星健康大数据与健康行为模式应用研究中心

BWS Question

1	To select the language of the questionnaire from the drop down box, the original survey language is Chinese
2	Hover your mouse over to get more information about the task
3	The current question (problem)
4	Shows the current question position out of all required questions
5	Here is the description of the health state
6	Here on the left, choose only ONE of the health state descriptions that you believe is the most desirable
7	Here on the right, choose only ONE of the health state descriptions that you believe is the least desirable
8	Your choices appear here to make sure that you have chosen exactly what you want to choose, please read them again before answering the confirmation question)
9	A confirmation question to make sure that you correctly choose what you believe is the ONE Most desirable AND the ONE least desirable description of the health stat)
10	These are the buttons to proceed to the next or previous page, you must answer both questions before going to the next page
11	A progress bar that shows how much of the questionnaire did you finish and how much is left (approximately)

1 Select Language ... Powered by Tongji Translate



天津大学  
TIANJIN UNIVERSITY

2 帮助

3 请您仔细阅读以下健康状态，如果您必须从中选择一个最倾向的和一个最不倾向的描述。  
您认为哪一个描述是最倾向的，哪一个描述是最不倾向的？

4 (第2题，共10题)

最倾向的 😊👍	健康状况	最不倾向的 😞🗨️
<input type="radio"/>	您有时会由于身体健康或情绪问题，而导致实际完成的工作或其他日常活动比想要完成的少	<input type="radio"/>
<input checked="" type="radio"/>	您很少会感到疲惫	<input type="radio"/>
<input type="radio"/>	您很少会感到沮丧或非常紧张	<input type="radio"/>
<input type="radio"/>	您的身体有非常轻度的疼痛	<input type="radio"/>
<input type="radio"/>	您的健康有一点限制您做高强度活动	<input checked="" type="radio"/>
<input type="radio"/>	您一直会由于身体健康或情绪问题，而干扰到您的社交活动	<input type="radio"/>

6 5 7

8 您选的最倾向的描述是您很少会感到疲惫  
您选的最不倾向的描述是您的健康有一点限制您做高强度活动

9 上面是否您的真实想法？

是  
 否

10 上一步

11 0% 100%

中国大陆居民健康水平与健康状态偏好网络调查研究

## TTO Question

1	To select the language of the questionnaire from the drop down box, the original survey language is Chinese
2	Hover your mouse over to get more information about the task
3	The current question (problem)
4	Here is the description of the first health state which is always full health
5	This is the slider bar representing the number of years you are expected to live in full health before you die, later the handle on this bar will change automatically depending on your opinion until you are satisfied
6	This is the slider bar representing the number of years you are expected to live in the second health state before you die, this is either 10 or 20 years
7	Here is the description of the second health state
8	Click on this button if you think that the first and the second health states are EQUAL, And it will also take you to the next page
9	Click on this button if you think that the Second health states is BETTER than the First health state
10	Click on this button if you think that the First health states is BETTER than the Second health state
11	This button is to go back to the previous page in case you wanted to recheck/change any of your answers
12	A progress bar that shows how much of the questionnaire did you finish and how much is left (approximately)

1 [Select Language] ... | Powered by Google Translate



2 帮助

3 问题1: 在屏幕显示的描述下, 您认为是生活一好, 还是生活二好, 还是它们大致相同?  
(请通过选择“生活一更好”或“生活二更好”来改变生活一的生存时间。当你认为二者差不多时, 请点击“大致相同”)

4 生活一: 完全健康

0年 1年 2年 3年 4年 5年 6年 7年 8年 9年 10年

5 生活一 5年后死亡

6 生活二 10年后死亡

7 生活二:  
您**很少**会由于身体健康或情绪问题, 而导致实际完成的工作或其他日常活动比想要完成的少  
您**大部分时间**会由于身体健康或情绪问题, 而干扰到您的社交活动  
您**大部分时间**会感到疲惫  
您**有时**会感到沮丧或非常紧张  
您的身体**没有**疼痛  
您的健康**有一点限制**您做**中等强度**活动

8 大致相同

9 生活二更好

10 生活一更好

11 上一步 下一步

12 0% 100%

中国城镇居民健康水平与健康状态转换网络调查研究

## Statistical Analysis

Descriptive statistics including:

- Demographics
- Duration of answering the survey
- Ease of understanding and answering each method

Inferential statistics (regression analysis) including::

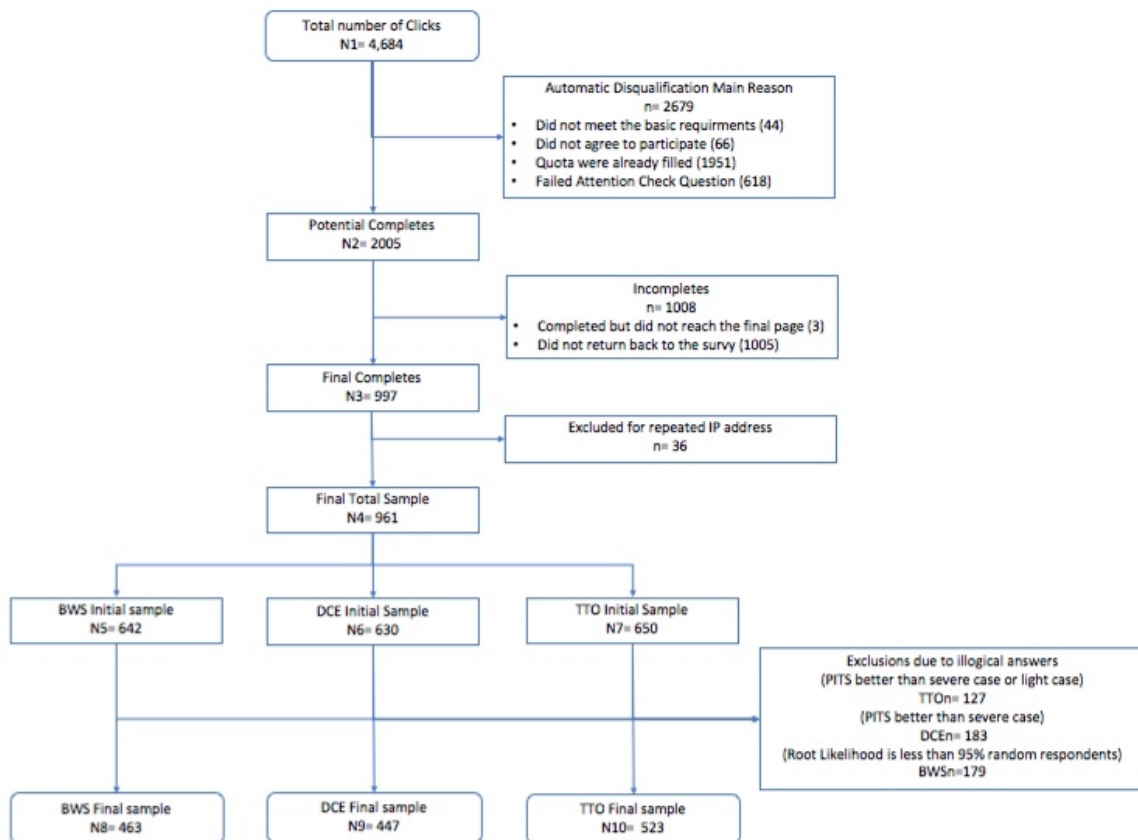
- Homogenous preference: Conditional Logit (BWS & DCE); Linear by Ordinal Least Square (TTO)
- Heterogenous preference: Hierarchical Bayes Mixed Logit (BWS & DCE); Hierarchical Bayes Mixed Effects Linear Model (TTO)

Self-Anchoring on death values for BWS & DCE to generate the scores

# RESULTS

## Study Sample

A cross-sectional quota sampled representative population of the Chinese general population in terms of age, gender, geographical allocation, educational level and household registration was surveyed online during the summer of 2019. A complete sample selection flow chart is presented in the following flowchart.



The main quota characteristics of the Final sample of the analysis was indifferent from the intended and general population sample as presented in the following table (Table 1)

<b>The socio-demographic characteristics of the study sample</b>			
<b>Quota category</b>	<b>China National adults' (%)</b>	<b>Intended sample N=990 (n)</b>	<b>Final sample* N=961 n (%)</b>
<i>Gender</i>			
Male	51.2%	507	490 (51.0%)
Female	48.8%	483	471 (49.0%)
<i>Age (years)</i>			
18-29	21.54%	213	212 (22.1%)
30-39	18.70%	185	178 (18.5%)
40-49	21.06%	209	201 (20.9%)
50-59	17.06%	169	163 (17.0%)
≥ 60	21.63%	214	207 (21.5%)
<i>Education Level</i>			
Primary or less	26.24%	260	247 (25.7%)
Middle school	40.28%	399	389 (40.5%)
High school and Vocational	17.18%	170	165 (17.2%)
University or higher	16.30%	161	160 (16.6%)
<i>Geographical Region</i>			
Northern	13%	129	126 (13.1%)
North East	8%	79	70 (7.3%)
Eastern	30%	297	292 (30.4%)

Central	28%	277	272 (28.3%)
South West	14%	139	132 (13.7%)
North West	7%	69	69 (7.2%)
<i>Household Registry</i>			
Urban	58.52%	579	588 (61.2%)
Rural	41.48%	411	373 (38.8%)

---

The average answering duration for the survey was 26.8 min (SD 24.5 and median of 21.2 min). Answering and understanding any of the three tasks was reasonably easy. The highest difficulty in answering and understanding a task was for the TTO (5.9% hard understanding & 8.1% hard in answering).

## Homogenous preference

Several CL and OLS models were compared for the elicitation techniques (results not shown). The best fit of these models was anchored on death (for the latent ordinal methods) then constrained to generate the model for utility values of the SF-6Dv2 tool. The best fit model for DCE was the model considering main effects only and weighting results by the answer to the opinions of the calm death sentence. The best fit model for the BWS was the best only choice weighted by considering all dimensions important or not. The best fit model for the TTO was the OLS with the addition of MOST (a Dummy variable equal to 1 if the health state has one or more dimension levels at the 2 most severe levels). The following table (Table 2) summarizes the coefficient estimates.

SF-6Dv2 attributes and levels		DCE			BWS		TTO		
		Coef. Ordinal	S.E	Coef. Anchored (disutility)	Coef. Ordinal	S.E	Coef. Anchored	Coef. Cardinal (disutility)	S.E
	Intercept	N/A	N/A	N/A	N/A	N/A	N/A	0.3312	0.0291
Physical Functioning (PF)	1	0.2934	0.0455	-0.0024	4.0911	0.1036	0.1667	0.0000	
	2	0.1463	0.0456	0.0590	2.9746	0.1136	0.1212	0.0305	0.0274
	3	0.1461	0.0458	0.0591	2.5489	0.1267	0.1039	0.0305	0.0274
	4	-0.1078	0.0465	0.1650	1.3072	0.1563	0.0533	0.1017	0.0351
	5	-0.4779	0.0481	0.3196	0.3986	0.1599	0.0162	0.2045	0.0328
Role Functioning (RF)	1	0.1227	0.0469	0.0688	4.0221	0.1044	0.1639	0.0000	
	2	0.0706	0.0462	0.0906	3.8283	0.1046	0.1560	0.0125	0.0160
	3	0.0704	0.0465	0.0907	2.9418	0.1165	0.1199	0.0125	0.0160
	4	-0.1318	0.0467	0.1750	1.8196	0.1373	0.0741	0.0207	0.0359
	5	-0.1320	0.0464	0.1751	1.4636	0.1457	0.0596	0.0394	0.0353
Social Functioning (SF)	1	0.1373	0.0458	0.0627	4.0495	0.1020	0.1650	0.0000	
	2	0.1371	0.0458	0.0628	3.7622	0.1063	0.1533	0.0077	0.0259
	3	0.0245	0.0462	0.1098	2.8979	0.1180	0.1181	0.0077	0.0259
	4	-0.1227	0.0460	0.1712	1.1077	0.1518	0.0451	0.0077	0.0259
	5	-0.1763	0.0467	0.1936	0.8880	0.1686	0.0362	0.0979	0.0321
Pain (PN)	1	0.5874	0.0520	-0.1252	4.3668	0.1056	0.1779	0.0000	
	2	0.4966	0.0517	-0.0873	3.0097	0.1207	0.1226	0.0121	0.0327
	3	0.4962	0.0513	-0.0871	2.9133	0.1257	0.1187	0.0233	0.0361
	4	-0.0128	0.0508	0.1254	1.5080	0.1639	0.0615	0.0825	0.0412
	5	-0.6469	0.0539	0.3901	-0.3389	0.1658	-0.0138	0.1816	0.0368
Mental Health (MH)	6	-0.9205	0.0591	0.5043	-0.7972	0.1646	-0.0325	0.2775	0.0359
	1	0.2937	0.0467	-0.0025	3.9571	0.1048	0.1612	0.0000	
	2	0.1966	0.0460	0.0380	3.9569	0.1015	0.1612	0.0096	0.0143
	3	0.0115	0.0466	0.1152	3.2669	0.1108	0.1331	0.0096	0.0143
	4	-0.2006	0.0458	0.2038	1.0514	0.1592	0.0428	0.0096	0.0143
Vitality (VT)	5	-0.3011	0.0465	0.2458	0.7007	0.1553	0.0286	0.0608	0.0329
	1	0.2909	0.0467	-0.0014	4.0540	0.1020	0.1652	0.0000	
	2	0.1812	0.0457	0.0444	3.9246	0.1031	0.1599	0.0324	0.0312
	3	0.0105	0.0458	0.1157	3.6425	0.1050	0.1484	0.0522	0.0377
	4	-0.1781	0.0462	0.1944	1.8067	0.1360	0.0736	0.0970	0.0349
	5	-0.3044	0.0468	0.2471	1.5171	0.1453	0.0618	0.1283	0.0336
	Death	-1.0360	0.0489	-1.0000			0.0000	N/A	N/A
	MOST							-0.1070	0.0446
Disparities (insignificant)		0 (6)			0 (0)		0 (17)		
	Log-likelihood, df	-3426.1938	26	-7329.5971	32	-2079.8120	22		
	Percent Certainty	19.5983		30.4746		26.0964			
	AIC	6904.3898		14721.1943		4203.6243			
	CAIC	7094.3966		14968.2934		4336.7020			

BIC	7068.5966		14937.2954		4536.7558		
ABIC	6985.7800		14838.7816				
Chi-Square	1670.3066		6425.4784				
Fit Statistic (RLH)	0.3836		0.3941		0.2657		
MAE					0.3776		
RMSE					0.4695		
pits Estimation	-0.6855		0.1700		-0.0326		
		Rank		Rank		Rank	
Attributes Importance	PF	19.10%	2	18.13%	2	23.70%	2
	RL	6.30%	6	12.56%	5	5.50%	6
	SF	7.77%	5	15.52%	4	7.80%	4
	PN	37.35%	1	25.35%	1	37.23%	1
	MH	14.73%	4	15.99%	3	5.78%	5
	VT	14.74%	3	12.45%	6	20.00%	3

## Heterogenous preference

Several HB and LHB models were compared for the elicitation techniques (results not shown). The best fit of these models was anchored on death (for the latent ordinal methods) then constrained to generate the model for utility values of the SF-6Dv2 tool. The best fit model for DCE was the model considering main effects only

without adding any covariates. The best fit model for the BWS was the best only choice with a covariate considering if all dimensions are important or not. The best fit model for the TTO was the LHB with the addition of MOST. The following table (Table 3) summarizes the coefficient estimates.

SF-6Dv2 attributes and levels		DCE			BWS		TTO		
		Coef. Ordinal	S.E	Coef. Anchored (disutility)	Coef. Ordinal	S.E	Coef. Anchored	Coef. Cardinal (disutility)	S.E
	Intercept	N/A	N/A	N/A	N/A	N/A	N/A	0.3494	0.0052
<i>Physical Functioning (PF)</i>	1	0.4875	0.0145	-0.0023	8.5529	0.2027	0.1631	0.0000	
	2	0.2540	0.0077	0.0206	6.0723	0.0983	0.1158	0.0133	0.0075
	3	0.1548	0.0070	0.0303	5.4655	0.0890	0.1042	0.0134	0.0088
	4	-0.1795	0.0083	0.0632	3.4066	0.1009	0.0650	0.0649	0.0084
	5	-0.7169	0.0135	0.1159	0.6903	0.1013	0.0132	0.1882	0.0080
<i>Role Functioning (RF)</i>	1	0.2026	0.0076	0.0256	8.9165	0.1810	0.1700	0.0000	
	2	0.1211	0.0034	0.0336	8.3224	0.1489	0.1587	0.0000	
	3	0.0832	0.0033	0.0374	5.8072	0.0937	0.1107	0.0000	
	4	-0.1729	0.0051	0.0625	4.0940	0.1111	0.0781	0.0000	
	5	-0.2339	0.0046	0.0685	3.3589	0.1177	0.0640	0.0625	0.0069
<i>Social Functioning (SF)</i>	1	0.2486	0.0063	0.0211	8.5141	0.2082	0.1624	0.0000	
	2	0.1473	0.0044	0.0311	8.1621	0.1568	0.1556	0.0278	0.0072
	3	0.0693	0.0102	0.0387	6.3451	0.1173	0.1210	0.0403	0.0089
	4	-0.1869	0.0071	0.0639	3.0009	0.1144	0.0572	0.0485	0.0085
	5	-0.2783	0.0102	0.0729	2.7880	0.1159	0.0532	0.1143	0.0080
<i>Pain (PN)</i>	1	1.0077	0.0357	-0.0534	8.9402	0.2239	0.1705	0.0000	
	2	0.8581	0.0176	-0.0387	6.3179	0.1054	0.1205	0.0038	0.0070
	3	0.8312	0.0164	-0.0361	5.7752	0.0958	0.1101	0.0038	0.0070
	4	-0.0833	0.0125	0.0537	3.4878	0.0891	0.0665	0.0690	0.0096
	5	-1.0413	0.0257	0.1478	-1.5354	0.1184	-0.0293	0.1369	0.0091
	6	-1.5725	0.0412	0.2000	-1.7232	0.1167	-0.0329	0.2322	0.0090
<i>Mental Health (MH)</i>	1	0.4649	0.0099	-0.0001	8.7341	0.1871	0.1665	0.0000	
	2	0.3152	0.0067	0.0146	8.3783	0.1695	0.1598	0.0000	
	3	-0.0443	0.0093	0.0499	6.5116	0.1152	0.1242	0.0000	
	4	-0.2572	0.0086	0.0708	2.1717	0.1120	0.0414	0.0000	
	5	-0.4786	0.0095	0.0925	1.8528	0.1123	0.0353	0.1471	0.0070

<i>Vitality (VT)</i>	1	0.3703	0.0101	0.0092	8.7847	0.1968	0.1675	0.0000
	2	0.2806	0.0109	0.0180	8.2525	0.1746	0.1574	0.0368 0.0070
	3	0.1133	0.0104	0.0344	7.2308	0.1091	0.1379	0.0368 0.0070
	4	-0.2973	0.0084	0.0747	3.9481	0.1033	0.0753	0.0368 0.0070
	5	-0.4669	0.0099	0.0914	3.3726	0.1095	0.0643	0.1141 0.0086
Death		-2.7171	0.1082	-1.0000			0.0000 N/A	N/A
MOST								-0.1076 0.0082
Disparities (insignificant) 0 (0)				0 (0)		0 (+4)		
Percent Certainty		0.5523		0.7004				
Fit Statistic (RLH)		0.6115		0.5846				
Avg. Residual Variance						0.0021		
pits Estimation		0.3588		0.1972		-0.1002		
Attributes Importance	Rank		Rank		Rank			
	PF	18.00%	2	18.76%	2	20.13%	2	
	RL	6.91%	6	13.26%	5	4.50%	6	
	SF	9.51%	5	13.66%	4	16.60%	3	
	PN	36.80%	1	24.99%	1	32.05%	1	
	MH	14.47%	3	16.42%	3	10.58%	5	
VT	14.32%	4	12.91%	6	16.14%	4		

## DISCUSSION

This online survey study generated models to elicit preference for SF-6Dv2 showed that the quality of the data from different elicitation techniques is not the same even when applying similar restrictions and controlling measures. It also suggested that an estimation of about 30% of the data to be of poor quality when calculating the sample size is a reasonable assumption. Furthermore, the quota requirements were achieved quickly (one month). The quota control and attention checks are critical, causing more than half of the exclusions beforehand, and the length of the survey also affects incompletes.

Homogenous preference Trading with death is important in deciding the values for DCE, considering all dimensions important affects the values in BWS, Online TTO results are very poor, and models' estimations based on it is not reliable and not reasonable even when restricting the data a lot to the most logical answers.

Heterogenous preference DCE results are robust and independent of the effect of inclination toward trading life and death, BWS is still affected by the view of the importance of all dimensions which considers this issue critically related to the estimation technique itself, TTO results remain unreasonable from several points of view as for example giving a positive value to the MOST coefficient which means that the presence of severe levels makes the health state more favorable.

Heterogenous assumptions outperformed the homogenous assumptions in all 3 methods of estimation even when the results are very poor, like in the TTO, as it engaged more information and resulted in better monotonic and significant coefficients.

The ordinal estimation methods had a better fit model than the cardinal methods. The coefficients were more monotonic, more significant, reasonable and even easier subjectively. The TTO data is poor and does not follow the assumptions of the Linear regression (ex. Normal distribution). This means that the answers were random and did not follow any logic. The mathematical solution for the models will complicate the understanding of the estimates and their applicability.

The comparison between the two ordinal methods is very tight, and both of them had a relatively good fit model and coefficients. BWS was slightly better

## CONCLUSION

Online surveys are a promising alternative to face-to-face interviewing methods for collecting and generating health preference values from the general population in China if the proper controls and precautions were considered. The sample should follow a specific quota to represent the general population or be directed purposefully to the group under investigation. Some subgroups have limited access to the internet. Ordinal elicitation methods are a better choice for online surveys than cardinal methods in China but still need quality measures to be included. Self-anchoring using death values from the same task is acceptable to convert latent ordinal values to 0-1 death-full health cardinal values, especially from online data when cardinal methods performance is poor.

# ABSTRACT

## Objectives

Online surveys are widely used to elicit preference in the market research and social sciences in general, but it is still not widely utilized in China by health economists. The experience from an online health states values elicitation survey that was conducted in China is shared herein. The relative performance of three different elicitation methods is reported to further understand their potential usability and conclude recommendations for better application.

## Methods

An online panel of the general population in China participated in a web survey. Each participant was randomized to complete two out of three preference elicitation technique tasks. 8 random Discrete Choice Experiment (DCE) tasks, in which a pairwise choice was firstly made between two health states based on Short Form 6 Dimension version 2 questionnaire (SF-6Dv2), followed by a comparison between the worse of them and death; 6 random Time Trade-Off (TTO) tasks, where a state was decided to be better or worse than death then a trade-off based on years spent in full health is made until the point of indifference; and 10 random Best-Worst Scaling (BWS) tasks, which is basically a forced decision of the best and worst level of one health state, where the three proposed techniques. Comparison to death was included for self-anchoring in the ordinal techniques. Models that assume homogeneous preference (Conditional Logit (CL) and linear regressions (OLS)) and preference heterogeneity (Mixed Logit using Hierarchical Bayes (HB) and Mixed Effects Linear Regression using Hierarchical Bayes (LHB)) were used for analyzing the results. Monotonicity, the number of significant levels and other model fit statistics were employed for the comparison.

## Results

A total of 961 participants (40.6% <40 years, 49 % female, 76.4% healthy) who passed the quality threshold were utilized in the analysis. Data from ordinal methods (DCE & BWS) were more consistent than those from the cardinal method (TTO), which also affected the generated models. Models assuming preference heterogeneity showed a better fit and performance over their comparators assuming homogeneous preference. Between the two cardinal methods, BWS had a better performance than DCE.

## Conclusion

Online surveys to elicit preference in China are acceptable for collecting general population data if quality assurance methods were followed strictly. Cardinal methods still perform worse than ordinal methods in online surveys and thus generate less reliable results.

# REFERENCES

1. Sue VM, Ritter LA. *Conducting Online Surveys*. 2nd ed. Sage; 2012.
2. Denscombe M. Web-Based Questionnaires and the Mode Effect: An Evaluation Based on Completion Rates and Data Contents of Near-Identical Questionnaires Delivered in Different Modes. *Soc Sci Comput Rev*. 2006;24(2):246-254. doi:10.1177/0894439305284522.
3. Evans JR, Mathur A. The value of online surveys. *Internet Res*. 2005;15(2):195-219. doi:10.1108/10662240510590360.
4. Mei B, Brown GTL. Conducting online surveys in China. *Soc Sci Comput Rev*. 2018;36(6):721-734. doi:10.1177/0894439317729340.
5. Ryan M, Scott DA, Reeves C, et al. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess Winch Engl*. 2001;5(5):1-186.
6. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *PharmacoEconomics*. 2008;26(8):661-677. doi:10.2165/00019053-200826080-00004.
7. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271-292. doi:10.1016/S0167-6296(01)00130-8.
8. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x.
9. Culyer AJ, ed. *Encyclopedia of Health Economics*. 1st ed. Elsevier; 2014. <https://www.sciencedirect.com/referencework/9780123756794/encyclopedia-of-health-economics>.
10. Knies S, Evers SMAA, Candel MJJM, Severens JL, Ament AJHA. Utilities of the EQ-5D: Transferable or Not? *PharmacoEconomics*. 2009;27(9):767-779. doi:10.2165/11314120-000000000-00000.
11. Roudijk B, Donders ART, Stalmeier PFM, et al. Cultural Values: Can They Explain Differences in Health Utilities between Countries? *Med Decis Making*. Published online July 2019:0272989X1984158.

doi:10.1177/0272989X19841587.

12. Kularatna S, Whitty JA, Johnson NW, Scuffham PA. Study protocol for valuing EQ-5D-3L and EORTC-8D health states in a representative population sample in Sri Lanka. *Health Qual Life Outcomes*. 2013;11:149. doi:10.1186/1477-7525-11-149.
13. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ*. 2002;11(5):447-456. doi:10.1002/hec.688.
14. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull*. 2012;103(1):21-44. doi:10.1093/bmb/lds020.
15. Whitty JA, Oliveira Gonçalves AS. A Systematic Review Comparing the Acceptability, Validity and Concordance of Discrete Choice Experiments and Best-Worst Scaling for Eliciting Preferences in Healthcare. *The Patient*. Published online November 25, 2017. doi:10.1007/s40271-017-0288-y.
16. Krucien N, Watson V, Ryan M. Is Best–Worst Scaling Suitable for Health State Valuation? A Comparison with Discrete Choice Experiments. *Health Econ*. Published online January 1, 2016:n/a-n/a. doi:10.1002/hec.3459.
17. Comans TA, Nguyen K-H, Mulhern B, et al. Developing a dementia-specific preference-based quality of life measure (AD-5D) in Australia: a valuation study protocol. *BMJ Open*. 2018;8(1):e018996. doi:10.1136/bmjopen-2017-018996.
18. Ratcliffe J, Cameron I, Lancsar E, et al. Developing a new quality of life instrument with older people for economic evaluation in aged care: study protocol. *BMJ Open*. 2019;9(5):e028647. doi:10.1136/bmjopen-2018-028647.